Agroinformatics
Herdon, Miklós
TÁMOP-4.1.2.A/1-11/1-2011-0009
University of Debrecen, Service Sciences Methodology Centre
Debrecen, 2013.
Tartalom

1. Agriculture Systems and Informatics ........................................ 1
   1. Introduction ............................................................................. 2
      1.1. The Importance of Agriculture to Development and the Contribution of ICTs .... 2
      1.2. ICTs, Critical Information Flows, and the Agricultural Knowledge System .... 2
   2. The Feasibility of ICT in Rural Areas ....................................... 4
      2.2. Technical Aspects of ICT Feasibility in Rural Areas ......................... 5
      2.3. Promising Emerging Technologies ......................................... 5
      2.4. Some major ‘ICT’ trends .................................................. 6
      2.5. Some high impact innovations .......................................... 6
         2.5.1. Hardware and Connectivity ........................................ 8
         2.5.2. Ubiquitous Telecommunication Infrastructure ..................... 9
         2.5.3. Utility or "Cloud" Computing ....................................... 9
         2.5.4. Software and Content Management .................................. 9
         2.5.5. Interaction with Biology ............................................. 10
         2.5.6. Biotechnology, Nanotechnology, Materials Science and ICTs ........... 10
   3. Information and Communication Technology in Agricultural Development .... 10
      3.1. Convergence of ICT with agricultural development ................. 10
      3.2. Areas of IT convergence ................................................. 11
      3.3. Drivers of ICT in Agriculture ........................................... 12
         3.3.1. Low-Cost and Pervasive Connectivity .............................. 12
         3.3.2. Adaptable and More Affordable Tools ............................ 12
         3.3.3. New Business Models and Public-Private Partnerships ............. 12
         3.3.4. Democratization of Information, the Open Access Movement, and Social Media 12
         3.3.5. Use Appropriate Technologies ....................................... 13
      4.1. Accessible and Affordable in Rural Areas .............................. 13
      4.2. Key Challenges and Enablers ......................................... 13
         4.2.1. Partnerships .................................................. 13
         4.2.2. Regulation and Policy Challenges ................................. 13
      4.3. Infrastructure .............................................................. 13
         4.3.1. Local Loop or “Last Mile” Connectivity ........................... 14
      4.4. Appliances ................................................................. 14
      4.5. Services ................................................................. 14
      4.6. Anytime, Anywhere: Mobile Devices and Services ..................... 15
   5. Applications in Agriculture .................................................. 15
   6. Questions .............................................................................. 17

2. Hardware Architectures ............................................................ 18
   1. Introduction ............................................................................. 18
      1.1. Data and Information .................................................. 18
      1.2. Information Processing Cycle ....................................... 18
   2. The Components of a computer ............................................ 18
      2.1. Input Devices .......................................................... 18
      2.2. Output Devices ........................................................ 19
      2.3. System Unit .............................................................. 19
      2.4. Storage Devices ......................................................... 19
      2.5. Communications Devices ............................................. 20
   3. The Computer Architecture .................................................. 20
      3.1. Basic Terms .................................................................. 21
      3.2. A data processor ........................................................ 22
   4. The physical PC ..................................................................... 22
      4.1. Internal drivers ............................................................. 22
      4.2. External drivers ............................................................ 22
      4.3. An example of an interface ............................................. 23
7.4. Smartphones ................................................................. 47
8. Functions of Operating Systems ........................................... 47
8.1. User Interface ............................................................. 47
8.2. Job Management .......................................................... 47
8.3. Task Management .......................................................... 48
8.4. Memory Management ....................................................... 48
8.5. File Management ............................................................ 48
8.6. Device Management ........................................................ 49
8.7. Security ............................................................................. 49
9. Categories of Operating Systems ............................................ 49
9.2. Single-User/Multitasking ................................................... 49
9.3. Multituser/Multitasking Operating Systems ......................... 50
9.4. Real-Time Operating Systems ........................................... 50
9.5. 16-, 32-, and 64-Bit OSs .................................................... 50
9.6. Yesterday’s Operating Systems ......................................... 51
9.7. UNIX – The OS for All Platforms ...................................... 51
9.8. Today’s Desktop OSs ....................................................... 51
9.9. Windows Vista ............................................................... 51
9.10. Windows 7 ................................................................. 52
9.11. Windows 8 ................................................................. 52
10. Questions ............................................................................ 52
4. Advanced Spread Sheet System Applications .......................... 53
1. Introduction ......................................................................... 53
2. Using Excel formulae ........................................................... 55
3. Excel functions ................................................................. 55
  3.1. Category Includes .......................................................... 56
  3.2. Copying and pasting formulas ........................................... 57
  3.3. The order of processing of formulae .................................. 57
  3.4. Priority Operator Description .......................................... 57
  3.5. Using relative and absolute cell references ......................... 58
  3.6. Advanced Excel Tips ...................................................... 61
4. Working with charts and graphs .............................................. 61
  4.1. Adding legends to your chart or graph ............................... 63
  4.2. Editing charts and graphs ............................................... 63
  4.3. Editing a chart’s data range ............................................. 64
5. Absolute vs. Relative References Beyond the Basics ................ 64
  5.1. Basics Reviewed ........................................................... 64
  5.2. Beyond the Basics ......................................................... 64
  5.3. Partially Absolute Cell References ..................................... 64
  5.4. Range Name and Sheet Name References ......................... 65
  5.5. AutoSum Drop Down Options ......................................... 65
6. Working with Styles ........................................................... 65
  6.1. Changing Defaults for Workbooks and Worksheets ............. 65
  6.2. Create a File You Always Start With ................................. 66
  6.3. Create a Template to Use for New Workbooks .................. 66
  6.4. Create a Template that is Opened Automatically ................. 66
7. Questions ............................................................................ 66
5. Database system ..................................................................... 67
  1. Introduction ....................................................................... 67
  2. Organized Collection of Symbols ........................................ 67
  3. Collection of Variables ...................................................... 67
  4. The Relational Database .................................................... 68
  5. “Relation” Not Equal to “Table” .......................................... 69
  6. Anatomy of a Relation ...................................................... 70
  7. The DBMS ................................................................. 71
8. Database Language ............................................................ 71
9. Relational Algebra – The Foundation ........................................ 72
10. Access-2007 ................................................................. 72
  10.1. Basic Terminology ....................................................... 72
1. Introduction ........................................................................................................ 103
2. A Comparison of the OSI and TCP/IP Reference Models ................................. 103
3. The IP Protocol ................................................................................................... 105
   3.1. IP Addresses ................................................................................................. 106
   3.2. The Application Layer ................................................................................ 107
   3.3. DNS – The Domain Name System ................................................................ 107
4. The DNS Name Space ......................................................................................... 108
5. Electronic Mail .................................................................................................... 109
   5.1. Architecture and Services ........................................................................... 111
   5.2. The User Agent ............................................................................................. 111
      5.2.1. Sending E-mail ....................................................................................... 111
      5.2.2. Reading E-mail ..................................................................................... 111
      5.2.3. MIME – The Multipurpose Internet Mail Extensions .......................... 111
   5.3. Final Delivery ................................................................................................. 112
      5.3.1. POP3 ..................................................................................................... 112
      5.3.2. IMAP .................................................................................................... 112
      5.3.3. Webmail ............................................................................................... 113
6. The World Wide Web .......................................................................................... 113
   6.1. Architectural Overview ................................................................................ 113
      6.1.1. The Client Side ....................................................................................... 114
      6.1.2. The Server Side ..................................................................................... 116
   6.2. URLs – Uniform Resource Locators ............................................................ 116
   6.3. Static Web Documents ................................................................................ 117
   6.4. HTML – The HyperText Markup Language ............................................... 117
   6.5. XML and XSL .............................................................................................. 119
   6.6. Dynamic Web Documents .......................................................................... 119
      6.6.2. Client-Side Dynamic Web Page Generation ......................................... 120
   6.7. HTTP – The HyperText Transfer Protocol .................................................. 120
   6.8. The Wireless Web ........................................................................................ 120
7. Multimedia .......................................................................................................... 121
   7.1. Introduction to Digital Audio ........................................................................ 122
   7.2. Audio Compression ....................................................................................... 122
   7.3. Streaming Audio ........................................................................................... 122
   7.4. Internet Radio ............................................................................................... 122
   7.5. Voice over IP ................................................................................................. 123
   7.6. Video on Demand ......................................................................................... 123
   7.7. The MBone – The Multicast Backbone ....................................................... 125
8. Questions ............................................................................................................. 125
9. Information Systems in Agriculture ................................................................... 127
   1. Introduction ...................................................................................................... 127
   2. Information systems and communication networks ........................................ 127
   3. Management Information Systems .................................................................. 128
   4. Internal Information Systems .......................................................................... 133
      4.1. On-Farm Information Systems – Computer Hardware ............................ 133
      4.2. On-Farm Database and Modelbase Applications ................................... 133
      4.3. User Interface ............................................................................................ 134
      4.4. The Decision Maker ................................................................................. 134
   5. External Information Systems .......................................................................... 134
      5.1. Internet ....................................................................................................... 135
      5.2. Satellite Data Transmission Systems ....................................................... 135
      5.3. Outside Advisors ....................................................................................... 135
   6. The Future ......................................................................................................... 135
      6.1. Structured Decisions .................................................................................. 135
      6.2. Structured and Unstructured Decisions .................................................... 137
      6.3. Summary .................................................................................................... 138
7. Farm Management Information System (FMIS) ................................................ 138
   7.1. Knowledge Management in Agriculture ...................................................... 139
   7.2. Adoption of Knowledge Management ........................................................ 142
2.4. People use Rules ................................................................. 201
2.5. Problem Solving Requires Focused Knowledge ....................... 202
  3.1. Data ............................................................................... 202
  3.2. User Interface ................................................................. 202
  3.3. User ............................................................................... 202
  3.4. Knowledge Base/Rule Base ............................................... 203
  3.5. Inference Engines ............................................................ 203
  3.6. Expert System Software .................................................... 203
4. What is an Expert System? ............................................................ 203
  4.1. Rule-based Knowledge Representation .................................... 204
  4.2. Activity/Task of the system ............................................... 204
  4.3. Definitions ....................................................................... 204
5. Characteristics of Expert System Applications ............................... 204
  5.1. Structured vs. Unstructured Tasks ........................................ 204
  5.2. Support vs. Replace .......................................................... 205
  5.3. Available Time .................................................................. 205
6. Applications .............................................................................. 205
7. Knowledge Acquisition ................................................................. 205
8. Knowledge Representation ............................................................. 205
  8.1. "If…then…” Rules ............................................................ 205
9. Explanation .............................................................................. 205
  9.1. Importance of Explanation Facilities ...................................... 206
  9.2. Trace Through the Rules ..................................................... 206
  9.3. Model-based Reasoning ...................................................... 206
  9.4. Dialog-Based Systems ....................................................... 206
  9.5. Verification and Validation .................................................. 206
10. Expert System Strengths and Limitations ...................................... 207
  10.1. Strengths ....................................................................... 207
  10.2. Limitations ..................................................................... 207
  11.1. Knowledge from Data ....................................................... 207
  11.2. Alternative Forms of Knowledge Representation .................... 208
  11.3. Alternative Problem Solving Approaches .............................. 208
  11.4. Expertise .................................................................... 208
  11.5. Uncertainty Representation ................................................. 208
  11.6. The Internet and Connecting Systems .................................. 208
  11.7. Ontologies ................................................................... 208
  11.8. Embedded Intelligence vs. Stand Alone Systems .................... 208
  11.9. Business Rules ................................................................ 209
12. Conclusion .............................................................................. 209
13. Questions ............................................................................... 209
13. eGovernment Services in Agriculture ........................................ 210
  1. Introduction ...................................................................... 210
  2. Definition ......................................................................... 210
  3. Main characteristics of electronic government .............................. 213
  4. The levels of maturity of eGovernment ..................................... 215
  5. Multi-channel access ............................................................ 215
  6. Customer-centred thinking ..................................................... 216
  7. The advantages of eGovernment ............................................. 217
    7.1. The supply and quality of information improves. ............... 217
    7.2. Procedure time decreases. ............................................. 217
    7.3. Administrative burdens decrease. .................................... 217
    7.4. Costs can be reduced. ................................................... 218
    7.5. Higher level services ..................................................... 218
    7.6. E-democracy opens up new opportunities for the public. .... 218
    7.7. Networks organised from the grass roots up are created through the self-organisation of citizens. .... 220
  8. Measuring the effects of eGovernment ....................................... 220
  9. Key terms ......................................................................... 221
Az ábrák listája

1.1. Figure 1.1: Information Flows in Agriculture .......................................................... 4
1.2. Figure 1.2: Growing Service System ......................................................................... 5
1.3. Figure 1.3: The Nature of Innovation ....................................................................... 7
1.4. Figure 1.4: Grid System in Agriculture ..................................................................... 8
2.1. Figure 2.1: CPU Test Program .................................................................................. 18
2.2. Figure 2.2: The Main Memory .................................................................................. 19
2.3. Figure 2.3: The Neumann Model ............................................................................. 20
2.4. Figure 2.4: The main parts of PC ............................................................................ 21
2.5. Figure 2.5: The Winchester ..................................................................................... 23
2.6. Figure 2.6: A specific example of an interface ......................................................... 23
2.7. Figure 2.7: The motherboard is the hub of all data exchange ................................... 24
2.8. Figure 2.8: A motherboard is a board covered with electronics ............................ 24
2.9. Figure 2.9: CPU Test Program ................................................................................ 25
2.10. Figure 2.10: The buses ......................................................................................... 26
2.11. Figure 2.11: Bridges connect the various buses together ....................................... 27
2.12. Figure 2.12: The Front Side Bus ............................................................................ 27
2.13. Figure 2.13: Chip Sets .......................................................................................... 27
2.14. Figure 2.14: ulink between Bridges ...................................................................... 28
2.15. Figure 2.15: The North Bridge .............................................................................. 28
2.16. Figure 2.16: The South Bridge .............................................................................. 28
3.1. Figure 3.1: Systems Today ...................................................................................... 31
3.2. Figure 3.2: Main modules of Operating System ...................................................... 33
3.3. Figure 3.3: Bootstrap process .................................................................................. 37
3.4. Figure 3.4: File Root System ................................................................................... 39
4.1. Figure 4.1: Tables and Rows ................................................................................... 53
4.2. Figure 4.2: Exploring Excel .................................................................................... 53
4.3. Figure 4.3: Features and descriptions ..................................................................... 53
4.4. Figure 4.4: Spreadsheet formulas ........................................................................... 55
4.5. Figure 4.5: Spreadsheet formulas ........................................................................... 55
4.6. Figure 4.6: Functions ............................................................................................. 56
4.7. Figure 4.7: Absolute reference ............................................................................... 38
4.8. Figure 4.8: Relative reference ............................................................................... 38
4.9. Figure 4.9: Common charts elements ..................................................................... 62
4.10. Figure 4.10: Chart styles ....................................................................................... 62
4.11. Figure 4.11: Formatting columns .......................................................................... 63
5.1. Figure 5.1: Anatomy of a relation ........................................................................... 70
5.2. Figure 5.2: Creating Table ..................................................................................... 73
5.3. Figure 5.3: Creating criteria for Query .................................................................... 73
5.4. Figure 5.4: Datasheet View ...................................................................................... 75
5.5. Figure 5.5: Creating Table ..................................................................................... 78
5.6. Figure 5.6: Creating Database from Template ....................................................... 79
5.7. Figure 5.7: Access Objects ...................................................................................... 80
5.8. Figure 5.8: Using Wizard in Creating Report .......................................................... 81
5.9. Figure 5.9: Modifying Report Layout ...................................................................... 82
6.1. Figure 6.1: A network with two clients and one server .......................................... 84
6.2. Figure 6.2: The client-server model involves requests and replies .......................... 84
6.3. Figure 6.3: In a peer-to-peer system there are no fixed clients and servers ............ 85
6.4. Figure 6.4: Combinations of wireless networks and mobile computing ............... 87
6.5. Figure 6.5: Classification of interconnected processors by scale .......................... 88
6.6. Figure 6.6: Two broadcast networks (a) Bus; (b) Ring .......................................... 88
6.7. Figure 6.7: A metropolitan area network based on cable TV ................................ 89
6.8. Figure 6.8: Relation between hosts on LANs and the subnet ................................... 90
6.9. Figure 6.9: A stream of packets from sender to receiver ......................................... 91
6.10. Figure 6.10: (a) Bluetooth configuration; (b) Wireless LAN ................................. 92
6.11. Figure 6.11: (a) Individual mobile computers; (b) A flying LAN ......................... 93
6.12. Figure 6.12: Layers, protocols, and interfaces ....................................................... 95
<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>14.8</td>
<td>Evaluation to IMS</td>
<td>231</td>
</tr>
<tr>
<td>14.9</td>
<td>Application in Vineyard</td>
<td>232</td>
</tr>
<tr>
<td>15.1</td>
<td>Agriculture and Food Sectors</td>
<td>234</td>
</tr>
<tr>
<td>15.2</td>
<td>Levels of Precision Farming</td>
<td>235</td>
</tr>
<tr>
<td>15.3</td>
<td>Wireless Sensor Network</td>
<td>236</td>
</tr>
<tr>
<td>15.4</td>
<td>Precision Farming through Satellite</td>
<td>237</td>
</tr>
<tr>
<td>15.5</td>
<td>Monitoring System for Farm Operating</td>
<td>237</td>
</tr>
<tr>
<td>15.6</td>
<td>Traceability Systems and ulinkage with PLF</td>
<td>239</td>
</tr>
<tr>
<td>15.7</td>
<td>Some information flows</td>
<td>241</td>
</tr>
<tr>
<td>15.8</td>
<td>ICT and the Rural Finance Ecosystem</td>
<td>248</td>
</tr>
<tr>
<td>15.9</td>
<td>Innovation Funnel</td>
<td>251</td>
</tr>
</tbody>
</table>
A táblázatok listája

1.1. Table 1.1: Evolving technologies for ICT applications in agricultural and rural areas ............ 6
5.1. Figure 5.1: An Organized Collection of Symbols ........................................................................ 67
5.2. Figure 5.2: A variable, showing its current value ...................................................................... 68
5.3. Figure 5.3: A relation, shown in tabular form ........................................................................... 68
5.4. Figure 5.4: Same relation as Figure 5.3. ................................................................................... 69
5.5. Figure 5.5: Not a relation ........................................................................................................ 70
Tárgymutató

AGROINFORMATICS

TEXTBOOK

WRITTEN BY
MIKLÓS HERDON

2013
1. fejezet - Agriculture Systems and Informatics

1. Introduction

There is scarcely a field of human activity today that has not been touched by the dramatic changes in Information and Communication Technologies (ICTs) that have taken place in the last 10-15 years. Agriculture and agriculture-related natural resource management are no exceptions. Information and the technologies that facilitate its use, exchange, and reliability have been important aspects of agriculture and agriculture-related natural resource management for centuries. Decisions on what to plant, when to plant it, how to cultivate and harvest, and where to store and sell and at what price have long depended on knowledge, communication, and information exchange. The importance of information and communication technologies to agriculture is not new, and many traditional methods of managing and communicating information will continue in the future. Some recent ICTs are offering new opportunities, however, to increase the timeliness and availability of critical information, improve its quality and relevance, and offer more cost-effective methods for empowering and ensuring feedback from previously marginalized communities. In addition, the emergence of global agricultural production chains interlinked by digital networks has important implications for the livelihoods of those presently outside of the system.

ICT for Development Strategic Plan defines information and communication technology as the combination of hardware, software, and the means of production that enable the exchange, processing, and management of information and knowledge. ICTs thus include technologies and methods for storing, managing, and processing information (e.g., computers, software, books, PDAs, digital and non-digital libraries) and for communicating information (e.g., mail and email, radio and television, telephones, cell phones, pagers, instant messaging, “the web”, etc.).

In everyday speech, ICTs commonly refer to electronic and digital devices and the software used for storing, retrieving, and communicating information. However, the poorest and most vulnerable populations may have little opportunity or capacity to use or benefit from ICTs so narrowly defined. Broadening the definition to include some older, more traditional technologies and methods (e.g., accounting ledgers, couriers, radio, television, face-to-face training) allows the discussion to focus on the needs of agricultural communities and applicability of new technologies while simultaneously including more technologies available to the rural poor.

1.1. The Importance of Agriculture to Development and the Contribution of ICTs

The agricultural sector is the most critical economic sector in many countries. More than half of the developing world’s population lives in rural areas and is economically dependent on the performance of agricultural production. Agriculture provides food, is generally the largest market for labor, provides tradable goods – hence foreign exchange – for the national economy, and contributes to government budgets through taxation. Long-term improvements in living standards for the rural poor require both resources and innovations to facilitate access to new markets and improve production capacity. ICTs have important roles in each of these areas. Improving agricultural performance is also a prerequisite for economic growth and creation of a stable environment for democracy. Agricultural innovation is understood today to be the result of an interacting constellation of agricultural actors: not just public agencies such as the extension network, but also private firms, NGOs, farmer associations, and others. In this context, ICTs are more than simply a tool to make each entity individually more productive; ICTs offer methods for weaving agricultural actors together into networks that can collectively identify, modify, act on, and implement relevant innovations.

1.2. ICTs, Critical Information Flows, and the Agricultural Knowledge System

The variety of new ICT tools for agriculture is impressive, but the tools need to be placed in an overall context of agricultural information and communication needs. By looking at the critical information needs of agriculture and farming communities, the focus can move away from a compendium of “neat gadgets” and their individual applications toward understanding of their overall role in promoting productive, equitable, and sustainable
agriculture. The key framework for this is the Agricultural Knowledge System (AKS), consisting of the organizations, sources of knowledge, methods of communication, and behaviors surrounding an agricultural process. Knowledge is not the same as information: knowledge includes information, understanding, insights, and other information that has been processed by individuals through learning and thought.

Information exchange in the local knowledge system is generally by non-digital means: face-to-face discussions, printed pamphlets, videocassettes, radio broadcasts, etc. Local communities may lack affordable power and communication systems to drive ICTs, or they may need investments in human capacity to maintain them. Increasingly, some communities will begin to have access through such services as cellular phones, rural use of battery or solar-powered personal digital assistants (PDAs) or local telecenters/cybercafés run out of local organizations.

Agricultural knowledge and information needs to be managed like any other key business input. Advances in ICTs have helped create an entirely new discipline, termed knowledge management. Effective knowledge management means that an organization or network of partners gets the right information to the right person at the right time in a user-friendly and accessible manner so that they can perform their jobs efficiently.

Development efforts must improve the capacity of the agricultural knowledge system to manage and disseminate knowledge effectively, particularly to small farming families and women. ICTs can play an important role in unlinking knowledge seekers to knowledge sources. Agricultural research, extension, and development organizations – public or private, for-profit or non-for-profit – are all part of an overall agricultural knowledge system unlinked by information and disseminating important agricultural information to farmers has been an integral part of agricultural development strategies for years. In an ICT-enabled approach, information dissemination from institutionalized knowledge sources will continue to be important, but the real transformation that ICTs make possible is to allow feedback and “return flows” of information from users that tell information suppliers whether the information they supply is useful or relevant and offer guidelines to improve it. Promoting knowledge feedback from rural communities does not necessarily require a connection to the Internet; paper surveys, mailed floppy disks, telephone voice menus, PDAs, and other methods are also options.

The role of the intermediate organization is that of a knowledge management organization whose purpose is to introduce change for the benefit of the clients. These organizations generally avoid issuing prescriptive recommendations; rather, they play an advisory and facilitating role. In short, intermediate organizations are organizations that provide management of information, communication, and knowledge that allow farmers and farmer groups to make better management decisions that will improve their long-term livelihoods.

The critical point is that these knowledge processes do not occur automatically. This is a new role for many support organizations and their staff. Intermediate organizations need training and support in problem diagnosis, problem solving, participatory decision-making, organizing, convening, and motivating all attributes of a successful facilitator. Such training should include how to seek information via the Internet, how to take that information and modify it for their (farming) clients, how to interact with farming households and facilitate knowledge use, and how to follow-up, gather information regarding farms’ experiences, and relay this back to the original information providers and institutionalized knowledge sources.

Designing ICT-enabled knowledge flows between these actors in any specific case requires careful consideration of the types of ICTs accessible by each group and the technological and conceptual packaging of information so that it can flow effectively from one user to the other. Effective ICT deployment explicitly considers the appropriate interfaces between the digital and non-digital worlds, so that those without access to digital ICTs can still benefit from an improved local information environment. From the perspective of the smallholder farmer, the key question is how to gain access to information and resources. These farmers need local support groups that will act as brokers between the available knowledge system and the individual needs of farming households. Developing economical local access for the rural poor and ensuring appropriate content is the essence of bridging the digital divide.

ICTs can accelerate agricultural development by providing more accessible, complete, timely, or accurate information at the appropriate moment to those making key on the ground decisions. Examples of such decisions are: what and when to plant; where to locate agricultural inputs (and at the best price); how to identify and respond to disease, pests, and drought; where to sell products; what new technology options exist for production, post- harvest, and soil fertility control; what agricultural credit programs are available; and how to access relevant government programs, including land titling.
Digital ICTs can add value over traditional methods when: Information is time sensitive (e.g., prices) Information requires significant customization to satisfy a client’s need (e.g., soils content, local policies). The information needed involves standardized calculations (e.g., credit evaluations). Knowledge requires significant back-and-forth interactivity over distances (e.g., locating a remote specialist for disease/pest diagnosis and treatment).

Information and knowledge flows that are important to farming households as they make key agricultural decisions throughout the year. These include knowledge and information about agricultural technologies and methods; the local natural resource base and geography; the policy environment, laws, and regulations; and market information (Figure 1.1).

1.1. ábra - Figure 1.1: Information Flows in Agriculture

2. The Feasibility of ICT in Rural Areas

Some development planners have been skeptical of the cost and benefits of ICT-enhanced strategies over traditional modes of agricultural development assistance since the beginning of the digital revolution. Equipment costs, the technical infrastructure requirements to support PCs, and electrical scarcity in rural areas may at first make ICT investments seem uneconomical. On closer examination, one can easily overestimate the costs and underestimate the aggregate benefits that ICTs can bring, particularly if one starts from the assumption that ICT interventions are limited to desktop computers connected to the Internet. This chapter discusses financial considerations, long-term sustainability, technical issues, rural power requirements and some emerging technologies that are likely to impact the costs, benefits, or applicability of ICTs in rural areas.


Investments in ICTs look quite inexpensive compared to the cost of large infrastructure investments such as hydroelectric dams and road systems. However, they can seem quite costly compared to the average income of a poor family. On balance, well-planned ICT-enhanced interventions tend to boost the impact and longevity of development assistance, while simultaneously assisting with monitoring and knowledge gathering in project activities. These effects occur because the number of uses for ICTs tends to increase over time as more users become familiar with the technology and as new ICT-based services or content become available.
Connectivity approaches seek to provide target groups with new or upgraded access to ICT equipment and communications capabilities so they may connect to information networks and process the information they need. In rural areas, these approaches often include power systems for running electronic equipment off of the electrical and communications grid. Connectivity provision can be the most difficult activity to undertake cost-effectively, because up-front equipment costs can be substantial. ICT hardware may depreciate rapidly, and servicing and maintenance usually demands that effective management practices be in place.

2.2. Technical Aspects of ICT Feasibility in Rural Areas

Some of the most important hardware considerations in ICT and ICT-enhanced projects were covered in the preceding section, which emphasized the importance of moving beyond a computers-laptops-Internet model when using ICT to integrate information systems for agriculture. When working with international agricultural researchers, national government agencies, universities, and other more “elite” organizations, desktop/laptop-cum-Internet and Internet portal approaches make sense, but more remote, less affluent, and less literate areas may demand alternative hardware systems that suit local needs and capacities.

A key concept for linking central computing centers to remote needs is server-side processing. Server-side models of information processing allow users with relatively basic equipment (e.g., older computers, PDAs, cell phones, pagers, or other “thin clients”) to take advantage of powerful computers on the far side of a telecommunications connection. The more powerful computers accept simple commands and small quantities of data, process it (often in conjunction with large data sets of its own), and communicate only the results back to the more simple equipment. In this manner, a relatively simple computer or PDA can access powerful GIS software and large databases over the web. The key to using these technologies effectively is that the software on the more powerful end must be designed to know that it will communicate with less powerful clients, and be able to handle requests from devices other than PCs.

2.3. Promising Emerging Technologies

Even after the technology bubble of the 1990s, many information and communication technologies are advancing rapidly even as technology prices continue to fall. These trends suggest that more and more ICTs will become affordable at any specific income level over time and the service sector is increasing (Figure 1.2). High-cost technologies concentrated in capital cities and regional centers are not necessarily inaccessible by remote communities, since server-side processing models can often facilitate shared access.

1.2. ábra - Figure 1.2: Growing Service System
Falling costs will help to increase the availability of ICTs, but the demand for technologies and the services they enable also depend on the quality of content, its transparency, and reliability. Ease of use and attention to user interface designs will prove essential to adoption and will require explicit integration into ICT-enabled agriculture projects. With these caveats, Table 1.1 identifies some important technologies entering the ICT and agriculture mainstream, as well as some others which are on the horizon and may become more widely applicable over the next few years.

1. táblázat - Table 1.1: Evolving technologies for ICT applications in agricultural and rural areas

<table>
<thead>
<tr>
<th>Some technologies in or entering the mainstream</th>
<th>Promising technologies on the horizon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database-driven websites</td>
<td>Wireless (VSAT, Wi-Fi, Bluetooth) Portable flash media</td>
</tr>
<tr>
<td>Digital Photography</td>
<td>DVD burning and design</td>
</tr>
<tr>
<td>Server-side/distributed computing</td>
<td>Biometrics</td>
</tr>
<tr>
<td>Cellular phones</td>
<td>Voice-recognition/text-to-speech</td>
</tr>
<tr>
<td>Short-Message Service (SMS) Instant Messaging</td>
<td>Translation software</td>
</tr>
<tr>
<td>Geographic Information Systems (GIS) PhotoVoltaic (PV) power for ICTs</td>
<td>Fuel Cells for ICTs</td>
</tr>
</tbody>
</table>

2.4. Some major ‘ICT’ trends

Arising from the discussions, participants highlighted the following significant trends that agricultural science will need to pay attention to:

- Information and communication technologies, devices and software are becoming much cheaper and more affordable, even in rural areas where ICTs are increasingly available.

- Connectivity is becoming more pervasive and ‘mobile’ – people can connect and interact in real time with other people and data across a broad range of wireless, mobile and other devices. And more and more of the devices are becoming smart and intelligent – capable of multiple operations.

- Geo-spatial and ‘neogeographic’ functionalities, applications and tools are spreading and becoming ubiquitous, offering pinpoint location and data collection and sharing possibilities.

- More and more services will be provided across the Internet through so called ‘cloud’ computing, obviating the need for sophisticated local ICT systems and capabilities.

- As the quantity of data and information grows, new ways to organize, navigate, mine, share, visualize, and ‘mash’ it up will emerge, creating new possibilities and services.

- Digital applications and tools are being applied to enable and extend traditional ‘human’ processes like communication, collaboration and analysis.

- Within agriculture and science, new thinking and approaches are emerging around ‘end user innovation’, focused on knowledge, value chains, innovation systems, etc.

- Scientists increasingly use and depend on ICTs in their daily work: computers, enhanced ICT literacy, and connectivity are part of the ‘basic’ package.

2.5. Some high impact innovations

Develop good M&E system around ICT use in agricultural science to be able to track, learn and adjust along the way. We need to know if these tools are really working toward more effective, efficient and impactful work (Figure 1.3).
1.3. ábra - Figure 1.3: The Nature of Innovation

Build up and support the right mix of personnel with the right skills to be able to work with ICTs in agricultural science for development. We need both new curriculum to support this as well as ongoing capacity building opportunities to keep people ‘on the ball.’ Incentives – if people are going to be engaging in clearly beneficial work to their institutes – producing non-traditional research outputs, then we need to find a way to recognise and reward them. Make sure when we introduce and use new tools that they have a clear purpose. ICTs need to advance us along the impact pathway.

Catching and successfully harnessing these ‘waves’ requires strategic investments in capacities, bandwidth and infrastructure, skills, tools and applications, and the adoption of an ‘open innovation’ mindset that breaks barriers, ulinks data and knowledge, and guarantees the public accessibility of goods generated and captured through science.

What are some of the trends and changes we can expect in the coming years?

• Increasingly ‘ubiquitous’ connectivity along value chains – We will all make use of a range of devices and platforms to access and share knowledge: From the web to phones, radio, video and text messaging. Most scientists will work in knowledge-rich environments; farming communities, probably using different devices, will be far more connected than at present. Multiple connectivity paths widen the potential reach of science.

• Increasingly ‘precise’ applications and tools – ICTs and digital signatures or labels of various types will be used to track products from producer to consumer; to monitor local soil, weather and market conditions; to tailor data and information services to the demands of a specific audience or individuals. Applications will come in many shapes and sizes, to suit even the most specialized needs.

• Increasingly ‘accessible’ data and information – Vast quantities of public data and information held by institutions and individuals will become visible and re-usable at the click of a device. More intermediary skills and applications will be needed to help harvest, make sense of, and add value to these layers of data and information.

• Increasingly ‘diverse’ set of applications available across digital clouds – The digital ‘identities’ of scientists and their collaborators will give them access to a wide range of online tools and applications, accessible from any location and across different devices, enabling collaboration across boundaries as never before. Local firewalls and server configurations conditions will not restrict global sharing.

• Increasingly ‘inter-connected’ tools and knowledge bases – Different communities and their knowledge will be able to connect and share with each other, along the research cycle and across disciplines, including people with different engagement in science such as farmers, traders, politicians. A whole new breed of products and services will emerge to inter-connect and represent diverse knowledge.
In general, the most significant impact of ICTs on agricultural technology generation will be in connecting and engaging communities in participatory agricultural innovation. Science will be able to come out of its ‘silos.’ New agricultural processes and technologies to solve agricultural problems will emerge through continuous innovation with user communities, thus eliminating many of the constraints that agricultural science, research and technology generation now face. The need for conventional extension from research stations to farmers’ fields will diminish. Agricultural innovations will best fit the needs of user communities. What are some of the changes needed to move in these directions? These include:

- Improve communications infrastructure and bandwidth, investing in lower-cost hardware, software and applications that connect science right along the development chain.
- Increase and improve formal education and training in information and communication sciences that contributes to innovation in the use of new ICTs in agriculture.
- Extend the generation and dissemination of data and information content as a ‘public good’ that is widely accessible and is licensed to be easily re-used and applied.
- Support applications that integrate data and information or foster the interoperability of applications and information systems, allowing safe and ethical access while protecting necessary rights.
- Encourage the effective uptake and use of data, information and knowledge, particularly focusing on capacity building dimensions necessary for the outputs of science to have impacts.
- Support innovation in the workflows, processes and tools used to create, share, publish, visualize, and connect the outputs of agricultural science and the people engaged in it.

2.5.1. Hardware and Connectivity

The Moore’s law that the number of transistors that can be placed inexpensively on an integrated circuit is growing exponentially, the number of transistors doubling approximately every two years, has so far held. The same law can be applied to processing speeds of microprocessors, memory capacity and the number of pixels that a digital camera can process. Memory storage capacities in magnetic and optical media have also increased exponentially and solid state drives are already commercialised. Connectivity between computers and through the Internet has similarly increased in bandwidth. The rates at which data can be transmitted, both within buildings and across long distances, grows without apparent limit and ever reducing costs. Parallel and Grid computing (Figure 1.4) has demonstrated huge potentials of processing power available for use on the desktop of an average computer user and this will be multiplied many folds with memristors (already prototyped), photonic and quantum computers (still in the research phase). We are seeing a boom in handheld devices that interface with existing systems.

1.4. ábra - Figure 1.4: Grid System in Agriculture
2.5.2. Ubiquitous Telecommunication Infrastructure

Flowing from the falling costs of all things digital, there has been a steady flow of investment into communications infrastructure around the world. Cell phone and broadband (wired and wireless) Internet networks carrying both voice and data are being deployed in even the poorest countries and with time will expand to cover most rural areas. These systems are sophisticated and manageable by both private and public entities, allowing agriculture and agricultural research to increasingly take communications for granted and being continuously improving in the years ahead.

2.5.3. Utility or "Cloud" Computing

The combination of progress in computing hardware, system software, and Internet communications has now enabled the construction of general-purpose data centres that can be reconfigured by command to support any software application in minutes. There are already data services that allow a user to have hundreds or thousands of computers at their command, and yet pay for them by the hour or minute, without owning or operating the hardware themselves. The costs are far less than even falling hardware prices would suggest, since the cost of the data centre can be shared among many "bursty" users. In effect, the data centre acts like a utility, providing as much computing as requested at just the times when needed. Since these data centres are invariably shared over the Internet, they are sometimes called computing "in the cloud." These "cloud" data centres are the natural repository for shared data sets, so that users in any location or institution can instantly access, analyze and interpret public information goods without the need to move the data to their own facilities. This can enable a researcher in any location to work with data as well as any other researcher, which can lead to new kinds of collaboration and new sources of project direction.

2.5.4. Software and Content Management

A far more important frontier achieved through more complex processors, processing speeds, memory capacity and connectivity has been the development of agents, sensors and devices such as radio frequency identification tags (RFIDs) that is now reshaping how humans work and interact creating huge potentials in terms of how we can mediate share and extract value from information and knowledge. Among the others, the semantic web and its related techniques and applications (e.g. ontologies) are currently working in this sense, trying to re-shape machine-to-machine interaction and the way computers retrieve, manage and share knowledge on the web.

The science of pragmatics – the practical interpretation and use of signs by agents or communities within particular circumstances and contexts – and going beyond conventional semantics, is now allowing ICTs to be used in much more supportive ways. This has been demonstrated in diverse areas such as health, scientific research and business management in modelling, simulation, forecasting and visualization and has implications
for agriculture. These potentials bring new challenges on how we understand this new pervasive computing landscape and how we can make use of collective and distributed form of intelligence.

2.5.5. Interaction with Biology

The interaction of ICT with biology, biotechnology, nanotechnology and new materials is enabling the development of high quality information that is created from diverse entities and sources and which is self organizing. This self-organizing collective intelligence – living information – presents new frontiers in effective use and application. Continuous advances in ICT and biology are enabling developments where the relationship between these two disciplines faces a paradigm shift; from ICTs that mimic biology to ICT that use biology for information processing. Progress in synthetic biology – the study of the design and building of novel biological functions and systems – is bringing progress in systematic design methodologies and manufacturing processes. The potential of interfacing ICT with biological systems at the micro/nano scale is now emerging.

2.5.6. Biotechnology, Nanotechnology, Materials Science and ICTs

It can be argued even with current knowledge that in future Bio- and Nano- technology, Material Sciences and ICTs together will define the core direction of agricultural science, research and technology by having impact on plant and animal breeding and improvement, agricultural production systems, risk management and aversion, sustainable use of natural resources, protecting the environment and agricultural market chains and in agricultural innovation in general.

3. Information and Communication Technology in Agricultural Development

Today a new paradigm of agricultural development is fast emerging: in both developing and developed countries the overall development of rural areas is expanding in new directions; old ways of delivering important services to citizens are being challenged; and traditional societies are being transformed into knowledge societies all over the world.

ICT in the revival of social organisations ICT can give a new impetus to the social organisations and productive activity of agriculture which, if nurtured effectively, could become transformational factors. The ‘knowledge’ itself will become a technology for overall agricultural development. Agricultural extension, in the current scenario of a rapidly changing world, has been recognised as an essential mechanism for delivering knowledge (information) and advice as an input for modern farming. However it has to escape from the narrow mindset of transferring technology packages to transferring knowledge or information packages. If this can be achieved, with the help of ICT, extension will become more diversified, more knowledge-intensive, and more demand driven, and thus more effective in meeting farmers’ information needs. ICT has many potential applications in agricultural extension. It can bring new information services to rural areas where farmers, as users, will have much greater control than before over current information channels. Access to such new information sources is a crucial requirement for the sustainable development of the farming systems.

3.1. Convergence of ICT with agricultural development

Broad basing agricultural extension activities; developing farming system research and extension; having location-specific modules of research and extension; and promoting market extension, sustainable agricultural development, participatory research, etc. are some of the numerous areas where ICT can play an important role. Several research studies conducted on extension organisations have revealed that the delivery of goods is effective when the grass roots extension worker covers a small area of jurisdiction, with multiple purposes (broad basing). The existing system of large jurisdictions, each with a narrow range of activities, is less effective. However, broad basing requires grass roots workers to be at the cutting edge of extension and master of many trades, which is not really possible. It can help here, by enabling extension workers to gather, store, retrieve and disseminate a broad range of information needed by farmers, thus transforming them from extension workers into knowledge workers. The emergence of such knowledge workers will result in the realisation of the much talked about bottom-up, demand driven technology generation, assessment, refinement and transfer. Agricultural extension systems in most developing countries are under-funded and have had mixed effects. Much of the extension information has been found to be out of date, irrelevant and not applicable to small farmers’ needs, leaving such farmers with very little information or resources to improve their productivity. ICT helps the extension system in re-orienting itself towards the overall agricultural development.
of small production systems. With the appropriate knowledge, small-scale producers can even have a competitive edge over larger operations. When knowledge is harnessed by strong organisations of small producers, strategic planning can be used to provide members with least-cost inputs, better storage facilities, improved transportation links and collective negotiations with buyers.

ICT can also play an important role in bringing about sustainable agricultural development when used to document both organic and traditional cultivation practices. Developing countries can create Traditional Knowledge Digital Libraries (TKDL) to collect and classify various types of local knowledge so that it can be shared more widely. These libraries could also integrate widely scattered references to Indigenous Technical Knowledge (ITK) systems in a retrievable form. Thus IT could act as a bridge between traditional and modern knowledge systems.

3.2. Areas of IT convergence

Applications of IT in support of agricultural and rural development fall into five main areas:

• economic development of agricultural producers;
• community development;
• research and education;
• small and medium enterprises development; and
• media networks.

Some agricultural development services that can be provided in the developing world, using ICT, are:

• online services for information, education and training, monitoring and consultation, diagnosis and monitoring, and transaction and processing;
• e-commerce for direct linkages between local producers, traders, retailers and suppliers;
• the facilitation of interaction among researchers, extension (knowledge) workers, and farmers;
• question-and-answer services where experts respond to queries on specialised subjects ICT services to block- and district-level developmental officials for greater efficiency in delivering services for overall agricultural development;
• up-to-date information, supplied to farmers as early as possible, about subjects such as packages of practices, market information, weather forecasting, input supplies, credit availability, etc.;
• creation of databases with details of the resources of local villages and villagers, site-specific information systems, expert systems, etc.;
• provision of early warning systems about disease/pest problems, information regarding rural development programmes and crop insurances, post-harvest technology, etc.;
• facilitation of land records and online registration services;
• improved marketing of milk and milk products;
• services providing information to farmers regarding farm business and management;
• increased efficiency and productivity of cooperative societies through the computer communication;
• network and the latest database technology;
• tele-education for farmers;
• websites established by agricultural research institutes, making the latest information available to extension (knowledge) workers and obtaining their feedback.
3.3. Drivers of ICT in Agriculture

Five main trends have been the key drivers of the use of ICT in agriculture, particularly for poor producers: (1) low-cost and pervasive connectivity, (2) adaptable and more affordable tools, (3) advances in data storage and exchange, (4) innovative business models and partnerships, and (5) the democratization of information, including the open access movement and social media. These drivers are expected to continue shaping the prospects for using ICT effectively in developing-country agriculture.

3.3.1. Low-Cost and Pervasive Connectivity

The pervasiveness of connectivity – to mobile phones, Internet, and other wireless devices – is due to a number of factors, including decreases in costs, increases in competition, and expansion of last-mile infrastructure. Several trends, working in tandem, are making ICT devices and services more affordable in ways that also extend access to small-scale producers.

The reach and affordability of broadband Internet is also improving dramatically – though somewhat slower – in developing regions.

3.3.2. Adaptable and More Affordable Tools

Mobile-based applications are also becoming more suitable for poor and isolated communities, especially though feature phones. Geospatial information is also becoming easier to access and use as mapping tools, such as Microsoft Earth or Google Maps, bring geographical data information to nonspecialist users. Scientists and development organizations have created substantial sets of georeferenced data on population, poverty, transportation, and any number of other public goods and variables through more affordable, usable geographic information systems available on standard PCs and mobile devices using web-based tools. Satellite images and similar representations have improved exponentially in quality and detail. These tools and remote sensors use less energy and require less human attention than in previous years.

Greatly increased data storage capacity and the ability to access data remotely and share it easily have improved the use of ICT in agriculture. Sharing knowledge and exchanging data have created opportunities to involve more stakeholders in agricultural research involvement facilitated by an improved e-learning environment and networking capacity.

Improvements in data storage and sharing have underlying causes. The capacity of hard drives and the speed of micro-processors have continued to rise, making it dramatically cheaper to store data. Cloud computing offers access to numerous shared computing resources through the Internet, including sharable tools, applications, and intelligently unlinked content and data.

3.3.3. New Business Models and Public-Private Partnerships

The development and use of many ICTs originated in the public sector but were quickly dominated by the private sector when their profit potential became clear. The public sector maintains great interest in ICT as a means of providing better public services that affect agriculture (for instance, land registration, forest management, and agricultural extension services), as well as for connecting with citizens and managing internal affairs. Private sector involvement in some of these efforts has enhanced the access, affordability, and adaptability of ICTs for development. Unlike other development strategies, which often struggle to survive or be scaled because the public sector cannot fund them, development strategies featuring ICTs have benefited from growing private sector interest and public demand.

New forms of business incubation and knowledge brokering are also contributing to ICT in agriculture. The private sector has a keen interest in investing in firms that come out of such incubation schemes, speculating on the ability of an innovative idea to expand into a highly profitable enterprise. Incubators identify additional investors and other suitable partners, including technical experts. In many instances, they develop enterprises through which private and public providers of agricultural services collaborate to deliver products more efficiently to farmers; in developing, sharing, and capitalizing on innovations for agricultural development, they almost always use ICT and often develop new ICT tools.

3.3.4. Democratization of Information, the Open Access Movement, and Social Media
The democratization of information and science facilitated by ICTs is also contributing to agriculture and rural development more broadly. Vast quantities of information held by institutions and individuals are becoming visible, publicly accessible, and reusable through the open access movement.

The expansion of open access software also enables grass-roots community organizations to share knowledge with one another. Social media, once used purely for entertainment, has great potential to be used for knowledge sharing and collaboration even in agriculture.

3.3.5. Use Appropriate Technologies

In designing ICT interventions, it is necessary to research and understand local information and communication practices, barriers to ICT-enabled empowerment, and priority information and communication needs of end users. Using conventional information and communication tools to address the needs of those who cannot access the ICT because of limitations related to literacy, isolation, and social norms is often required.

4. Making ICT Infrastructure, Appliances, and Services More

4.1. Accessible and Affordable in Rural Areas

The idea that wider access to and use of ICTs throughout a country will reduce inequalities in income and quality of life between rural and urban residents is compelling. Creating affordable ICT services in rural areas is a complex challenge. In these areas, the “last mile” of telecommunications infrastructure is provided at a very high cost that may not be justified by the resulting use and effects of the telecommunications network. Affordable access to ICTs in rural areas can be frustrated at the supply as well as the demand end of the service-provision chain. To supply ICTs and related services in rural areas, the main challenge is the high level of capital and operating expenses incurred by service providers.

Recognizing the equity implications of access to ICTs, governments have adopted regulatory policies to enable the rollout of ICT infrastructure and the supply of services in rural areas, and they have addressed low rural demand by introducing locally relevant content in the form of e-government and e-agriculture services.

4.2. Key Challenges and Enablers

4.2.1. Partnerships

Considering the multilayered nature of the problem of ensuring affordable rural access to infrastructure, devices, and services, partnerships among organizations with different specialties, capacities, and profit motives appear to be a key way to improve access and affordability. Partnerships serving as critical mechanisms for improving rural ICT access can take the form of partnerships within the public sector, negotiated public-private partnerships, private agreements among stakeholders in the telecommunications sector, or informal understandings between service providers and stakeholders at the community level.

4.2.2. Regulation and Policy Challenges

Although the evolution of ICTs in developing countries has far to go, it has moved significantly forward in the past decade. The rapid expansion of mobile phone networks and market uptake of Global System for Mobile Communication (GSM) technologies following liberalization and deregulation are the most frequently cited examples of this evolution. “Fixed-mobile convergence” is the increasingly seamless connectivity among wired and wireless networks, devices, and applications, which permits users to send and receive data regardless of device and location. Convergence is the result of converting content formats (text, images, audio, video), devices for creating and communicating this content, and telecommunications infrastructure to digital standards. Device convergence allows devices to support different functionalities and different network access technologies. Service convergence means that end users are able to receive comparable services via different devices and technologies for accessing networks.

4.3. Infrastructure
The lag between the arrival of complementary infrastructure and public services and the establishment of wired ICT infrastructure in rural areas can be considerable, but the introduction of wireless, especially mobile, infrastructure is bound neither by the presence of roads nor by access to the electricity grid. Rural infrastructure development needs to be considered in light of the different opportunities offered by wired and wireless technologies and the fixed-mobile convergence occurring throughout the ICT sector. Rural infrastructure development needs to be considered in light of the different opportunities offered by wired and wireless technologies and the fixed-mobile convergence occurring throughout the ICT sector.

Telecommunications networks comprise a hierarchy of ulinks that connect users at the “edge” of a network to its “core,” also called the “backbone” (the high-capacity ulinks between switches on the network). The backhaul portion of a network consists of the intermediate ulinks between subnetworks at the users’ end and the core network.

Even though wireless is accepted as an economical option for delivering “last mile” connectivity, backhaul traffic is usually carried via fiber-optic networks because of their high capacity. Connectivity is often limited by the limited penetration of the fixed-line backhaul that supports it. The delivery of connectivity to rural areas lacking fixed-line backhaul involves balancing concerns about ICT access, connection quality, and the expenditures and delays entailed in rolling out fixed lines and supporting infrastructure. The benefits of wireless backhaul technology are worth considering in such cases.

A comparison between traditional fixed-line telephone services and voice over IP (VoIP) clearly demonstrates the difference between the two types of networks. NGNs completely separate the packet-switched transport (connectivity) layer and the service layer, enabling any available fixed-line carriage infrastructure to be used efficiently for any service.

4.3.1. Local Loop or “Last Mile” Connectivity

The delivery of network access in the “last mile” is the most costly and challenging element of rural deployments. The technology options for delivering wired local loop broadband connectivity include the rollout of xDSL, cable, and fiber to the home infrastructure. Wireless options include the rollout of mobile (2G, 3G, 4G), wireless broadband (WiMAX, Wi-Fi, WLAN), and satellite very small aperture terminal (VSAT) infrastructure. Within cell-based (mobile) wireless standards, all users connect to a single base station, and the transmission bandwidth has to be shared among all users in the cell’s coverage area.

The “digital dividend” has been widely hailed as the solution to urban-rural inequities in digital ICT access. The “digital dividend” is the reassignment of operational frequencies that become available following the switch from analog to digital television broadcasting.

4.4. Appliances

From a user’s perspective, device convergence has two main aspects. First, users can access content in different formats (audio, data, location data, pictures, maps, text) and with different dynamic properties, produced by different authors, on the same device. Second, users can take advantage of different options (radio, GSM, Wi-Fi, Bluetooth, satellite) for accessing that content.

Portable devices, including but not limited to mobile phones, are starting to allow users dual (or multiple) mode flexibility. For example, dual connectivity (Wi-Fi/GSM and Bluetooth/GSM) enables mobile phones to conduct both VoIP and standard mobile calls. Dedicated telephone devices are able to process VoIP phone calls using Session Initiation Protocol, as well as regular phone calls using analog signals. Gains in processing power allow functions with higher technology requirements to work on smaller devices (high-end smartphones and Netbook appliances). Conversely, bulkier stationary devices such as the desktop computer have evolved functionalities traditionally associated with more portable devices, such as VoIP telephony and on-demand radio and TV broadcasts.

4.5. Services

Services entail much more than access to hardware; they encompass affordable access to locally relevant rural content through connectivity providers, content creators and disseminators, information intermediaries, social facilitators, information literacy educators, and the governance channels steering the performance of these services. The service layer reflects the synergies (or lack thereof) among network infrastructure, connectivity modalities, access devices, and content. Traditionally, rural information services focused on providing
broadcasting ("push") content, such as rural radio programming, but the ubiquity of mobile devices enables the sourcing and sharing ("pull") of rural content. The presence of mobile technology as an authoring tool in rural areas presents an untapped opportunity to engage rural users in authoring content, thereby increasing the demand for existing rural infrastructure. Mobile devices, in combination with broadcasting technologies such as radio, enable rural residents to participate in public discourse and influence decision making.

4.6. Anytime, Anywhere: Mobile Devices and Services

This module describes current knowledge, innovative practices, opportunities, and challenges in using mobile phones to benefit agriculture.

Mobile phones may help to increase income, improve the efficiency of markets, reduce transaction costs, and offer a great opportunity for innovative interventions, especially in service delivery. Yet to realize the full potential of enhanced communication of market information, the use of mobiles must be coupled with additional investments (in roads, education, financial services, and so forth).

The rise of the mobile phone has been one of the most stunning changes like other technologies before it, the mobile phone is likely to be the subject of inflated expectations and hopes. The newest smartphones are far more sophisticated than the more affordable models populating poor regions, but those simple phones are still leaps and bounds ahead of devices that were cutting edge a decade ago and they are entirely relevant to agriculture.

In many countries, agriculture accounts for the overwhelming majority of rural employment. The manifold benefits that accompany improvements in agricultural productivity are well known: Farmers’ incomes rise, food prices fall, and labor is freed for additional employment. In some instances productivity improvements have proven elusive, as climate change and uncertain commodity prices have worsened agrarian conditions for many rural communities.

Technical innovation, most prominently demonstrated in the Green Revolution, has been key to improving agricultural markets in the developing world. Mobile phones, despite their recent entry into agrarian communities, are already helping those communities improve their agricultural activities.

Advances throughout the mobile phone ecosystem tend to act as a positive feedback loop. This "virtuous circle" of innovation enables a number of benefits, even for smallholder farmers:

• Access. Mobile wireless networks are expanding as technical and financial innovations widen coverage to more areas.

• Affordability. Prepaid connectivity and inexpensive devices, often available second hand, make mobile phones far cheaper than alternatives.

• Appliances. Mobile phones are constantly increasing in sophistication and ease of use. Innovations arrive through traditional trickle-down effects from expensive models but have also been directed at the bottom of the pyramid.

• Applications. Applications and services using mobile phones range from simple text messaging services to increasingly advanced software applications that provide both livelihood improvements and real-time public services.

• The proliferation of mobile phones across the globe has impinged on agriculture in various ways. Mobiles are being used to help raise farmers’ incomes, making agricultural marketing more efficient, lowering information costs, reducing transport costs, and providing a platform to deliver services and innovate. Whether the potential of these trends can be realized more widely, especially in rural areas and in an equitable way, is uncertain. Every aspect of the technology is changing rapidly; the public sector, private sector, and private citizens are constantly experimenting with new applications for it; and governments are grappling with any number of strategies to ease the digital divide.

5. Applications in Agriculture

New automation, ICT and GIS technologies provide solutions for steering and controlling mobile working units in site-specific production systems as a way to fulfil requirements for a safe, efficient, environment friendly and traceable production. However, enhanced quality and efficient performance of work tasks require the organizing
of a user-centric on-line support which is based on open system solutions. A missing unink in the envisioned system is the lack of a refined and integrated analysis of the acquired data and the transformation of these data into information and knowledge useful for decision making. Currently, major parts of the information collected by sensors or by manual registrations are not used, due to data logistic problems. Costs of the time spent for handling and managing the data, in many cases, outweigh the economical benefits using the data. Thus, for example, the use of wireless communication is very much in the demand in the future. The required user-friendly system does not exist currently, and companies are not able to develop these kinds of open systems alone.

The biggest challenge for farmers in the future will be to effectively manage information on and off their farms to improve economic viability and reduce environmental impact. Farms of tomorrow have to be able to meet environmental and societal standards through advanced technologies and ICT tools. This will allow the integration of public goods, provided by agriculture. The FutureFarm project will provide and demonstrate the scientific and technical prerequisites for easy and reliable information management on future farms. Due to the high variability of farming regions, farm types, crops and technological adoption, the project will also deliver typologies of potentials in the development of future farms, of feasible technical development threads and of policies within the EU-member states, relevant to information management and to introduction or promotion of innovations on farms.

Precision Farming research and development over many years has adapted and developed information and communication technologies for farming systems. With its sensors, satellite positioning systems, PC-based geographical information systems and automated control it is the technical core of an information-driven or ‘information intensive’ crop production. We have the ability to apply these techniques on a cheap and simple basis in order to meet the internal and external requirements of management and information for all kinds of farms and farming systems. Currently most of the required sub-systems and technologies exist as separate entities. FutureFarm will analyze the necessary prerequisites and design concepts to set up the required technical environments, information platforms and procedures for core activities of future farmers in the European Union. Standardized information flow is a technical prerequisite to comply with private or public standards and thus to communicate indirectly but in a new mode with consumers and citizens. This will also allow better integration of local specific information and thus will account for the diversity of European agriculture.

Using case studies and practical examples this project will integrate existing research into a coherent information management system (Farming Management Information System, FMIS). This system will allow not only supporting socioeconomic viability with environmental considerations but also go towards meeting evolving standards for compliance. The general validity of the system will be tested on four farms in the EU using examples of cash crop production as well as internal use of ‘bio-energy’ produced on the farm. Studying the potentials to use field robots in order to increase labor- and energy-efficiency, or to achieve controlled precision in complying with environmental demands in crop production, rounds off the project. The applicability and the necessary steps to adapt the system to the variability of farm types in the EU will be analyzed and typified. Recommendations for the development of future farms from the central view of information management will be elaborated with stakeholders from farms and regions as well as with the European companies providing information technology for agriculture. Besides researchers from universities and non-university research centers some project partners are European SMEs. Four selected farms are participating in the project to ensure R+D and communication within the project with a view for the practical conditions of farmers in the EU and to have access to real world situations and data.

Developing codes of good farming practice, diversifying markets and production systems as well as European standards of sustainable agricultural production systems require implementation of more elaborate management strategies. These have to respect specific ecological conditions, demands from the rural regions and those from the value-added chains. On top of that, these strategies have to be simple, but flexible enough to be adapted easily to changing economic or environmental conditions and they need proof of their compliance. Beyond that, the demand for information about the production processes is growing, both from the perspective of the value-added chains (traceability) as well as from regional stakeholders in order to fulfil multifunctional objectives by farming. An important prerequisite for farmers to comply with all these different demands is to easily have sufficient and timely information available for decision making or providing documentary evidence. The rapid development of technologies for information and communication, new sensors as well as the vast potentials for providing geo-referenced data (remote-sensing, on-line sensors, public databases etc.) also allows farmers to access new and high quality data and use them as specific information in decision making or process documentation. With automated data acquisition and handling in a farm management information system the farmers can comply with a rapid growing demand of standards in the management for further tracing the production processes.
Precision Farming (PF) in Europe uses new technologies in information handling and management as well as in managing the spatial and temporal variability found on all farms. Such explicit information use improves economic returns and reduces environmental impact. Precision farming is very data intensive and historically ulinked with site specific activities and management on the field. It has become very clear in recent years that PF is not limited to site-specific farming. The use of techniques and methods that form precision farming can provide a wealth of information and tools to handle and apply information properly for any type of farm in any region. This information-driven approach can be used to help improve crop management strategies and proof of compliance through documentation.

The introduction of advanced ICT technologies into agriculture will also be a significant progress in all efforts for measurements oriented payments within agro-environmental programs and related efforts to enforce environmentally sound systems in land use within the EU. This includes also the Best Management Practice according to the cross compliance scheme.

6. Questions

1. What type of organizations play role in agricultural material and information relationship?
2. What is the role of Collaborative Working Environment in agriculture?
3. What are the main working phases in farms?
4. Who needs ICT for agriculture development?
5. What are the building bloks of precision agriculture?
6. What is the ambient intelligent?
2. fejezet - Hardware Architectures

1. Introduction

A computer is an electronic device, operating under the control of instructions stored in its own memory, that can accept data, process the data according to specified rules, produce results, and stores the results for future use.

1.1. Data and Information

Computers process data into information. Data is a collection of unprocessed items, which can include text, numbers, images, audio, and video. Information conveys meaning and is useful to people. Many daily activities either involve the use of or depend on information from a computer.

1.2. Information Processing Cycle

Computers process data (input) into information (output). Computers carry out processes using instructions, which are the steps that tell the computer how to perform a particular task. A collection of related instructions organized for a common purpose is referred to as software. A computer often holds data, information, and instructions in storage for future use. Some people refer to the series of input, process, output, and storage activities as the information processing cycle.

Most computers today communicate with other computers. As a result, communications also has become an essential element of the information processing cycle.

2. The Components of a computer

A computer contains many electric, electronic, and mechanical components known as hardware. These components include input devices, output devices, a system unit, storage devices, and communications devices (Figure 2.1).

2.1. ábra - Figure 2.1: CPU Test Program

2.1. Input Devices
An input device is any hardware component that allows you to enter data and instructions into a computer. Five widely used input devices are the keyboard, mouse, microphone, scanner, and Web cam. A computer keyboard contains keys you press to enter data into the computer. For security purposes, some keyboards include a fingerprint reader, which allows you to work with the computer only if your fingerprint is recognized. Mouse, you control movement of a small symbol on the screen, called the pointer, and you make selections from the screen. A microphone allows you to speak into the computer. A scanner converts printed material (such as text and pictures) into a form the computer can use.

### 2.2. Output Devices

An output device is any hardware component that conveys information to one or more people. Three commonly used output devices are a printer, a monitor, and speakers.

A printer produces text and graphics on a physical medium such as paper. A monitor displays text, graphics, and videos on a screen. Speakers allow you to hear music, voice, and other audio (sounds).

### 2.3. System Unit

The system unit is a case that contains the electronic components of the computer that are used to process data. The circuitry of the system unit usually is part of or is connected to a circuit board called the motherboard.

Two main components on the motherboard are the processor and memory. The processor, also called a CPU (central processing unit), is the electronic component that interprets and carries out the basic instructions that operate the computer. Memory consists of electronic components that store instructions waiting to be executed and data needed by those instructions (Figure 2.2). Although some forms of memory are permanent, most memory keeps data and instructions temporarily, which means its contents are erased when the computer is shut off.

2.2. ábra - Figure 2.2: The Main Memory

![Figure 2.2: The Main Memory](image)

### 2.4. Storage Devices

Storage holds data, instructions, and information for future use. For example, computers can store hundreds or millions of customer names and addresses. Storage holds these items permanently.

A computer keeps data, instructions, and information on storage media. Examples of storage media are USB flash drives, hard disks, optical discs, and memory cards. A storage device records (writes) and/or retrieves (reads) items to and from storage media. Drives and readers/writers, which are types of storage devices, accept a specific kind of storage media. For example, a DVD drive (storage device) accepts a DVD (storage media). Storage devices often function as a source of input because they transfer items from storage to memory.

A USB flash drive is a portable storage device that is small and lightweight enough to be transported on a keychain or in a pocket. The average USB flash drive can hold about 4 billion characters. You plug a USB flash drive in a special, easily accessible opening on the computer.
A hard disk provides much greater storage capacity than a USB flash drive. The average hard disk can hold more than 320 billion characters. Hard disks are enclosed in an airtight, sealed case. Although some are portable, most are housed inside the system unit. Portable hard disks are either external or removable. An external hard disk is a separate, freestanding unit, whereas you insert and remove a removable hard disk from the computer or a device connected to the computer.

An optical disc is a flat, round, portable metal disc with a plastic coating. CDs, DVDs, and Blu-ray Discs are three types of optical discs. A CD can hold from 650 million to 1 billion characters. Some DVDs can store two full-length movies or 17 billion characters.

Blu-ray Discs can store about 46 hours of standard video, or 100 billion characters.

Some mobile devices, such as digital cameras, use memory cards as the storage media. You can use a card reader/writer to transfer the stored items, such as digital photos, from the memory card to a computer or printer.

2.5. Communications Devices

A communications device is a hardware component that enables a computer to send (transmit) and receive data, instructions, and information to and from one or more computers or mobile devices. A widely used communications device is a modem.

Communications occur over cables, telephone lines, cellular radio networks, satellites, and other transmission media. Some transmission media, such as satellites and cellular radio networks, are wireless, which means they have no physical lines or wires.

3. The Computer Architecture

The modern microcomputer has roots going back to USA in the 1940’s. Of the many researchers, the Hungarian-born mathematician, John von Neumann (1903-57), is worthy of special mention. He developed a very basic model for computers which we are still using today (Figure 2.3).

2.3. ábra - Figure 2.3: The Neumann Model

Von Neumann divided a computer’s hardware into 5 primary groups:

- CPU
- Input
- Output
- Working storage
- Permanent storage
This division provided the actual foundation for the modern PC, as von Neumann was the first person to construct a computer which had working storage (what we today call RAM), and the amazing thing is, his model is still completely applicable today.

Today we talk about multimedia PC’s, which are made up of a wealth of interesting components. Note here that modems, sound cards and video cards, etc. all function as both input and output units. But this doesn't happen simultaneously, as the model might lead you to believe. At the basic level, the von Neumann model still applies today (Figure 2.4).

![Figure 2.4: The main parts of PC](image)

### 3.1. Basic Terms

I’m soon going to start throwing words around like: interface, controller and protocol. These aren’t arbitrary words. In order to understand the transport of data inside the PC we need to agree on various jargon terms. I have explained a handful of them below.

**Binary data**: data, be it instructions, user data or something else, which has been translated into sequences of 0’s and 1’s.

**Bus width**: the size of the packet of data which is processed (e.g. moved) in each work cycle. This can be 8, 16, 32, 64, 128 or 256 bits.

**Band width**: the data transfer capacity. This is measured in, for example, kilobits/second (Kbps) or megabytes/second (MBps).

**Cache**: a temporary storage, a buffer.

**Chipset**: a collection of one or more controllers. Many of the motherboard’s controllers are gathered together into a chipset, which is normally made up of a north bridge and a south bridge.

**Controller**: a circuit which controls one or more hardware components. The controller is often part of the interface.

**Hubs**: this expression is often used in relation to chipset design, where the two north and south bridge controllers are called hubs in modern design.

**Interface**: a system which can transfer data from one component (or subsystem) to another. An interface connects two components (e.g. a hard disk and a motherboard). Interfaces are responsible for the exchange of data between two components. At the physical level they consist of both software and hardware elements.

**I/O units**: components like mice, keyboards, serial and parallel ports, screens, network and other cards, along with USB, firewire and SCSI controllers, etc.
Clock frequency: the rate at which data is transferred, which varies quite a lot between the various components of the PC. Usually measured in MH.

Clock tick (or clock cycle): a single clock tick is the smallest measure in the working cycle. A working cycle (e.g., the transport of a portion of data) can be executed over a period of about 5 clock ticks (it "costs" 5 clock cycles).

Logic: an expression I use to refer to software built into chips and controllers. E.g., an EIDE controller has its own "logic", and the motherboard’s BIOS is "logic".

MHz (Megahertz): a "speed" which is used to indicate clock frequency. It really means: million cycles per second. The more MHZ, the more data operations can be performed per second.

North bridge: a chip in the motherboard which serves as a controller for the data traffic close to the CPU. It interfaces with the CPU through the Front Side Bus (FSB) and with the memory through the memory bus.

ProtocolsElectronic: traffic rules which regulate the flow of data between two components or systems. Protocols form part of interfaces.

South bridge: a chip in the motherboard which works together with the north bridge. It looks after the data traffic which is remote from the CPU (I/O traffic).

3.2. A data processor

The PC is a digital data processor. In practise this means that all analogue data (text, sound, pictures) gets translated into masses of 0’s and 1’s. These numbers (binary values) exist as tiny electrical charges in microscopic circuits, where a transistor can take on two states: charged or not charged. This is one picture of a bit, which you can say is either turned on or off.

There can be billions of these microscopic bits hidden inside a PC, and they are all managed using electronic circuits (EDP stands for electronic data processing). For example, the letter "A" (like all other characters) can be represented by a particular 8-digit bit pattern. For "A", this 8-digit bit pattern is 01000001.

When you type an "A" on your keyboard, you create the digital data sequence, 01000001. To put it simply, the "A" exists as a pattern in eight transistors, where some are "turned on" (charged) and others are not. Together these 8 transistors make up one byte.

The same set of data can be stored in the video card’s electronics, in RAM or even as a magnetic pattern on your hard disk. The set of data can also be transferred to a printer, if you want to print out your text. The printer electronically and mechanically translates the individual bits into analogue letters and numbers which are printed on paper. In this way, there are billions of bytes constantly circulating in your PC, while ever it is switched on. But how are these 0’s and 1’s moved around, and which components are responsible?

4. The physical PC

The PC is made up of a central unit (also called a system unit) and some external devices. The central unit is a box (a cabinet), which contains most of the computer’s electronics (the internal devices). The external devices are connected to the central unit (shown below) using cables.

4.1. Internal drivers

Motherboard: CPU, RAM, cache, ROM circuits containing the BIOS and startup programs. Chipsets (controllers). Ports, busses and slots. EIDE interface, USB, AGP, etc.

Drives: Hard disk(s), diskette drive, CD-ROM, DVD, etc.

Plug-in cards: Graphics card (video adapter), network card, SCSI controller. Sound card, video and TV card. Modem and ISDN card.

4.2. External drivers
Keyboard, mouse, joystick, screen, printer, scanner, speakers, external drives, tape drive, MIDI, units modem, digital camera.

The PC processes data. It performs calculations and moves data between the various components. It all happens at our command, and we want on the happen fast. The PC can be viewed as a series of more or less independent subsystems, which can each be developed to permit greater capacity and higher speed. We constantly need new standards, because of the new, faster, interfaces, busses, protocols (which we all work out together), delivering better performance.

The PC is the sum of all these subsystems. At each boundary between one subsystem and another, we find an interface. That is, an electrical system which connects the two subsystems together and enables them to exchange data.

The concept of an interface is a little abstract, as it most accurately refers to a standard (a set of rules for the exchange of data). In practice, an interface can consist of, for example, two controllers (one at each end of the connection), a cable, and some software (protocols, etc.) contained the controllers.

The controllers are small electronic circuits which control the movement of data to and from the device.

An interface connects two hardware devices. An interface can consist of controllers with built-in software, cables, etc. There are many interfaces in the PC, because there are many subsystems which have to be connected. Each interface is normally tailor-made for the job, and tuned to achieve maximum bandwidth (data transfer capacity) between the two components.

4.3. An example of an interface

Later on the guide I want to explore the EIDE interface in more detail, but I will use it here as a specific example of an interface. Keep your attention focused on the concept of an interface.

If we want to connect a hard disk (Figure 2.5) to a motherboard, this is achieved using an EIDE interface. If we look more closely at this interface, it can be divided into a series of subcomponents. The interface consists of both hardware and logic: the most important being the two EIDE controllers. One is integrated into the hard disk’s electronics, and the other is integrated into the motherboard, where it forms part of the chipset’s south bridge.

2.5. ábra - Figure 2.5: The Winchester

Underneath the hard disk you can see a small printed circuit board. This incorporates the controller functions which work together with the corresponding controller in the PC’s motherboard.

The advantage of this system is that the hard disk can be connected directly to the motherboard with a cable. But the cable still runs from one controller to the other.

The two controllers work according to a common standard, which is the ATA standard. This standard includes a set of protocols which are continually being developed in new versions. Let’s say our specific hard disk can use the ATA/100 protocol. That means the controller on the motherboard has to also be compatible with ATA/100, and the cable as well. When all that is in place, we have a working ATA interface (Figure 2.6).

2.6. ábra - Figure 2.6: A specific example of an interface
All traffic originates from or ends up in the motherboard; which is appropriately called the most important component of the PC (Figure 2.7).

2.7. ábra - Figure 2.7: The motherboard is the hub of all data exchange

I will show you pictures of the individual components of the motherboard later, but this is what it looks like as a total unit (Figure 2.8).

2.8. ábra - Figure 2.8: A motherboard is a board covered with electronics
5. The CPU

CPU stands for Central Processing Unit. There can be several processors in a computer, but one of them is the central one – the CPU.

The reason the CPU is called a processor is because it can work with data. And it has two important jobs:

• It can do calculations.
• It can move data.

The CPU is very fast at doing both jobs. The faster the CPU can do calculations and move data, the faster we say the PC is. What follows is a short description of how to achieve faster data processing. Read it, and see if you understand all the concepts. There are three ways to improve a PC’s performance:

• Higher clock frequencies (which means more clock ticks per second).
• Greater bus width.
• Optimising the core of the processor and other components so that the maximum amount of work is done for each clock tick.

All this can lead to better bandwidth, which is required throughout the PC. The entire development process is focused around the motherboard, and especially the CPU. But all of the electronics has to be able to keep up with the high pace, and that is what makes the motherboard so fascinating.

The CPU is physically quite small. At its core is an electronic circuit (called a die), which is no bigger than your little fingernail.

Despite its small size, the CPU is full of transistors. The die in a Pentium 4 CPU contains 125 million transistors, all squashed together into a very tight space. It is about 1 cm x 1 cm in size.

5.1. CPU testing programs

Finally, let me just mention some small utility programs which you can download from the Internet (e.g. search for “WCPUID” or “CPU-Z” on www.google.com, and you’ll find it). The programs WCPUID and CPU-Z, reveals lots of information about your CPU, chipset, etc. They are used by motherboard nerds. Here CPU-Z reports that the Pentium 4 processor is a ”Prescott” model (Figure 2.9).

2.9. ábra - Figure 2.9: CPU Test Program
5.2. The CPU and the motherboard

The heart and soul of the PC’s data processing is the CPU. But the processor is not alone in the world, it communicates with the rest of the motherboard. There will be many new terms introduced in the following sections, so remember that you can find definitions for all the abbreviations in the back of the guide.

6. Busses

Data packets (of 8, 16, 32, 64 or more bits at a time) are constantly being moved back and forth between the CPU and all the other components (RAM, hard disk, etc.). These transfers are all done using busses. The motherboard is designed around some very powerful data channels (or pathways, as they are also called). It is these busses which connect all the components to each other (Figure 2.10).

2.10. ábra - Figure 2.10: The busses

The busses are the data channels which connect the PC’s components together. Some are designed for small transfers, others for large ones. There is not just one bus on a motherboard; there are several. But they are all connected, so that data can run from one to another, and hence reach the farthest corners of the motherboard.

We can say that a bus system is subdivided into several branches. Some of the PC components work with enormous amounts of data, while others manage with much less. For example, the keyboard only sends very few bytes per second, whereas the working storage (RAM) can send and receive several gigabytes per second. So you can’t attach RAM and the keyboard to the same bus.
Two busses with different capacities (bandwidths) can be connected if we place a controller between them. Such a controller is often called a bridge, since it functions as a bridge between the two different traffic systems (Figure 2.11).

**2.11. ábra - Figure 2.11: Bridges connect the various busses together**

The entire bus system starts close to the CPU, where the load (traffic) is greatest. From here, the busses work outwards towards the other components. Closest to the CPU we find the working storage. RAM is the component which has the very greatest data traffic, and is therefore connected directly to the CPU by a particularly powerful bus. It is called the front side bus (FSB) (Figure 2.12) or (in older systems) the system bus.

**2.12. ábra - Figure 2.12: The Front Side Bus**

The PC’s most important bus looks after the “heavy” traffic between the CPU and RAM. The busses connecting the motherboard to the PC’s peripheral devices are called I/O busses. They are managed by the controllers.

**7. The chip set**

The motherboard’s busses are regulated by a number of controllers. These are small circuits which have been designed to look after a particular job, like moving data to and from EIDE devices (hard disks, etc.).

A number of controllers are needed on a motherboard, as there are many different types of hardware devices which all need to be able to communicate with each other. Most of these controller functions are grouped together into a couple of large chips, which together comprise the chip set (Figure 2.13).

**2.13. ábra - Figure 2.13: Chip Sets**

The two chips which make up the chipset, and which connect the motherboard’s busses.

The most widespread chipset architecture consists of two chips, usually called the north and south bridges. This division applies to the most popular chipsets from VIA and Intel. The north bridge and south bridge are connected by a powerful bus, which sometimes is called a ulink channel (Figure 2.14).
2.14. ábra - Figure 2.14: ulink between Bridges

The north bridge and south bridge share the work of managing the data traffic on the motherboard.

### 7.1. The north bridge

The north bridge is a controller which controls the flow of data between the CPU and RAM, and to the AGP port. The north bridge has a large heat sink attached to it (Figure 2.15). It gets hot because of the often very large amounts of data traffic which pass through it. All around the north bridge you can see the devices it connects.

2.15. ábra - Figure 2.15: The North Bridge

The north bridge and its immediate surroundings. A lot of traffic runs through the north bridge, hence the heat sink.

The AGP is actually an I/O port. It is used for the video card. In contrast to the other I/O devices, the AGP port is connected directly to the north bridge, because it has to be as close to the RAM as possible. The same goes for the PCI Express x16 port, which is the replacement of AGP in new motherboards.

### 7.2. The south bridge

The south bridge incorporates a number of different controller functions (Figure 2.16). It looks after the transfer of data to and from the hard disk and all the other I/O devices, and passes this data into the ulink channel which connects to the north bridge. You can clearly see that the south bridge is physically located close to the PCI slots, which are used for I/O devices.

2.16. ábra - Figure 2.16: The South Bridge

### 8. SCSI, USB and Firewire
In this chapter I need to discuss three I/O buses. They are very different, but they still belong together:

- **SCSI** is an older but advanced I/O bus which has especially been used for hard disks, CD-ROM drives, scanners and tape units.

- **USB** is a modern bus which can be used for a host of devices, and which has had a powerful breakthrough on the PC front in recent years.

- **FireWire** is a modern, high-speed I/O bus which is especially used for digital video cameras (DV), scanners and external hard disks.

As I mentioned, the three buses are very different, but they also overlap with each other. For example, you can buy CD burners with all three types of interface.

### 8.1. SCSI

SCSI (Small Computer System Interface) is an advanced controller technology, which is especially used in high-end PC’s. These can be network servers or just powerful workstations, for which there are a number of different SCSI standards. The SCSI bus can transfer up to 160 MB/second, which is more than the PCI bus can deliver.

A SCSI system is built around a central controller, called the host adapter, which is almost a tiny computer in its own right. The adapter can be quite expensive if, for example, it has to be used with very fast hard disks. However there are simpler SCSI adapters, for example, those sold with SCSI based scanners.

The best known manufacturer of SCSI adapters is Adaptec. It used to also be quite common for motherboards to have built-in SCSI controller often of high quality.

A host adapter can control a number of SCSI devices, which are connected in a long series (a chain). Every device is allocated an identification number, and a terminator has to be put in at both ends of the SCSI chain. This is done, for example, using a jumper on one of the devices.

### 8.2. RAID

RAID stands for Redundant Array of Inexpensive Disks. It is a disk technology that connects together a serious of standard hard disks to form an advanced, error correcting system, which is used in servers.

The system is virtually an extension of the SCSI standard, and was first used in 1987. Since then, ATA-based RAID systems have been developed which use the much cheaper ATA or SATA disks in an equivalent configuration (see the discussion later in the guide). The trick is, that you can spread your data over several disks. With a RAID chain of hard disks, you gain two advantages:

- Greater security. The data is on several disks. If one disk goes down, the other disks contain the same data.

- Faster data transfer. The RAID controller writes and reads from several disks at the same time. This means the transfer speed can be doubled or tripled using RAID. When the user has read or written his file, the controller finishes the job itself, so that the complete file is located on all the attached disks.

There are several RAID categories. A RAID controller has to be used, which is a special SCSI adapter.

### 8.3. USB

USB stands for Universal Serial Bus and is an I/O standard originally developed by seven companies – Compaq, Digital, IBM, Intel, Microsoft, NEC and Northern Telecom (see the USB Implementers Forum at www.usb.org).

USB is a cheap serial I/O bus with an open specification. This means that anyone can produce USB products, without having to pay licences to anyone.

USB has been the biggest and most welcome innovation in PC design seen for many years. It is an expansion bus which allows a vast amount of PC equipment to be connected. It’s suddenly possible to connect loads of...
different gadgets to the PC and using just one type of connector! And USB devices can also be used both with Macintosh computers and PC’s yet another advantage.

USB has been advertised since 1994, and for many years it was called the Useless Serial Bus. But starting in 1999, production finally surged forward, and there are now thousands of different USB gadgets.

USB unifies all the different connections for keyboard, mouse, scanner, joystick, digital camera, and perhaps printer, onto a shared bus connected using a common connector type.

8.4. USB – the technology

From a technical viewpoint, the following can be said about USB:

• The transfer speed is limited to a maximum of 12 Mbit/sec. in USB version 1.1. It is therefore primarily used for equipment which doesn’t require a large bandwidth.

• USB version 2.0 has a bandwidth of 40 MB pr. second, and is used in all modern computers. USB version 2.0 is backwards compatible. The same type of connector is used, and old devices can be connected to the new controllers.

• USB is a serial connection using just four conductors (in contrast to the 50 or so used for a PCI device). This makes manufacturing much easier and cheaper.

• The USB cable can also supply power to the devices. This means that scanners, for example, don’t have to have their own power supply. The maximum cable length is 5 meters.

• Up to 127 USB devices can be connected to the PC using USB hubs.

• There are no IRQ’s to be concerned with. USB devices can be connected “On the fly”, without restarting the PC.

There has to be a USB host controller in the PC in order to be able to connect the devices. This controller can be bought separately, as an adapter, but most motherboards have one built into the chipset’s south bridge. There are typically two or four USB connectors on the motherboard, but you can have many more USB devices than this if you connect an extra hub (e.g. integrated in a screen).

8.5. Hard disks, ATA and SATA

I am now finally going to describe the ATA interface, which has been mentioned several times earlier in the guide. The ATA interface is used for hard disks. We have also seen, that ATA devices use bus mastering to exchange data directly with RAM. But what does this interface actually consist of? And why do new standards for hard disks keep coming out? We are going to look at that now.

9. Questions

1. What are the main computer hardware components?

2. What is the configuration?

3. What are the main functions of the main memory?

4. Describe the input/output devices!

5. What are the main functions of the processor?

6. Describe the Neuman architecture!
3. fejezet - Operating Systems

1. Introduction

An operating system controls the way in which the computer system functions. In order to do this, the operating system includes programs that:

- initialize the hardware of the computer system,
- provide basic routines for device control,
- provide for the management, scheduling and interaction of tasks,
- maintain system integrity and handle errors.

There are many types of operating systems, the complexity of which varies depending upon what type of functions are provided, and what the system is being used for. Some systems are responsible for managing many users on a network. Other operating systems do not manage user programs at all. These are typically found in hardware devices like petrol pumps, airplanes, video recorders, washing machines and car engines (Figure 3.1).

3.1. ábra - Figure 3.1: Systems Today

Windows NT Workstation is known as a general-purpose operating system. This is because it provides the ability to run a number of different programs, such as games, word processing, business applications and program development tools.

An operating system for a security control system (such as a home alarm system) would consist of a number of programs. One of these programs would gain control of the computer system when it is powered on, and initialize the system.

The first task of this initialize program would be to reset (and probably test) the hardware sensors and alarms. Once the hardware initialization was complete, the operating system would enter a continual monitoring routine of all the input sensors. If the state of any input sensor changed, it would branch to an alarm generation routine.

1.1. Input and Output devices

Input and output devices are components that form part of the computer system. These devices are controlled by the operating system. Input devices provide input signals such as commands to the operating system. These commands received from input devices instruct the operating system to perform some task or control its behavior. Typical input devices are a keyboard, mouse, temperature sensor, air-flow valve or door switch. In the previous example of our simple security control system, the input devices could be door switches, alarm keypad
panel and smoke detector units. Output devices are instruments that receive commands or information from the operating system. Typical output devices are monitor screens, printers, speakers, alarm bells, fans, pumps, control valves, light bulbs and sirens.

### 1.2. A single-user operating system

We are all familiar with the concept of sitting down at a computer system and writing documents or performing some task such as writing a letter. In this instance there is one keyboard and one monitor that you interact with. Operating systems such as Windows 95, Windows NT Workstation and Windows 2000 professional are essentially single user operating systems. They provide you the capability to perform tasks on the computer system such as writing programs and documents, printing and accessing files.

Consider a typical home computer. There is a single keyboard and mouse that accept input commands, and a single monitor to display information output. There may also be a printer for the printing of documents and images.

In essence, a single-user operating system provides access to the computer system by a single user at a time. If another user needs access to the computer system, they must wait till the current user finishes what they are doing and leaves.

Students in computer labs at colleges or University often experience this. You might also have experienced this at home, where you want to use the computer but someone else is currently using it. You have to wait for them to finish before you can use the computer system.

### 1.3. A multi-user operating system

A multi-user operating system lets more than one user access the computer system at one time. Access to the computer system is normally provided via a network, so that users access the computer remotely using a terminal or other computer.

In the early days of large multi-user computers, multiple terminals (keyboards and associated monitors) were provided. These terminals sent their commands to the main multi-user computer for processing, and the results were then displayed on the associated terminal monitor screen. Terminals were hard-wired directly to the multi-user computer system.

Today, these terminals are generally personal computers and use a network to send and receive information to the multi-user computer system. Examples of multi-user operating systems are UNIX, Linux (a UNIX clone) and mainframes such as the IBM AS400.

The operating system for a large multi-user computer system with many terminals is much more complex than a single-user operating system. It must manage and run all user requests, ensuring they do not interfere with each other. Devices that are serial in nature (devices which can only be used by one user at a time, like printers and disks) must be shared amongst all those requesting them (so that all the output documents are not jumbled up). If each user tried to send their document to the printer at the same time, the end result would be garbage. Instead, documents are sent to a queue, and each document is printed in its entirety before the next document to be printed is retrieved from the queue. When you wait inline at the cafeteria to be served you are in a queue. Imagine that all the people in the queue are documents waiting to be printed and the cashier at the end of the queue is the printer.

### 1.4. Operating system utilities

The operating system consists of hundreds of thousands of lines of program code and stored on hard disk. Portions of the operating system are loaded into computer system memory (RAM) when needed. Utilities are provided for:

- Managing Files and Documents;
- Development of Programs and Software;
- Communicating between people and with other computer systems;
Operating Systems

- Managing user requirements for programs, storage space and priority.

1.5. Operating System Interfaces

In addition, the operating system provides each user with an interface that accepts, interprets and executes user commands or programs. This interface is commonly called a SHELL or command line interpreter (CLI). In some systems this might be a simple text mode line-by-line entry using keywords (such as MSDOS or UNIX), in other systems it might be highly graphical using windows and a pointing device such as a mouse (such as X-Windows).

1.6. Advantages and problems of multi-user operating systems

The advantage of having a multi-user operating system is that normally the hardware is very expensive, and it lets a number of users share this expensive resource. This means the cost is divided amongst the users. It also makes better use of the resources. Since the resources are shared, they are more likely to be in use than sitting idle being unproductive.

One problem with multi-user computer systems is that as more users access it, the performance becomes slower and slower. Another disadvantage is the cost of hardware, as a multi-user operating system requires a lot of disk space and memory. In addition, the actual software for multi-user operating systems tend to cost more than single-user operating systems.

1.7. Multi-tasking operating system

A multi-tasking operating system provides the ability to run more than one program at once. For example, a user could be running a word processing package, printing a document, copying files to the floppy disk and backing up selected files to a tape unit. Each of these tasks the user is doing appears to be running at the same time.

A multi-tasking operating system has the advantage of letting the user run more than one task at once, so this leads to increased productivity. The disadvantage is that more programs that are run by the user, the more memory that is required.

2. Parts of an Operating System

In this section we look at that part of the operating system that is responsible for running programs, called the real-time executive or kernel.

An operating system for a large-scale computer that is used by many people at once is a very complex system. It contains many millions of lines of instructions (commands that the computer executes) written by programmers. To make operating systems easier to write, they are constructed as a series of modules (programs), each module responsible for one function (Figure 3.2).

3.2. ábra - Figure 3.2: Main modules of Operating System
• Typical modules in a larger multi-user operating system could be:
  
  • Kernel (also known as the executive);
  
  • Process manager;
  
  • Scheduler;
  
  • File manager.

2.1. The real-time executive

The core of all operating systems is called a REAL TIME EXECUTIVE (also known as the kernel). Some of the functions that it performs are:

• switching between programs,

• hardware device control and programming,

• memory management,

• process management,

• scheduling (deciding what programs to run),

• inter-process communication,

• processing of exceptions and interrupts.

Our simple security monitoring system would not have all of the above, as it would probably be a single task system, running only one program. As such, it would not need to perform scheduling of more than one program or allow communication to take place between programs (called inter-process communication). Memory management would be unnecessary, as the program would easily fit into the available memory of the computer.

An operating system designed to handle a large number of people would need a real-time executive that performs all of the above. User programs are generally stored on disk, thus need to be loaded into memory before being executed. This presents the need for memory management, as the memory of the computer would need to be searched for a free area in which to load a persons program into. When the user was finished running
the program, the memory consumed by it would need to be freed up and made available for another user when
required.

Process scheduling and management is also necessary, so that all programs would be executed and run fairly.
There is no point if a program by a specific user runs to such an extent that it denies the running of any other
programs, making every other user wait. In addition, some programs might need to be executed more frequently
than others, for example, checking network communications or printing. Some programs may need to be
temporarily halted, then restarted again later, so this introduces the need for inter-program communication.

2.2. Computer program

Programs are a series of instructions to the computer. When a software programmer (a person who writes
programs to run on a computer system) develops a program, it is converted into a long list of instructions that is
executed by the computer system.

In operating systems we talk more of a process (part of a program that is in some stage of execution) than a
program. This is because in modern operating systems, only a portion of a program is loaded at any one time.
The rest of the program sits waiting on a disk unit till it is needed. This saves memory space.

Processors execute computer programs. A processor is a chip in the computer that executes program
instructions. Processors execute millions of instructions per second.

2.3. Run more than one program at once

Some systems run only a single process at a time, other systems run multiple processes at once. Most computer
systems are single processor based, and a processor can only execute one instruction at a time, so how is it
possible for such a single processor system run multiple processes? The simple answer is that it doesn’t. The
processor of the computer runs one process for a short period of time, then is switched to the next process and so
on. As the processor executes millions of instructions per second, this gives the appearance of many processes
running at once.

2.4. Co-operative and preemptive switching

In a computer system that supports more than one process at once, some mechanism must be used to switch
from one task to another. There are two main methods used to perform this switching:

• Co-operative switching means that a task that is currently running will voluntarily give up the processor at
some time, allowing other processes to run;

• Preemptive switching means that a running task will be interrupted (forced to give up) and the processor
given to another waiting process.

The problem with co-operative switching is one process could hang and thus deny execution of other processes,
resulting in no work being done. An example of a co-operative system was Windows 3.1.

Pre-emptive scheduling is better. It gives more response to all processes and helps prevent (or reduce the
number of occurrences of) the dreaded machine lockup.

A multi-user operating system allows more than one user to share the same computer system at the same time. It
does this by time-slicing the computer processor at regular intervals between the various programs run by each
user.

In this example, there are five people that share the processor hardware and main memory on a time basis.
Consider a 486 Intel processor running at 50MHz. This processor is capable of about 6 million instructions per
second.

If we decided that we would share the hardware by letting each user run for 1/5th of a second, this would mean
each user could execute about 1.2 million instructions each time they have the processor.
We start off by giving the first user (which we will call Bart) the processor hardware, and run Bart's program for 1/5th of a second. When the time is up, we intervene, save Bart's program state (program code and data) and then start running the second user program (for 1/5th of a second).

This process continues till we eventually get back to user Bart. To continue running Bart's program, we restore the programs code and data and then run for 1/5th of a second.

### 2.5. The Process

A process or task is a portion of a program in some stage of execution. A program can consist of several tasks, each working on their own or as a unit (perhaps periodically communicating with each other).

Each process that runs in an operating system is assigned a process control block that holds information about the process, such as a unique process ID (a number used to identify the process), the saved state of the process, the process priority and where it is located in memory.

The process priority is used to determine how often the process receives processor time. The operating system may run all processes with the same priority, or it may run some processes more often that others. Processes that have been waiting a long time for execution by the processor may have their priority increased so that they will be more likely to be executed in the future.

A process in a computer system may be in one of a number of different possible states, such as:

- **ready** – if it can run when the processor becomes free;
- **running** – it currently has the processor;
- **blocked** – it cannot run when the processor becomes free.

When a running process is interrupted by the processor after completing its allotted time, its state is saved in its process control block, its process state changed to ready and its priority adjusted.

When a running process accesses an input or output device, or for some reason cannot continue, it is interrupted by the processor, the process state and associated data is saved in the associated process control block. The process state is changed to blocked and the priority adjusted.

When the scheduler decides the next task to run, it changes the process state of the selected process to running and loads the saved data associated with that process back into the processor.

### 2.6. Background and foreground process

Multi-tasking systems support foreground and background processes (tasks). A foreground task is one that the user interacts directly with using the keyboard and screen. A background task is one that runs in the background (it does not have access to the screen or keyboard). Background tasks are usually used for printing. Windows NT Workstation and Windows 95/98 assign a higher priority to foreground tasks.

The operating system may be loaded into the computers memory in two ways:

- it is already present in ROM (so is permanent, immediately accessible and difficult to update),
- it is loaded from disk when the computer is turned on.

If the operating system is already present in ROM (for systems like industrial controllers, petrol pumps etc.), it will gain control immediately the processor is powered on. This method is best suited for small appliances and hand held devices where the operating system is relatively simple and small.

For more complex systems, the operating system is usually stored on secondary media (such as disk), and is loaded into the computer memory (RAM) when the computer is powered on. Advantages of this type of system are that changes to the operating system are easier to make and implement.

### 2.7. The Bootstrap Process
The bootstrap process describes the task of initially loading the operating system from disk into RAM (Figure 3.3). A small routine stored in ROM, called the BOOTSTRAP LOADER or IPL (Initial Program Loader), reads a special load routine from the diskette.

### 3.3. ábra - Figure 3.3: Bootstrap process

![Bootstrap process diagram](image)

1: Request OS Loader routine from disk unit
2: Transfer OS loader into memory
3: OS loader loads rest of OS into memory

In floppy based system, this routine is normally located on Track 00, sector 00 (or 01), and is called the boot sector. The code contained in the sector is transferred into RAM, then executed. It has the sole responsibility for loading the rest of the operating system into memory.

### 2.8. Different types of Operating Systems

Operating systems are divided into categories that define their characteristics. Modern systems may use combinations of those described below.

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BATCH</strong></td>
<td>The earliest type, allowed only one program to run at a time. The program was entered into the computer, then run till completed. The data used by the program could not be modified whilst the program was running. Any errors in the program or data mean starting all over again.</td>
</tr>
<tr>
<td><strong>INTER-ACTIVE</strong></td>
<td>These allow the modification and entry of data whilst the program is running. Typical systems are airline reservations and languages such as BASIC.</td>
</tr>
<tr>
<td><strong>TIME-SHARING/MULTI-USER</strong></td>
<td>These share the computer system amongst more than one user, and employ pre-emptive scheduling techniques.</td>
</tr>
<tr>
<td><strong>MULTI-TASKING</strong></td>
<td>More than one process may be executed at once. The processor is switched rapidly between the processes. A user may run more than one process at a time.</td>
</tr>
<tr>
<td><strong>REAL-TIME</strong></td>
<td>Primarily used in process control, telecommunications, etc. The OS monitors various inputs which affect the execution of processes, changing the computers model of the environment, thus affecting the outputs, within a guaranteed time period (usually &lt; 1 second).</td>
</tr>
<tr>
<td><strong>MULTI-PROCESSOR</strong></td>
<td>A computer that has more than one processor dedicated to running processes.</td>
</tr>
</tbody>
</table>
Embedded systems are also known as dedicated systems. This is because they only perform a specific task, and cannot run a wide variety of programs like a home computer (which we previously identified as a general purpose system). Windows NT workstations are an interactive, multitasking multiprocessor operating system. Windows 98 is an interactive, multitasking operating system. Linux is an interactive, multitasking, multiprocessor, multi-user operating system.

### 2.9. The shell

A shell is a program that handles user input and output. It provides routines for handling user input from a keyboard or mouse, as well as routines for displaying information on the terminal screen. A shell also provides a mechanism to interpret user commands and run additional programs that users request. In program called command.com was the shell in MS-DOS. In Windows 98 and NT workstation this shell is still present (it is called the command prompt and you run it by accessing the command prompt icon). The UNIX shell is called the bourne shell, and is a program called sh. There are other UNIX shell programs, notably the kourne shell and the C shell.

### 3. File Management

An operating system must provide a number of operations associated with files so that users can safely store and retrieve data.

Typical operations are: Open, Close, Create, Copy, Rename, List.

In addition, operations on single data elements within a file are supported by: Read, Write, Seek.

#### 3.1. The File Control Blocks

File control blocks (FCB), sometimes referred to as file descriptors, are data structures that hold information about a file. When an operating system needs to access a file, it creates an associated file control block to manage the file.

The structure of the file control block differs between operating systems, but most file control blocks include the following parts:

- Filename;
- Location of file on secondary storage;
- Length of file;
- Date and time or creation or last access.

#### 3.2. File names

Each operating system uses a specific convention or practice for naming files. MS-DOS Uses eight character file names, a dot, then a three-character extension that denotes the type of file. Filenames are not case-sensitive. UNIX Filenames can be up to 254 characters long and are case-sensitive. Windows Filenames can be up to 255 characters long and are not case-sensitive.

#### 3.3. File types

File types refer to classifying the content of the file, such as a program, text file, executable program or data file. In Windows operating systems, the file type is derived from the filename extension. Typical file types and their extensions are:
<table>
<thead>
<tr>
<th>File Extension</th>
<th>File Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>.bas</td>
<td>basic source program</td>
</tr>
<tr>
<td>.c</td>
<td>c source program</td>
</tr>
<tr>
<td>.dll</td>
<td>system library</td>
</tr>
<tr>
<td>.doc</td>
<td>Word document</td>
</tr>
<tr>
<td>.exe</td>
<td>executable program</td>
</tr>
<tr>
<td>.txt</td>
<td>text file</td>
</tr>
</tbody>
</table>

Windows associates applications (programs) with specific file types. For example, the default application that opens to process a file of type .txt is the Notepad editor.

### 3.4. Track of files

The hard disk is comprised of a large number of sequentially numbered sectors. As files are created, free sectors are allocated to hold the file contents and marked as allocated. To keep track of the sectors and whether they are allocated or free, and to which file they belong, the operating system maintains a number of tables.

### 3.5. Root file system

When the operating system is first installed, it creates a root file system on the disk that specifies how many sectors are available and how they will be allocated. The root file system is a table of entries like a directory. In general, this is a fixed size, and once full, no more entries can be added. Each entry can be either a file or another directory table. The following table depicts this structure (Figure 3.4).

#### 3.4. ábra - Figure 3.4: File Root System

![File Root System Diagram](image)

### 3.6. File systems supported by Windows operating systems

The Windows operating system supports the following file systems:

<table>
<thead>
<tr>
<th>File System</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAT</td>
<td>The MS-DOS operating system introduced the File Allocation Table system of keeping track of file entries and free clusters. Filenames where restricted to eight characters with an addition three...</td>
</tr>
</tbody>
</table>
characters signifying the file type. The FAT tables were stored at the beginning of the storage space.

<table>
<thead>
<tr>
<th>Operating System</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAT32</td>
<td>An updated version of the FAT system designed for Windows 98. It supports file compression and long filenames.</td>
</tr>
<tr>
<td>NTFS</td>
<td>Windows NT introduced the NT File System, designed to be more efficient at handling files than the FAT system. It spreads file tables throughout the disk, beginning at the center of the storage space. It supports file compression and long filenames.</td>
</tr>
</tbody>
</table>

### 3.7. Access-control lists and file permissions

In multi-user operating systems, files may be accessed by multiple users. Permission rights associated with folders (directories) and files are used to protect or restrict access to files. In UNIX these rights are known as Read, Write and Execute. In Windows NT and Windows 2000 (using the NTFS file-system), additional file permissions are available.

### 3.8. Symbolic ulink or shortcut

A symbol ulink is a filename that ulinks to another file. Consider the case on a UNIX box where three different mail packages are available. The administrator only wants to reference the mail using the command “mail”, so the filename is made to point to (or reference) the desired mail package. When the administrator runs the command “mail”, the appropriate mail package that is referenced by the symbolic ulink runs. In Windows, a similar capability is known as a shortcut.

### 4. The Windows Operating System

Windows (from Windows 2000 onward) is a significant example of what has become the new wave in microcomputer operating systems (other examples are Linux and MacOS). Windows was driven by a need to exploit the processing capabilities of today’s 32-bit and 64-bit microprocessors, which rival mainframes of just a few years ago in speed, hardware sophistication, and memory capacity. One of the most significant features of these new operating systems is that, although they are still intended for support of a single interactive user, they are multitasking operating systems. Two main developments have triggered the need for multitasking on personal computers, workstations, and servers. First, with the increased speed and memory capacity of microprocessors, together with the support for virtual memory, applications have become more complex and interrelated. For example, a user may wish to employ a word processor, a drawing program, and a spreadsheet application simultaneously to produce a document. Without multitasking, if a user wishes to create a drawing and paste it into a word processing document, the following steps are required:

- Open the drawing program;
- Create the drawing and save it in a file or on a temporary clipboard;
- Close the drawing program;
- Open the word processing program;
- Insert the drawing in the correct location.

If any changes are desired, the user must close the word processing program, open the drawing program, edit the graphic image, save it, close the drawing program, open the word processing program, and insert the updated image. This becomes tedious very quickly. As the services and capabilities available to users become more powerful and varied, the single-task environment becomes more clumsy and user unfriendly. In a multitasking environment, the user opens each application as needed, and leaves it open. Information can be moved around among a number of applications easily. Each application has one or more open windows, and a graphical interface with a pointing device such as a mouse allows the user to navigate quickly in this environment. A second motivation for multitasking is the growth of client/server computing. With client/server computing, a personal computer or workstation (client) and a host system (server) are used jointly to accomplish a particular application. The two are ulinked, and each is assigned that part of the job that suits its capabilities. Client/server can be achieved in a local area network of personal computers and servers or by means of a ulink between a user.
system and a large host such as a mainframe. An application may involve one or more personal computers and one or more server devices. To provide the required responsiveness, the OS needs to support high-speed networking interfaces and the associated communications protocols and data transfer architectures while at the same time supporting ongoing user interaction. The foregoing remarks apply to the desktop versions of Windows. The Server versions are also multitasking but may support multiple users. They support multiple local server connections as well as providing shared services used by multiple users on the network. As an Internet server, Windows may support thousands of simultaneous Web connections.

4.1. Architecture

Its modular structure gives Windows considerable flexibility. It is designed to execute on a variety of hardware platforms and supports applications written for a variety of other operating systems. As of this writing, desktop Windows is only implemented on the Intel x86 and AMD64 hardware platforms. Windows server also supports the Intel IA64 (Itanium).

As with virtually all operating systems, Windows separates application-oriented software from the core OS software. The latter, which includes the Executive, the Kernel, device drivers, and the hardware abstraction layer, runs in kernel mode. Kernel mode software has access to system data and to the hardware. The remaining software, running in user mode, has limited access to system data.

4.2. Operating System Organization

Windows has a highly modular architecture. Each system function is managed by just one component of the OS. The rest of the OS and all applications access that function through the responsible component using standard interfaces. Key system data can only be accessed through the appropriate function. In principle, any module can be removed, upgraded, or replaced without rewriting the entire system or its standard application program interface (APIs). The kernel-mode components of Windows are the following:

• Executive: Contains the base OS services, such as memory management, process and thread management, security, I/O, and interprocess communication.

• Kernel: Controls execution of the processor(s). The Kernel manages thread scheduling, process switching, exception and interrupt handling, and multiprocessor synchronization. Unlike the rest of the Executive and the user level, the Kernel’s own code does not run in threads.

• Hardware abstraction layer (HAL): Maps between generic hardware commands and responses and those unique to a specific platform. It isolates the OS from platform-specific hardware differences. The HAL makes each computer’s system bus, direct memory access (DMA) controller, interrupt controller, system timers, and memory module look the same to the Executive and Kernel components. It also delivers the support needed for symmetric multiprocessing (SMP), explained subsequently.

• Device drivers: Dynamic libraries that extend the functionality of the Executive. These include hardware device drivers that translate user I/O function calls into specific hardware device I/O requests and software components for implementing file systems, network protocols, and any other system extensions that need to run in kernel mode.

• Windowing and graphics system: Implements the graphical user interface (GUI) functions, such as dealing with windows, user interface controls, and drawing. The Windows Executive includes components for specific system functions and provides an API for user-mode software. Following is a brief description of each of the Executive modules.

• I/O manager: Provides a framework through which I/O devices are accessible to applications, and is responsible for dispatching to the appropriate device drivers for further processing. The I/O manager implements all the Windows I/O APIs and enforces security and naming for devices, network protocols, and file systems (using the object manager).

• Cache manager: Improves the performance of file-based I/O by causing recently referenced file data to reside in main memory for quick access, and by deferring disk writes by holding the updates in memory for a short time before sending them to the disk.
• Object manager: Creates, manages, and deletes Windows Executive objects and abstract data types that are used to represent resources such as processes, threads, and synchronization objects. It enforces uniform rules for retaining, naming, and setting the security of objects. The object manager also creates object handles, which consist of access control information and a pointer to the object. Windows objects are discussed later in this section.

• Plug-and-play manager: Determines which drivers are required to support a particular device and loads those drivers.

• Power manager: Coordinates power management among various devices and can be configured to reduce power consumption by shutting down idle devices, putting the processor to sleep, and even writing all of memory to disk and shutting off power to the entire system.

• Security reference monitor: Enforces access-validation and audit-generation rules. The Windows object-oriented model allows for a consistent and uniform view of security, right down to the fundamental entities that make up the Executive. Thus, Windows uses the same routines for access validation and for audit checks for all protected objects, including files, processes, address spaces, and I/O devices.

• Virtual memory manager: Manages virtual addresses, physical memory, and the paging files on disk. Controls the memory management hardware and data structures which map virtual addresses in the process’s address space to physical pages in the computer’s memory.

• Process/thread manager: Creates, manages, and deletes process and thread objects.

• Configuration manager: Responsible for implementing and managing the system registry, which is the repository for both system wide and per-user settings of various parameters.

• Local procedure call (LPC) facility: Implements an efficient cross-process procedure call mechanism for communication between local processes implementing services and subsystems. Similar to the remote procedure call (RPC) facility used for distributed processing.

4.3. User-Mode Processes

Four basic types of user-mode processes are supported by Windows:

• Special system processes: User mode services needed to manage the system, such as the session manager, the authentication subsystem, the service manager, and the logon process.

• Service processes: The printer spooler, the event logger, user mode components that cooperate with device drivers, various network services, and many, many others. Services are used by both Microsoft and external software developers to extend system functionality as they are the only way to run background user mode activity on a Windows system.

• Environment subsystems: Provide different OS personalities (environments). The supported subsystems are Win32/WinFX and POSIX. Each environment subsystem includes a subsystem process shared among all applications using the subsystem and dynamic unlink libraries (DLLs) that convert the user application calls to LPC calls on the subsystem process, and/or native Windows calls.

• User applications: Executables (EXEs) and DLLs that provide the functionality users run to make use of the system. EXEs and DLLs are generally targeted at specific environment subsystems; although some of the programs that are provided as part of the OS use the native system interfaces (NTAPI).

Windows is structured to support applications written for multiple OS personalities. Windows provides this support using a common set of kernel mode components that underlie the protected environment subsystems. The implementation of each subsystem includes a separate process, which contains the shared data structures, privileges, and Executive object handles needed to implement a particular personality. The process is started by the Windows Session Manager when the first application of that type is started. The subsystem process runs as a system user, so the Executive will protect its address space from processes run by ordinary users. A protected subsystem provides a graphical or command-line user interface that defines the look and feel of the OS for a user. In addition, each protected subsystem provides the API for that particular operating environment. This means that applications created for a particular operating environment may run unchanged on Windows, because the OS interface that they see is the same as that for which they were written. The most important
subsystem is Win32. Win32 is the API implemented on both Windows NT and Windows 95 and later releases of Windows 9x. Many Win32 applications written for the Windows 9x line of operating systems run on NT systems unchanged. At the release of Windows XP, Microsoft focused on improving compatibility with Windows 9x so that enough applications (and device drivers) would run that they could cease any further support for 9x and focus on NT. The most recent programming API for Windows is WinFX, which is based on Microsoft’s .NET programming model. WinFX is implemented in Windows as a layer on top of Win32 and not as a distinct subsystem type.

4.4. Client/Server Model

The Windows operating system services, the protected subsystems, and the applications are structured using the client/server computing model, which is a common model for distributed computing and which is discussed in Part Six. This same architecture can be adopted for use internal to a single system, as is the case with Windows. The native NT API is a set of kernel-based services which provide the core abstractions used by the system, such as processes, threads, virtual memory, I/O, and communication. Windows provides a far richer set of services by using the client/server model to implement functionality in user-mode processes. Both the environment subsystems and the Windows user-mode services are implemented as processes that communicate with clients via RPC. Each server process waits for a request from a client for one of its services (for example, memory services, process creation services, or networking services). A client, which can be an application program or another server program, requests a service by sending a message. The message is routed through the Executive to the appropriate server. The server performs the requested operation and returns the results or status information by means of another message, which is routed through the Executive back to the client.

Advantages of a client/server architecture include the following:

- It simplifies the Executive. It is possible to construct a variety of APIs implemented in user-mode servers without any conflicts or duplications in the Executive. New APIs can be added easily.

- It improves reliability. Each new server runs outside of the kernel, with its own partition of memory, protected from other servers. A single server can fail without crashing or corrupting the rest of the OS.

- It provides a uniform means for applications to communicate with services via RPCs without restricting flexibility. The message-passing process is hidden from the client applications by function stubs, which are small pieces of code which wrap the RPC call. When an application makes an API call to an environment subsystem or service, the stub in the client application packages the parameters for the call and sends them as a message to a server subsystem that implements the call.

- It provides a suitable base for distributed computing. Typically, distributed computing makes use of a client/server model, with remote procedure calls implemented using distributed client and server modules and the exchange of messages between clients and servers. With Windows, a local server can pass a message on to a remote server for processing on behalf of local client applications. Clients need not know whether a request is serviced locally or remotely. Indeed, whether a request is serviced locally or remotely can change dynamically based on current load conditions and on dynamic configuration changes.

5. The Unix Operating System

This was the first hint that UNIX would be an operating system for all computers. The next important milestone was the rewriting of UNIX in the programming language C. This was an unheard-of strategy at the time. It was generally felt that something as complex as an operating system, which must deal with timecritical events, had to be written exclusively in assembly language. Reasons for this attitude include the following:

- Memory (both RAM and secondary store) was small and expensive by today’s standards, so effective use was important. This included various techniques for overlaying memory with different code and data segments, and self-modifying code.

- Even though compilers had been available since the 1950s, the computer industry was generally skeptical of the quality of automatically generated code. With resource capacity small, efficient code, both in terms of time and space, was essential.

- Processor and bus speeds were relatively slow, so saving clock cycles could make a substantial difference in execution time.
• The C implementation demonstrated the advantages of using a high-level language for most if not all of the system code. Today, virtually all UNIX implementations are written in C.

These early versions of UNIX were popular within Bell Labs. In 1974, the UNIX system was described in a technical journal for the first time [RITC74]. This spurred great interest in the system. Licenses for UNIX were provided to commercial institutions as well as universities. The first widely available version outside Bell Labs was Version 6, in 1976. The follow-on Version 7, released in 1978, is the ancestor.

General UNIX Architecture

Of most modern UNIX systems. The most important of the non-AT&T systems to be developed was done at the University of California at Berkeley, called UNIX BSD (Berkeley Software Distribution), running first on PDP and then VAX computers. AT&T continued to develop and refine the system. By 1982, Bell Labs had combined several AT&T variants of UNIX into a single system, marketed commercially as UNIX System III. A number of features was later added to the operating system to produce UNIX System V.

The OS is often called the system kernel, or simply the kernel, to emphasize its isolation from the user and applications. It is the UNIX kernel that we will be concerned with in our use of UNIX as an example in this book. UNIX also comes equipped with a number of user services and interfaces that are considered part of the system.

5.1. Modern Unix Systems

As UNIX evolved, the number of different implementations proliferated, each providing some useful features. There was a need to produce a new implementation that unified many of the important innovations, added other modern OS design features, and produced a more modular architecture.

5.2. System V Release 4 (SVR4)

SVR4, developed jointly by AT&T and Sun Microsystems, combines features from SVR3, 4.3BSD, Microsoft Xenix System V, and SunOS. It was almost a total rewrite Common facilities of the System V kernel and produced a clean, if complex, implementation. New features in the release include real-time processing support, process scheduling classes, dynamically allocated data structures, virtual memory management, virtual file system, and a preemptive kernel. SVR4 draws on the efforts of both commercial and academic designers and was developed to provide a uniform platform for commercial UNIX deployment. It has succeeded in this objective and is perhaps the most important UNIX variant. It incorporates most of the important features ever developed on any UNIX system and does so in an integrated, commercially viable fashion. SVR4 runs on processors ranging from 32-bit microprocessors up to supercomputers.

5.3. BSD

The Berkeley Software Distribution (BSD) series of UNIX releases have played a key role in the development of OS design theory. 4.xBSD is widely used in academic installations and has served as the basis of a number of commercial UNIX products. It is probably safe to say that BSD is responsible for much of the popularity of UNIX and that most enhancements to UNIX first appeared in BSD versions. 4.4BSD was the final version of BSD to be released by Berkeley, with the design and implementation organization subsequently dissolved. It is a major upgrade to 4.3BSD and includes a new virtual memory system, changes in the kernel structure, and a long list of other feature enhancements. One of the most widely used and best documented versions of BSD is FreeBSD. FreeBSD is popular for Internet-based servers and firewalls and is used in a number of embedded systems. The latest version of the Macintosh operating system, Mac OS X, is based on FreeBSD 5.0 and the Mach 3.0 microkernel.

5.4. Solaris 10

Solaris is Sun’s SVR4-based UNIX release, with the latest version being 10. Solaris provides all of the features of SVR4 plus a number of more advanced features, such as a fully preemptable, multithreaded kernel, full support for SMP, and an object-oriented interface to file systems. Solaris is the most widely used and most successful commercial UNIX implementation.
UNIX System V makes use of a simple but powerful process facility that is highly visible to the user. UNIX follows the model, in which most of the OS executes within the environment of a user process. UNIX uses two categories of processes: system processes and user processes. System processes run in kernel mode and execute operating system code to perform administrative and housekeeping functions, such as allocation of memory and process swapping. User processes operate in user mode to execute user programs and utilities and in kernel mode to execute instructions that belong to the kernel. A user process enters kernel mode by issuing a system call, when an exception (fault) is generated, or when an interrupt occurs.

6. The Linux Operating System

An operating system (OS) is actually a collection of programs that gives a computer the critical functionality that lets you use it for work and entertainment. This book will explore the common operating systems used on microcomputers, but before we do, let’s answer a few general questions you may have: What is a microcomputer? What’s inside a microcomputer? What types of microcomputers are in use today? This section answers these questions. A microcomputer is a computer built around a microprocessor, a special integrated circuit (IC) that performs the calculations, or processing, of the computer. An IC, commonly called a chip, is a small electronic component made up of transistors (tiny switches) and other miniaturized parts. The first patents for integrated circuits were issued in 1959 to Jack Kilby of Texas Instruments and Robert Noyce of Fairchild Semiconductor Corporation. Today there are many ICs in a computer in addition to the microprocessor or central processing unit (CPU), also simply called a processor. A microcomputer is small enough and cheap enough to be dedicated to the use of a single person. This was a revolutionary idea in the 1970s, when microcomputers first became available; in the previous decades, computers were physically very large because bulky vacuum tubes served the purpose that microscopic transistors now serve.

6.1. Inside a Microcomputer

Our friend Brianna uses a PC at work and a Macintosh at home, and she will soon take night classes in which she will use a laptop PC that she carries to and from school. She wants to learn more about the computers that she uses each day, beginning with the hardware.

6.2. The Basic PC

Each computer that Brianna and the rest of us use is a metal and plastic hardware contraption composed of many components, some of which allow us to interact with the computer. In computerese, we call interaction with a computer input/output (I/O). When you send something into the computer – for instance, when you enter information via the keyboard or have your word processing program read a file from a disk – it is being input. When something comes out of the computer, such as the text and graphics you see on the display screen or the printed results on paper, it is being output from the computer.

Regardless of the brand of microcomputer you use, the common hardware components are basically the same. In general, they include at least one CPU per computer, a central circuit board called a motherboard, and random access memory (RAM) that acts as the main memory for holding active programs and their data. There are one or more chips called ROM BIOS, a keyboard, a pointing device (mouse or some sort of touch pad), disk drives, and other peripheral devices such as printers, scanners, and cameras.

6.3. More About ROM BIOS

Central to any motherboard, and as critical to the operation of a PC as the CPU, is the system ROM BIOS (read only memory basic input output system). System ROM BIOS contains the program code that informs the processor of the devices present and how to communicate with them. Additionally, most components and peripherals have their own ROM BIOS with program code for operating that component. The ROM BIOS for most components is limited to small programs for basic communication between the operating system and the component. Supplementing or replacing the ROM BIOS – even parts of the central system ROM BIOS – are device drivers, special programs installed into an operating system. Each device driver contains code for controlling a component. These are an extension of BIOS, usually allowing much more control of a device than the small programs stored in that device’s ROM BIOS.

You can see evidence of the system ROM BIOS and other BIOSes in a PC if you carefully watch the screen as you power up the computer. Each ROM BIOS that powers up with the computer will perform a test. The system
ROM BIOS test is known as the power on self-test (POST) and much of the information you see on the screen is the result of the POST and the tests of additional ROM BIOSs on the computer’s components.

7. Types of Microcomputers

In the 1970s, very few computers were small enough to sit on a desk. One computer, the Control Data 160A, actually was the desk! But people wanted computers on their desks, so what they used was a dumb terminal consisting of a CRT (cathode-ray tube) and a keyboard connected to a large mainframe computer. (Large stand-alone computers are called mainframes.) In the late 1970s, when computers became small enough to sit on a desk, someone coined the word microcomputer and it became widely used to describe the early forms of these computers. We’re now more likely to use the term personal computer (PC), which applies to computers that comply with hardware standards set and supported by Microsoft, Intel (the largest computer chip manufacturer), and to a lesser extent, other companies. We call these the Microsoft/Intel standards (also called Wintel). However, many important microcomputers don’t comply with these standards most notably, those from Apple and the small handheld computers such as smartphones that are so popular today.

7.1. Desktops and Laptops

A desktop computer is a computer designed to spend its useful life in one location on a desk. A portable computer is designed to be easily carried from one location to another. There are many sizes and types of portable computers. Our discussion of operating systems in this book is limited to the most common operating systems that run on modern desktop and laptop computers. They include Windows and Linux, both of which run on PCs, and Mac OS X, which runs on Apple computers. Early Macintosh computers came with sophisticated graphics abilities that made them attractive to users requiring high-quality graphical and multimedia support. Part of Apple’s early marketing plans simply gave computers to schools, so today the Macintosh has ardent supporters in education and in the graphics businesses. Eventually Microsoft/Intel compatible computers closed the gap in graphics and multimedia capabilities. In the decades since the introduction of the IBM PC, the majority of desktop and laptop computers used in private and public organizations complied with the Microsoft/Intel standard, with Macintosh computers a distant second. In recent years Apple desktop and portable computers have made great gains in market share. In fact, according to IDC, a market research firm, Apple is third in portable computer sales if you count the company’s iPad models along with the MacBook laptop models. If you exclude iPads, Apple is fourth in sales.

7.2. Servers

You can also use a PC or Mac as a server, a computer that plays several important roles in a network. In all of these roles, it provides services to other computers, which is why it is called a server. We call a computer on the receiving end of these services a client. We may base server functions on the same hardware components found in desktop computers, but they are beefed up considerably for a server to which hundreds or thousands of users must connect – resulting in the computer equivalent of a heavyweight versus a lightweight boxer. The cost, which can run in the thousands of dollars for a server versus only a few hundred dollars for an average PC, reflects this difference. What kind of services does a server provide? When we use a server to store all of the data files of the users in a department or company, we call it a file server. If a server has one or more printers connected to it that it shares with users on the network, we call it a print server. If a server can offer multiple services at the same time. Other servers may offer messaging services (e-mail and fax), Web services, and many, many others. It takes specialized software to provide each type of server service, and complementary client software to request each type of service.

7.3. Handheld Devices

Computers today include a long list of various devices that don’t have computer in their name, such as handheld devices, games, stoves, refrigerators, TVs, and DVD and Blu-ray players. Many handheld devices, often proprietary, comply with no, or very few, standards in their design, but they are still computers because they contain processors. They include a wide variety of products ranging from simple handheld computers to multifunction mobile devices, such as the devices used in grocery stores to track inventory. Many handheld devices run proprietary OSs, while others run scaled-down versions of desktop OSs. The most popular handheld devices are wireless phones, called smartphones, that not only allow voice communications, but also let you connect to the Internet, view your e-mail, and do many other things on their small color screens. Examples of
smartphones are Apple Computer’s iPhone, RIM’s BlackBerry, Microsoft’s Windows Phone, and various models by Palm, Nokia, HTC, Samsung, LG, and others. Operating systems designed specifically for use on smartphones include Google’s Android, Palm’s webOS, Apple’s iOS, and Windows Mobile.

7.4. Smartphones

Although you will not really study handheld devices in this book, to satisfy your curiosity about smartphones try this:

- Use an Internet search engine, such as www.google.com, and search on the term “smartphones.” Browse through the sites you find in the search engine. Results will vary, but some likely sites are www.htc.com, www.apple.com/iphone, and www.blackberry.com.

- What OSs do the latest smartphones use?

The functions of an operating system for an operating system between an application and the hardware, since the single-purpose software interacted with all of the hardware of the system. When programmers (also known as “developers”) write an application, they design the application to interact with the operating system and make all requests for hardware services through the operating system. To do this, they must write the program to use the correct commands to request operating system services. The operating system, in turn, interacts with the hardware on behalf of the application and fulfills the requests the application made. Today, because our computers are no longer such dedicated systems, the operating system performs several functions. We’ll study them next.

8. Functions of Operating Systems

Justin works part-time in a legal office and is a full-time student at a community college, where he is in the computer information systems (CIS) track. Recently, he took a Saturday community education class in computer graphics. At work, he has Windows 7 on his desktop computer; at home he uses Windows XP; and in his graphics class, he used a Macintosh. His next class at the community college will involve working with Linux. Although Justin’s experience might seem extreme, it illustrates that you are likely to encounter different desktop operating systems at work, school, and home. In addition, as computers proliferate, it becomes more important to learn the common characteristics that they share. Justin spends most of his time on each computer he uses working in one or another specific application, such as a word processor, a graphical drawing program, or an Internet browser. However, he must also perform tasks outside of these applications, such as logging onto the computer, launching each application, managing files, and even troubleshooting the occasional problem that may arise with the computer. But each of these different computers requires different ways of doing things so he wants to gain a better understanding of the OSs so that he can both perform better on the job and feel more comfortable while working on the various computers. He wants to learn what an OS is and what functions it performs, as the following sections describe. An operating system is the software (or group of programs) that acts as the central control program for the computer. As such, it is loaded (or “booted up,” a derivation of the adage “lifting yourself by your own bootstraps”) when the computer is turned on. Its main component, the kernel, always remains in memory while the computer is running, managing low-level (close-to-the-hardware) OS tasks. The operating system acts as an intermediary between the applications and the hardware. Until we made the transition from single-purpose machines to multipurpose machines, there was no need.

8.1. User Interface

The user interface is the software layer, sometimes called the shell, through which the user communicates with the OS. The OS, in turn, communicates with the computer. Thus, the user interface includes the command processor, which loads programs into memory, as well as the many visual components of the operating system (what you see when you look at the display). On a computer running DOS or Linux (without a graphical shell), this visual component consists of a character-based command line that provides only sparse amounts of information. This is the command-line interface (CLI). On the other hand, Apple’s Mac and Microsoft’s Windows operating systems provide an information-rich graphical user interface (GUI), fully integrated into the operating system. It is through this GUI that you communicate with the OS and the computer. The GUI offers menus and graphical icons (small graphics) that allow you to use a pointing device to select programs to run and to perform many other tasks, such as opening a word processor file. Although you do not have to memorize arcane commands, working within a GUI does require you to learn the meaning of the various graphical pieces that make up the GUI and how to navigate among these pieces to find your programs and data. In addition, you
must learn how to make a program become active (to start it running) so that you can get your work or play done. Notice the icons and other graphical components, such as the bar at the bottom containing the button showing the Microsoft logo and the cursor. In a GUI the cursor is sometimes replaced by a graphical pointer that can have a variety of shapes that you can move around by manipulating a pointing device – usually a mouse, trackball, or touch pad (on a laptop). The pointer allows you to select or manipulate objects in the GUI, which serves as a metaphor for commands. For example, to delete an item, you pick it up and put it into the recycle bin. By contrast, in a CLI you would type a command such as “delete myfile.txt.”

8.2. Job Management

Job management is an operating system function that controls the order and time in which programs are run. Two examples of programs that may perform this function are a scheduling program that schedules other programs or batch files to run on a certain day and time, and a print program that manages and prioritizes multiple print jobs.

8.3. Task Management

Task management is an operating system function found in multitasking operating systems. Multitasking implies that a computer is simultaneously running two or more programs (tasks) at the same time. In reality, a computer cannot run more tasks simultaneously than the number of processors that exist within the computer. Until recently, most desktop and laptop computers had only a single processor, switching between tasks so fast that it seems to be executing a number of tasks at the same time. New processors can have multiple CPUs within a single chip, so they have true multitasking. Task management controls the focus (where the system’s attention is at any given moment). It also allows the user to switch between tasks by giving the focus to the application the user brings to the foreground. In Windows, this application runs in the current window – the window that is on top of other windows on the screen and the window that receives input from the keyboard when the user types.

8.4. Memory Management

Memory management is an operating system function that manages the placement of programs and data in memory, while keeping track of where it put them. In the case of advanced operating systems, such as Windows XP and later Windows versions, this involves a scheme for making optimal use of memory. Virtual memory allows more code and data to be in memory than what the actual physical system memory can hold. Using a memory management OS component called the virtual memory manager, these operating systems move code and data, as necessary, to a portion of the disk defined as virtual memory, meaning that this disk space is used as if it were memory, not just disk storage space. The OS performs this transfer for code and data that are part of any program that currently does not have the user’s attention because this unneeded-now information does not have to be kept in RAM for immediate use, so other programs that do need to use the memory now can do so.

8.5. File Management

File management, also referred to as data management, is an operating system function that allows the operating system to read, write, and modify data, while managing the logical storage of the data. Each operating system has at least one scheme of logical organization, called a file system. A file system is the logical structure used on a storage device for the purpose of storing files, as well as the code within an operating system that allows the OS to store and manage files on a storage device. The logical structure is written to a storage device when an operating system uses a technique called formatting. The operating system maps the logical organization of the file system to physical locations on the storage device, most often a conventional hard disk drive or solid state drive so that it can store and retrieve the data. The logical structure of a file system stores metadata, which is data about the stored files.

Normally, a single storage device will have only a single file system, residing in a single area defined as a partition, but most operating systems allow a storage device to have more than one partition. A partition may be an entire drive volume or just a portion of a drive, and it is assigned some identifier, for example, a drive letter like C. In DOS and Windows, the drive letter is followed by a colon, so that a complete drive name may be C:. We often call this a logical drive. Within the logical structure of a file system, data is organized into entities called files that are saved to storage devices (usually disks). File management also allows users to organize their files, using other special files that act as containers. These special files, called folders or directories, can contain files as well as other folders. There are file systems for magnetic media (hard drives), file systems for optical...
media, and special file systems for working over the network. Each of the operating systems discussed in this
book supports several types of file systems. Examples of file systems for magnetic media include the FAT file
system that originated with both PC DOS and MS-DOS; the ext2, ext3, JFS, VFS, and ReiserFs file systems for
Linux; the FAT32 and NTFS file systems for Windows; and the Mac OS files systems, HFS and HFS+.

8.6. Device Management

The device management function controls hardware devices by using special software called device drivers,
which are installed in the operating system. Device drivers are unique to the device and the manufacturer of the
device creates them to work with a specific operating system. For instance, a printer or video adapter will come
with drivers for several different operating systems. The device driver contains the commands understood by the
device and uses these commands to control the device in response to requests it receives from the operating
system. You need a component-specific device driver for each unique hardware component with which the
operating system interacts.

8.7. Security

The security function of an operating system provides password-protected authentication of the user before
allowing access to the local computer and may restrict what someone can do on a computer. For example,
Rachel is the accounting clerk in a small company. She has confidential information on her computer, and she
doesn’t want just anyone to be able to walk up to her computer and access the information stored there. Rachel
can set up her computer so that anyone getting into it must log on with a user name and password from a user
account. A user account is nothing more than a name and an associated password stored inside the PC.

9. Categories of Operating Systems

Operating systems are organized into four categories, three of which are based on the number of simultaneous
tasks and the number of simultaneous users that can be served, while one category, real-time, is based on an
entirely different set of characteristics. The categories are:

• Single-user/single-tasking;
• Single-user/multitasking;
• Multiuser/multitasking;
• Real-time.

We’ll discuss each in turn.


A single-user/single-tasking operating system is one that allows only a single user to perform a single task at a
time. A task is a function such as reading a file from disk, performing a math calculation, printing a document,
or sending a request over the Internet to a Web server. Small and simple OSs can only manage a single task at a
time.

Examples of single-tasking OSs are MS-DOS and the Palm OS, used on the palmOne handheld computers.
Because they take up very little space on disk or in memory when they are running, a single-tasking OS does not
require a powerful and expensive computer. The current OSs on the iPhone and the iPad fall into this category
because they only allow a user to run one user application at a time.

9.2. Single-User/Multitasking

An operating system that allows a single user to perform two or more functions at once is a single-
user/multitasking operating system. Early versions of both Microsoft Windows and the Macintosh operating
systems were examples of this category. Thanks to these OSs and the applications that run on them, people
could accomplish more in less time, increasing their productivity. For instance, a single user can have two or
more programs open, share data between programs, and can instantly switch between them.
9.3. Multiuser/Multitasking Operating Systems

A multiuser/multitasking operating system allows multiple users to run programs simultaneously. We have had such operating systems for decades on mainframes and on today’s network servers, sometimes called terminal servers. This is not at all the same as connecting to a network server for the sake of accessing files and printers. When a computer connects to a file and print server to access document files to edit, the client computer performs the processing work locally. Not so with a multiuser server OS that gives each user a complete operating environment on the server that is separate from all other users. A terminal client is software on the user’s computer that establishes a connection to the terminal server. In a multiuser/multitasking operating system server environment, all or most of the computing occurs at the server. When connected to a multiuser/multitasking server, each user may have either a full-fledged PC or a thin client (a minimally configured PC). The terminal client runs software under Windows, Mac OS, or Linux, but with far lower hardware requirements than it would need if it ran all the processes locally. The terminal server providing a multiuser/multitasking OS to client computers may be a Unix, Linux, NetWare, or a special version of Microsoft Windows server. The multiuser capabilities of today’s OSs carry over to the user environment. Multiple users can have accounts on a single computer – desktop or server – and when each user logs on he sees his desktop with all his favorite gadgets and other settings, and he has access to his files and installed programs.

The desktop operating systems we survey in this book – Windows XP, Windows Vista, Windows 7, Linux, and Mac OS X – all are multiuser/multitasking operating systems with one caveat: only one user at a time can interactively log into the computer. However, a logged on user can simply log out, and leave their entire session of open applications and windows intact, and another user can log in and run a separate session, thus switching the current user. Therefore, Windows OSs support multiple but not simultaneous interactive users. All of the desktop OSs surveyed in this course support multiple simultaneous user connections over a network.

9.4. Real-Time Operating Systems

A real-time operating system (RTOS) is a very fast and relatively small OS. Often embedded, meaning it is built into the circuitry of a device and not normally loaded from a disk drive, a real-time operating system runs real-time applications. It may support multiple simultaneous tasks or it may only support single tasking. A real-time application responds to certain inputs extremely quickly – thousandths or millionths of a second (milliseconds or microseconds, respectively). They run medical diagnostics equipment, life-support systems, machinery, scientific instruments, and industrial systems. Real-time OSs are used in aviation and aerospace where the code must run within the allotted time. Output must be provided in time to make a course correction (in the case of aviation software), make a flow correction (in the case of a ventilator), or provide a warning to an operator. Examples of real-time operating systems include BlueCat Linux by LynuxWorks, QNX Neutrino by QNX Software Systems, Operating System Embedded (OSE), pSOS, and Windows CE. Real-time embedded systems are everywhere. Devices with real-time embedded systems control the movement of surveillance cameras suspended by cables over a sports arena. Other devices gather data from racecars and transmit the data and live video images of the race from each car. Real-time embedded systems are also used in unmanned military ground vehicles, and you’ll find real-time embedded systems in very large-scale machines, such as the huge sack, parcel, and large parcel sorting machines found in U.S. Postal Service bulk mail centers. A real-time operating system’s speed and ability to work with special real-time application programs defines it.

9.5. 16-, 32-, and 64-Bit OSs

We call an operating system that can take advantage of the addressing and processing features of a processor an x-bit OS. The original MS-DOS was a 16-bit OS, as was Windows 3.0 and its sub-versions. Windows 95, Windows 98, and Windows Millennium Edition were really hybrids, with mostly 32-bit pieces but some 16-bit pieces for downward compatibility. Windows XP had a 64-bit version, but it was not widely used, and you are unlikely to encounter it in the wild. The Windows versions, Mac OS X, and Linux OSs we discuss in this book are available in both 32-bit and 64-bit versions. All things being equal, the 64-bit version of an operating system will be faster than its 32-bit counterpart, but the biggest difference between the 32-bit and 64-bit versions of Windows is in the address space used by both system RAM and other RAM and ROM in your computer (see Table 1.1). Windows 64-bit does not use the maximum theoretical address space of a 64-bit CPU, which is 264, or 9.2 quintillion (nine followed by 18 digits). A 64-bit operating system requires 64-bit applications, although Microsoft has offered ways to support older applications in each upgrade of Windows, as described in this book. To determine if a Windows Vista or Windows 7 computer is running a 32-bit or 64-bit version, type “system” in the Start menu’s Start Search box (simply Search in Windows 7) and select “System” from the resulting list of Programs. The System Type field will say “32-bit Operating System” or “64-bit Operating System.” If you
purchase a new computer today with either Windows or the Mac OS preinstalled, it is most likely to be a 64-bit OS.

9.6. Yesterday’s Operating Systems

Sometimes people think that they can simply take the newest and best computer or other gadget and make it work without understanding anything about how it came to be. Well, they can. But they probably can’t fix it, modify it, or use it effectively without understanding how and why it came to be in the form it’s in now. One really can’t understand current PC technology without having a grasp of older PC technology. In other words, studying history is important to understand how we arrived at today. We’ll begin with arguably the oldest OS still in use today, with beginnings that predate microcomputers. Then we’ll explore the history of computers leading to today’s PCs and Mac desktop computers and the operating systems that evolved for each of these hardware platforms.

9.7. UNIX – The OS for All Platforms

In the context of this book, we have put UNIX under Yesterday’s Desktop OSs because UNIX has a longer history than any other popular operating system. It grew out of an operating system developed for an early Digital Equipment Corporation (DEC) computer and went through several generations of changes before it emerged from the Bell Labs Computing Science Research Center (Bell Labs) as UNIX version 6 in 1975. This was a portable operating system for minicomputers and mainframe computers distributed via government and commercial licenses and inexpensive academic licenses. A portable operating system is one that you can use on a variety of computer system platforms, with only minor alterations required for the underlying architecture. The University of California at Berkeley licensed UNIX, modified it, and distributed it to other schools as Berkeley Software Distribution (BSD) version 4. Later versions have followed. The schools paid licensing fees to Bell Labs. Students and others improved on and added to UNIX, freely sharing their code with each other. This tradition still prevails today with such versions of UNIX as Free BSD, Net BSD, Open BDS, and Open Solaris. Commercial versions of UNIX today include AIX, OpenServer (derived from SCO UNIX), and HP/UX. In certain environments UNIX is still very much alive. The current commercial versions of UNIX include Sun Microsystems’ Solaris, Hewlett-Packard’s HP-UX, IBM’s AIX, and Compaq’s Tru64 UNIX. These versions are high-end server applications and quite expensive, as are the computers they run on. There are also many open source versions of UNIX, including FreeBSD and NetBSD. Open source is a certification standard of the Open Source Initiative (OSI) through which a program’s source code (the original language in which a program is written) is made available free of charge to the general public. Learn more about open source at www.opensource.org. Even with these free versions available, however, it is worthwhile to buy one of the modestly priced packages from companies that charge small fees just for the value they have added to the OS in the form of additional software, installation and configuration instructions, and documentation. In addition to portability to many computer platforms, UNIX supports time-sharing and multiuser systems, and some versions run on personal computers. UNIX can be found today on very large computer systems (still referred to as mainframes) and on Intel desktop systems, as well as on a variety of midsize computers. Most versions of UNIX also offer several different user interfaces. Some use character mode, like the traditional shells, such as the Bourne shell and the C shell. Others use a graphical interface such as GNOME or KDE. Even fierce UNIX advocates do not see UNIX taking over the desktop any time soon. However, it is an excellent server operating system, because it uses resources carefully, allowing you to load only the services currently needed. It is also very secure, and versions of UNIX are present on many of the world’s Internet servers.

9.8. Today’s Desktop OSs

Today’s desktop operating systems include Windows Vista, Windows 7, Mac OS X, and Linux. The latest versions of all of these OSs are multiuser/multitasking operating systems, with support for virtual memory and security, and each comes in editions that support either 32-bit or 64-bit processors.

9.9. Windows Vista

Microsoft released the first retail version of Windows Vista early in 2007. Seen more as an upgrade of Windows XP, it included improvements in how Windows handles graphics, files, and communications. The GUI has a new look compared to previous versions of Windows, and, on computers that can support it, a feature called Aero, which provides translucent windows, live thumbnails, live icons, and other enhancements, further
enhances the GUI. Windows Vista was not widely adopted due to problems with speed, high hardware requirements, and the annoyance of a new security feature, called User Account Control (UAC).

9.10. Windows 7

Released in October 2009 Windows 7 is the best version of Windows to date. It includes several improvements to Windows Vista – correcting the shortcomings that kept Windows Vista from being widely accepted. Windows 7 is faster than Windows Vista in several ways, from starting up, to going into and out of sleep mode, to recognizing new devices when you connect them. Windows 7 has many new features. The short list includes a redesigned desktop with a new taskbar that has many new features of its own, such as jump lists.

9.11. Windows 8

At this writing, the next version of Windows, rumored to be named Windows 8, is projected to come out in late 2011 or early 2012. While all is speculation, the expected new features include improved support for wireless connectivity and other new features such as stereoscopic 3-D, facial detection, USB 3, and hardware acceleration. Experts expect Windows 8 to support virtualization – perhaps with a built-in hypervisor, which is software that provides the simulated hardware of a virtual machine. It may use virtualization of applications to separate running applications from the core of the OS. Applications virtualized remotely on a server could leave the OS with a much smaller footprint on the desktop. While 32- and 64-bit support is a given, some speculate that Windows 8 may go further with 128-bit support. Windows File Systems Depending on the version of Windows you use, you can select from two or three supported file systems when you format a disk. The main file systems supported for hard drives today are FAT32 and NTFS. Windows recognizes several formats for optical drives.

10. Questions

1. What is an operating system?
2. What are the goals of an operating systems?
3. What are the components of a computer system?
4. What does it mean a batch system?
5. Describe a personal computer system!
6. Give examples for distributed computing systems!
4. fejezet - Advanced Spread Sheet System Applications

1. Introduction

The term ‘spreadsheet’ comes from traditional accounting practice. It was used to describe the format used in book-keeping ledgers, in which expenditure categories were arranged as columns, and amounts were added in the relevant columns, with each row representing a transaction. This organisation of rows and columns is carried over into today’s software (Figure 4.1).

4.1. ábra - Figure 4.1: Tables and Rows

Every cell in a spreadsheet is like that programmable calculator! Every cell can contain any formula, of almost any complexity, and reference numbers in other cells. This means that once you have defined your formula, merely changing the numbers in the other cells allows you to freely experiment with your data, instantly. A cell is not restricted to numbers. In fact, a cell can contain any of:

• Text;
• Numbers;
• Logical values (true or false);
• Formulae (that is, calculations), which include references to other cells.

4.2. ábra - Figure 4.2: Exploring Excel

4.3. ábra - Figure 4.3: Features and descriptions
The cell selector allows you to highlight any cell by name. This is often faster than scrolling around a large worksheet. The Excel Window and menu bar houses most of Excel’s commands (Figure 4.2, 4.3).

- File: Saving and printing;
- Edit: Cutting and pasting, filling and clearing cells, deleting cells, columns and rows, searching;
- View: Viewing a document in different ways, enabling and disabling toolbars, headers and footers;
- Insert: Inserting cells, rows, columns, worksheets, charts, functions, objects, diagrams;
- Format: Applying styling and formatting to cell contents, rows and columns, defining styles;
- Tools: Checking spelling, mathematical consistency, protecting cell contents and whole sheets, high-level tools, options;
- Data: Sorting, filtering, grouping, data tables;
- Window: Handling multiple windows;
- plus of course a Help menu.

The formula bar provides you with somewhere to edit the data in a cell when the cell is highlighted, a real help when you are working with long formulae:

- Column titles are alphabetic, starting A, B and so on, through AA, AB to IV, 256 in all.
- The toolbars—Excel’s toolbars, the Standard toolbar and the Formatting toolbar, displayed on the same row. You can display the toolbars in full on two separate rows if you want. Excel’s toolbars, of which there are many, are designed to give you quick access to often-user commands. You can create your own toolbars, too, just as you can with Word.
- The task pane is displayed here because we’ve just done a File ➪ New command.
- The horizontal scroll bar allows you to move forward and backward through the columns in a worksheet.
- The vertical scroll bar (here obscured by the task pane) allows you to scroll a worksheet vertically through its rows.
- The bottom border of the window contains information about the status of the current cell. It displays Enter if you are typing data into a cell, otherwise it displays Ready. Tips are also displayed here when you are engaged in an editing task such as copying a group of cells.
- The worksheet tabs allow you to switch between the different worksheets in a workbook.
- The worksheet selector widgets allow you to scroll the worksheet tabs if necessary.
• Row titles are numeric, from 1 through to 65,536—that’s 16 million cells to a single worksheet!

2. Using Excel formulae

You have already learned that a cell whose contents start with ‘=’ is interpreted by Excel as a formula. Excel will try to calculate the results of any formula it finds and displays the results in the cell (Figure 4.4). But how can you create formulae? Excel offers you four ways:

• Entering a formulae directly into a cell.
• Entering a formula using the formula bar.
• Building up formula by clicking on the cells you wish to include.
• Pasting Excel functions into a formula.

4.4. ábra - Figure 4.4: Spreadsheet formulas

3. Excel functions

Excel functions (Figure 4.5, 4.6) provide the real mathematical power of Excel. A function in Excel is an expression that calls a piece of code dedicated to a specific purpose. For example, the formula: =SUM(A1:A4) calls the function SUM() to return the total of the cells A1, A2, A3 and A4, that is: A1+A2+A3+A4. Excel contains over two hundred functions, which allow you to calculate results that would be far too complex and tedious to program into a worksheet by hand. You can see this if you select Insert Function… in a blank worksheet, then set Or select a category in the Insert Function dialog to All. Many of them you will never use, as they are dedicated to complex mathematical calculations that you are unlikely to encounter—at least, not yet. Some, such as the SUM() function described above, are essential (Figure 4.5). Excel’s functions are grouped by purpose, as the pop-up menu adjacent to the Or select a category field in the Insert Function dialog shows.

4.5. ábra - Figure 4.5: Spreadsheet formulas
3.1. Category Includes

- Database: a set of functions for calculating data from an embedded database, or ‘look up’ list. Excel allows tables of data to be embedded in a worksheet.

- Date and Time Functions: to convert or display anything to do with dates, hours, minutes and seconds, for example NOW(), which returns the current date and time.

- Financial: a set of functions to calculate common financial values, such as the total cost of a loan, the future value of an investment, or the required interest rate for a loan.

- Information: a set of functions that are mainly concerned with returning information about the state of other cells. For example ISBLANK(), which returns FALSE if a cell or range of cells has contents, else TRUE.
• Logical: a set of functions for combining logical expressions, such as AND(), OR(), IF(), and which return the values TRUE or FALSE.

• Lookup and Reference: a set of functions for extracting data from look-up tables within a worksheet, or information about the current cell. Examples of the latter are ROW() and COLUMN(), which return the row and column numbers of the cell containing the current formula (i.e. “What row or column am I in?”).

• Maths and Trig: a set of functions to calculate common mathematical and trigonometrical values, such as sine, tangent, cosine, square root, sum of squares.

• Statistical: a comprehensive set of functions to calculate values used in statistical analysis, such as average, maximum, minimum or n-th largest of a set of numbers, as well as more complex functions such as the Squared, Poisson distribution and Student’s t-distribution tests.

• Text: a set of functions to process text, for example to make one length of text from text in multiple cells, to convert numbers to text, or to convert text to upper or lower case.

Excel displays the Function Arguments dialog. If all is well, it will select the range of cells C3:C7 for you, as Figure 3.1 shows. Note that Excel has already calculated the result of the SUM() function and displayed it in the dialog. The button adjacent to the Number fields allows you to select a range of cells by clicking and dragging. Try it now to see how it works.

6. When you have finished experimenting, click on OK to close the Function Arguments dialog. Entering a function like this might seem a bit long-winded for something as simple as SUM(), but it’s really useful for functions with more, or more complex, arguments, or for functions with which you’re not familiar. Keep this workbook open for the moment, as we’ll add to it in the next step.

3.2. Copying and pasting formulas

Soon you will be entering a set of data for 10 students. But first you must complete the Gradebook Template by creating formulas that will produce a TOTAL or SUM of the scores for EACH student, as well as a PERCENTAGE (a score out of 100) for each student in the class. Before you carry out the exercise, here is a description of what is involved. Fig. 4.13 illustrates the process.

3.3. The order of processing of formulae

When you write formulae in Excel, you need to remember that it has a predefined order of priority for processing mathematical expressions. What we mean is that:

\[=3+4*12\]

in Excel give the answer 51—that is, Excel gives the multiplication a higher priority than the addition, so does it first. So this expression is equivalent to:

\[=3+(4*12)\]

and not:

\[=(3+4)*12\]

which would give the answer 84. Excel uses the following order of priority when executing formulae:

3.4. Priority Operator Description

Highest Colon, comma Cell references, for example ‘C3:C6’

- Negation, for example ‘-1’

% Percentage, for example ‘20%’

\(^\text{Exponentiation}, \text{for example ‘}2^3\text{’ (this means ‘}2\text{ cubed’, i.e. }2^2\times2\text{)}\)
* and / Multiplication and division
+ and - Addition and subtraction
& Join text strings (‘concatenation’)

Lowest = < > <= >= <> Comparison: equal, less than, greater than, less than or equal, greater than or equal, not equal

You can override this order of priority by using brackets. Excel will first evaluate the expression in the innermost pair of brackets, using the priority shown above, then the next pair of brackets, and so on. If it finds two mathematical operators with the same priority, such as multiplication and division, it evaluates the formula from left to right.

### 3.5. Using relative and absolute cell references

We have mentioned relative references above. Now you can see examples for using relative and absolute cell references below.

#### 4.7. ábra - Figure 4.7: Absolute reference

![Absolute Reference Diagram]

#### 4.8. ábra - Figure 4.8: Relative reference

![Relative Reference Diagram]
For example:

C3:C5

Suppose however that you don’t want Excel to do this. Consider the case in which a cell contains a number that you always want Excel to use, no matter how formulae that reference it are copied or moved. Such a value might be something like a currency conversion, or any fixed value you want to use in other calculations. To demonstrate this, we’re going to extend our shopping list so that it displays prices in both pounds sterling and euros:

1. Reopen your shopping list workbook, if it’s not still open from the previous section.
2. First, add the titles ‘Item’, ‘Pounds’ and ‘Euros’ in cells B2 to D2.
3. Drag to select these cells again, then click on the button in the formatting toolbar to set the titles to bold.
4. In cell F2, enter:
   
   Euros per Pound Tab 1.52118 Return
   
   (or substitute the current conversion rate)

5. You can probably only see part of what you typed, as column F will be too narrow to display the entire phrase. Move the mouse cursor over the boundary between the titles for columns F and G, then click and drag to make column F wide enough.

6. Click in cell D3 and enter the following:

   =C3*G2 Return

   This calculates the price of your eggs in Euros, and the result is displayed, which will be €1.901475 if you used the exchange rate of €1.52118/£1.

7. Now click again in cell D3, and drag to copy its contents to cells D4 to D6. Not quite what you expected, maybe? As you can see if you select cells D4, D5 or D6, Excel has changed the reference to cell G2, which contains your conversion rate, to G3, G4 and G5. However, this is not what you want to happen—you want Excel to use the contents of cell G2 for all the conversions. Here’s how to stop this happening…
8. Select cell D3 again. Using the formula bar, change the cell’s contents to:

= C3*$G$2

This form of cell reference, ‘$G$2’, is known as an absolute reference. The ‘$’ signs tell Excel never to change the cell reference, no matter how often it is moved or copied—it will always reference cell G2.

9. Repeat step 7. This time you should get correct results in euros for all your items.

10. To complete this exercise, we’ll visit the formatting dialog to set the decimal spaces of the euro figures to 2.

Select cells D3 to D6, then select Format → Cells… In the Format Cells dialog, select Number. The number of decimal places should default to 2, so just click on OK. We’ll have more to say about cell formatting in Step 4. Before you leave this step, try the following exercises on your own:

• Copy cells C8 to D8 and C10:C11 to D10:D11 to see how Excel handles absolute function references.

• Change the euro conversion rate by changing the value in G2 and watch Excel work for you!

You can use the ‘$’ notation to make either the column, the row, or both, references absolute. For example, a cell reference of ‘$G$2’ would ensure that Excel never changed the column, but could change the row, when such a reference was moved or copied.

When you are editing a formula in the formula bar, the F4 key allows you to toggle between all the combinations of absolute, row-absolute, column absolute and relative references. Excel is usually clever enough to work out which reference to change.

**Using mixed cell references**

We have described how you express a range of cells in Excel. For example, the formula:

=SUM(A1:A4)

is the same as

=SUM(A1,A2,A3,A4)

You might wonder how you express multiple ranges. For example, suppose you wanted to tell Excel to calculate the sum of cells A1 through A4 and D2 through D6? It’s easy—you do it like this:

=SUM(A1:A4, D2:D6)

Try this now for yourself, using a blank worksheet to experiment with. You can select non-adjacent ranges of cells such as this by:

• Clicking and dragging to make the first selection

• Holding down the Ctrl key

• Clicking and dragging to make the second selection.

You can ‘nest’ Excel functions. For example:

=SUM(A1:A12,SUM(B1:B12))

means the same as:

=SUM(A1:A12)+SUM(B1:B12)

The system is copying the formula in cell J10 to cell J14 in relation to ("relative to") cell J10. In other words, just as the formula in J10 sums the values stored in cells D10 through H10, so the formula copied to cell J14
will sum the values relative to cells D14 through H14. Does that make sense? If so, give yourself a pat on the back! If not, don't despair. Read it over a couple of times. The alternative to a Relative Reference, by the way, is an Absolute Reference. You will need to use an Absolute Reference shortly, at which time you will more easily understand what it means in the context of the exercise. Bet you can't wait to check it out!

Press Ctrl-s again to save your work so far (are you getting into the habit of doing this?)

Filling down (copying the formula into the rest of the TOTAL column)

Excel provides a neat tool to duplicate the contents of cells into a set of adjacent cells. For the sake of this exercise we will assume you will have just ten students in your class. You are going to duplicate the formula that is in cell J14 into the other nine cells below it. As before, Excel will automatically adjust the cell addresses so that they are appropriate (relative) to each student's record.

### 3.6. Advanced Excel Tips

**Auto Fill**

Auto fill can be accomplished either by issuing the Edit-Fill-Series command and selecting one of the options from the dialog box that appears. The Edit-Fill-Series dialog box is particularly useful when defining a fill series of dates where the stop date is many cells away. It is also useful when you want to create a trend line using a linear or growth series.

**The Edit - Fill - Series Dialog Box**

Another way to auto-fill a range is to grab the “fill handle” with the mouse pointer and to drag it in the direction that you want to fill. The “fill handle” is the square on the lower right corner of a highlighted range. When you hover over the “fill handle” with the mouse, the cursor changes from a big + to a much thinner -.

**Auto Fill with the Fill Handle**

You can fill in several types of series by selecting cells and dragging the “fill handle.” By dragging the fill handle of a cell, you can copy the contents of a cell to other cells in the same row or column. If a selection contains a number, date, or time period, you can extend the series. For example, if you want to fill in a series of dates where the dates go from the last date of one month to the last date of the next month, simply enter in the first two month ends and then drag the range down with the fill handle.

**AutoFill with the Fill Handle and the Control Key**

The AutoFill feature has built-in defaults that can sometimes bother you. To get the AutoFill to increment a number or not increment a text label with a number, simply press the Control Key along with using AutoFill, and you will get the opposite results. In other words, by pressing the Control Key along with the AutoFill, you are temporarily flipping the default settings opposite what they normally are.

### 4. Working with charts and graphs

Chart is a term that has come to us from America. It tends now to be applied to any graphical representation of data that is not a traditional graph. For example, the terms bar chart and pie chart are very common in business:

- A bar chart represents numerical quantities as vertical or horizontal bars in which the length of the bar represents the size of each value

- A pie chart represents proportions of a whole, or percentages, as slices of a flat cylinder, or ‘pie’, in which the size of the slice represents that quantity’s proportion of the whole.

Excel applies the term series to the data used in the charts and graphs above. This is short for data series, and all it means is a set of related data. For example, your shopping list consists of a list of pairs of information, in this case items’ names and their prices. This is what Excel refers to as a data series. Similarly, the list of numerical values in the graph example is a data series. Each individual data item in a series is referred to as a data point. Excel has no problems plotting a chart or graph with more than one data series.
Creating a simple chart

Let’s see just how easy it is to create a chart. We’re going to produce the pie chart shown on the preceding page, which is based on your shopping list workbook. Excel makes it really easy to do this by providing a Chart Wizard.

Embedded versus floating charts

Whether you choose to embed a graph or chart in a worksheet, or to give it its own worksheet, depends largely on the design of your application and how you intend it to be used:

- Embedding a chart in a worksheet allows you to see the chart change as the data changes. This is useful if you are using the worksheet to analyse data. However, it can obscure your working area unless you are running Excel on a monitor with a large screen. It also means that you cannot easily print just the chart.

- Giving a chart its own worksheet leaves you with more working space on your data worksheet, but means that you cannot see both the data and the chart at the same time. It does make it easy to print just the chart, however.

Creating other types of chart

You saw in the first dialog of the Chart Wizard that Excel can create many different types of chart and graph. Some are purely decorative, others place specific requirements on your data, and do not make sense if the data does not meet those requirements. For example, X/Y scatter plots are not meaningful unless each data point has at least two numerical values. More exotically, ‘bubble’ plots, shown on the right, relate three sets of numerical values, in which the size of the bubble is used to represent the third variable. In the next exercise, we’ll use a bar
graph format to display both the pounds and euros data from your shopping list. This is thus a 2-series chart, as it displays two sets of numerical data:

4.1. Adding legends to your chart or graph

You have already seen how to use the options in the Chart Wizard to add legends to a chart, and how to edit the formats of axes legends. Excel has many options for adding legends to charts. These tend to be specific to the type of chart or graph you are using. These are available from the Chart Options dialog, which differs between different chart and graph types.

4.11. ábra - Figure 4.11: Formatting columns

4.2. Editing charts and graphs

Editing items in charts and graphs is very simple. All you have to do is: Select the item to be edited, right-click and:

• Select the relevant Format option, or
• Select Clear to delete the item, or

Changing a chart’s type

If none of the options for a particular chart type seem to be right, Excel allows you to change the type of a chart even after you have placed and edited it. Try this now:

1. Reopen the shopping list example workbook if it is not already open.
2. Right-click on the embedded pie chart in Sheet1, and select Chart Type…

3. Select a completely different chart type, such as the fifth option under Cylinder.

Click on OK.

4. Use the techniques you learned in the previous section to change the font to Arial Bold.

Click on the chart object to select it, then drag the top border up to make the chart taller.

4.3. Editing a chart’s data range

You may find, once you’ve created a chart or graph that you need to change the source data for the chart. For example, if the data is contained in a table within a worksheet, and you add to it, you will want to include the new data in the chart. Excel makes this easy to do. You have two options, both controlled from the Source Data dialog:

- Redefine the data range by clicking and dragging
- Redefine the data range by directly editing the data series specification. To see how this works, we’ll change the data range on the pie chart in our shopping list workbook:

5. Absolute vs. Relative References Beyond the Basics

5.1. Basics Reviewed

If you are going to copy formula, you need to understand the difference between absolute cell references and relative cell references. For example, if you enter the formula =C5 in cell C6, the reference is read by Excel as “Add what is one cell above the current cell.” Copying this formula to any other cell will result in a formula that adds what is in the cell above the cell that is being copied to. For example, if you copy the formula to cell M90 the formula will read =M89. You can make a cell reference an absolute reference by placing $ symbols in front of the column reference and the row reference. For example, if you enter the formula =$C$5 in cell C6 the formula will remain =$C$5 regardless of where you copy it. As new users quickly learn, understanding how to use absolute references is essential to most accounting and other financial spreadsheets.

5.2. Beyond the Basics

Using the F4 key to Create Absolute References

One approach to making a cell reference absolute is to simply type in the $ symbols. This is time consuming and usually requires the user to type the $ symbol twice. An easier approach is to simply press the F4 key while building the formula. This can be done both when you are typing in a cell reference and when you are using the mouse to build a formula by pointing to the cell reference.

5.3. Partially Absolute Cell References

In some worksheets, it is useful to have the cell reference be “partially absolute.” This means that the cell reference is either absolute with respect to column or relative with respect to row or vice versa.

In the previous screen capture, the formula has been constructed with partially absolute references so that it can be copied both down and across. Unfortunately, Excel does not allow you to use the drag function to copy something both down and across at the same time, so you will have to copy down and then copy across.

Note: When you are building partially absolute references, you can still use the F4 key to insert the $ symbol. You simply press the key more than one time to alter the nature of the reference. For example, if you press the F4 key twice when building a cell reference in a formula, you get a cell reference that is absolute with respect to row but relative with respect to column. If you press the F4 key three times, you get a cell reference that is absolute with respect to column but relative with respect to row.
5.4. Range Name and Sheet Name References

Range names (formally referred to as defined names in Excel) and sheet names are always absolute references and do not need the $ symbol. IMHO was an unwise choice by the designers of Excel because it takes away the option of having these references be either absolute or relative. Having sheet name references always be absolute, for example, prevents you from building a formula that points to a certain cell in the sheet immediately above the current sheet. This technique is useful in building spreadsheet models that have totals that continue from one sheet to the next.

5.5. AutoSum Drop Down Options

In Excel 2002 and Excel 2003, the AutoSum icon (look on the standard toolbar) has a drop down box that allows you to choose other commonly used mathematical functions.

6. Working with Styles

If you have not worked with styles before when using MS Word, you are really in for a treat. In Word, styles are drop down menu selections that can include formatting for font, spacing, borders and underlines, italics, color, and a whole lot more – all by just making one selection. Once you have defined a style (for example, Heading 1), you can later change the attributes of that style and all of the text that is formatted using that particular style will change to the new format. This is actually just a small glimpse into the benefits that styles bring to MS Word, but it’s a good start. In Excel, styles work a bit differently. One style allows you to set the appearance attributes for anything that is entered into a cell, both text and numbers, and also allows you to set the protection attribute of a cell. To bring up the Excel style dialog box, use the Alt – ‘ key sequence or issue the Format – Style command from the main menu.

There are a number of pre-programmed styles in Excel, and you can create your own styles by typing a new name in after Style name: in the top of the dialog box. Then, when you want to use that style, you can simply choose the new style from the list. Rather than having to open the style dialog box every time you want to use a style, you can add the style icon to your formatting toolbar.

This is all very well and good until you start a new spreadsheet and find that the new style you liked so much is missing from the new blank spreadsheet. It is also missing from all your old spreadsheets that were created prior to learning about and creating custom styles. Custom style settings are not global settings but, rather, are embedded in the particular workbook in which you create them. However, there are solutions to both issues of getting your new styles into existing spreadsheets and of wanting to have your styles appear automatically in new blank spreadsheet files.

Let’s tackle the task of getting the custom styles into existing spreadsheet files that you are going to continue to use for some time. Let’s assume that you have created several styles you really like and would like to transfer these styles to an existing file. First, open the file that has the custom styles. Next, open the older file to which you want to transfer the custom files. While you are in the older file, issue the Format – Style… command from the main menu.

Notice that the Style dialog box has a Merge button on the following graphic. You can use this button to open the Merge Styles dialog box. Unlike MS Word, which allows you to browse for files that have your styles, you have to have the other file that has your custom styles in it already open. It is a little cumbersome, but it is a lot easier to merge styles from one file to another than it is to create them all over again with every old file where you want your newly created styles.

What about the problem of wanting to have your styles automatically appear in new blank MS Excel Workbooks? The solution to that issue is to create a new default template that contains your custom styles. Then, every time you start with a new blank file, your custom files will be there waiting for you.

6.1. Changing Defaults for Workbooks and Worksheets

For many of the settings you use frequently, such as formats, macros, sheet names, and ranges, you may want to create a Master Template. The simple way to do this is to start with a blank worksheet and fix up all the items you want to have present in all future worksheets and then save the file to be used as the starting point of all new worksheets.
6.2. Create a File You Always Start With

A simple approach is to create a blank file that has all the settings you would like and to save it to use each time you create a new file. If you save the file with a name like _blank.xls and place it in the default directory, it will be relatively easy to find. Whenever you want to start a new worksheet, you simply open this file, which will be the first file in the default directory. It is probably a good idea to change the file attribute of this file to “read only” so you won’t forget to rename the file before you save it.

6.3. Create a Template to Use for New Workbooks

Excel templates are just like regular Excel files except they have the .xlt file extension rather than the .xls file extension. When you open a template file using the File – New command, Excel assumes that you don’t want to save the new file over the original template file. Excel also assumes that you want the file to have an .xls file extension when it is saved rather than the .xlt extension. The template scenario solves two problems with the simpler approach described previously. First, you don’t have to remember in what directory it is stored or waste time searching around for your template file. It is always available from the File – New menu. Second, you don’t ever make the mistake of saving the file and writing over your startup file because the template setup assumes you don’t want to write over the template.

To setup a template, start by making all the changes you normally have to make each time you create a new file. These alterations might include changing the margins, column widths, or footers to meet your preferences and your businesses standards. Then, save the file as a template in the templates directory.

6.4. Create a Template that is Opened Automatically

To permanently change the default worksheet you get when you start Excel or to start a new file, save your template as a file named book.xlt in the Excel XLStart folder. Excel will use this template as the default model from then on. You can create a similar default file for individual worksheets using the filename sheet.xlt in the same directory.

7. Questions

1. What are the main features of a spreadsheet system?
2. What is the structure of a spreadsheet?
3. Describe the formulas!
4. What are spreadsheet analyses for?
5. What are the spreadsheet functions?
6. What are the relative and absolute references?
7. How can we create charts?
5. fejezet - Database system

1. Introduction

We can find many definitions of this term if we look around the literature and the Web. At one time (in 2008), Wikipedia offered this: “A structured collection of records or data.”. According to other definitions, a database is an organized, machine-readable collection of symbols, to be interpreted as a true account of some enterprise. A database is machine-updatable too, and so must also be a collection of variables. A database is typically available to a community of users, with possibly varying requirements.

The organized, machine-readable collection of symbols is what you “see” if you “look at” a database at a particular point in time. It is to be interpreted as a true account of the enterprise at that point in time. Of course it might happen to be incorrect, incomplete or inaccurate, so perhaps it is better to say that the account is believed to be true.

The alternative view of a database as a collection of variables reflects the fact that the account of the enterprise has to change from time to time, depending on the frequency of change in the details we choose to include in that account.

The suitability of a particular kind of database (such as relational, or object-oriented) might depend to some extent on the requirements of its user(s). When E.F. Codd developed his theory of relational databases (first published in 1969), he sought an approach that would satisfy the widest possible ranges of users and uses. Thus, when designing a relational database, we do so without trying to anticipate specific uses to which it might be put, without building in biases that would favour particular applications. That is perhaps the distinguishing feature of the relational approach, and you should bear it in mind as we explore some of its ramifications.

2. Organized Collection of Symbols

For example, the table in Figure 5.1 shows an organized collection of symbols.

5.1. táblázat - Figure 5.1: An Organized Collection of Symbols

<table>
<thead>
<tr>
<th>StudentId</th>
<th>Name</th>
<th>CourseId</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Anne</td>
<td>C1</td>
</tr>
<tr>
<td>S1</td>
<td>Anne</td>
<td>C2</td>
</tr>
<tr>
<td>S2</td>
<td>Boris</td>
<td>C1</td>
</tr>
<tr>
<td>S3</td>
<td>Cindy</td>
<td>C3</td>
</tr>
</tbody>
</table>

Can you guess what this tabular arrangement of symbols might be trying to tell us? What might it mean, for symbols to appear in the same row? In the same column? In what way might the meaning of the symbols in the very first row (shown in blue) differ from the meaning of those below them?

Do you intuitively guess that the symbols below the first row in the first column are all student identifiers, those in the second column names of students, and those in the third column course identifiers? Do you guess that student S1’s name is Anne? And that Anne is enrolled on courses C1 and C2? And that Cindy is enrolled on neither of those two courses? If so, what features of the organization of the symbols led you to those guesses?

Remember those features. In an informal way, they form the foundation of relational theory. Each of them has a formal counterpart in relational theory, and those formal counterparts are the only constituents of the organized structure that is a relational database.

3. Collection of Variables
Now look at Figure 5.2, a slight revision of Figure 5.1. ENROLMENT

**5.2. táblázat - Figure 5.2: A variable, showing its current value**

<table>
<thead>
<tr>
<th>StudentId</th>
<th>Name</th>
<th>CourseId</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Anne</td>
<td>C1</td>
</tr>
<tr>
<td>S1</td>
<td>Anne</td>
<td>C2</td>
</tr>
<tr>
<td>S2</td>
<td>Boris</td>
<td>C1</td>
</tr>
<tr>
<td>S3</td>
<td>Cindy</td>
<td>C3</td>
</tr>
<tr>
<td>S4</td>
<td>Devinder</td>
<td>C1</td>
</tr>
</tbody>
</table>

We have added the name, ENROLMENT, above the table, and we have added an extra row.

ENROLMENT is a variable. Perhaps the table we saw earlier was once its value. If so, it (the variable) has been updated since then the row for S4 has been added. Our interpretation of Figure 5.1 now has to be revised to include the sentence represented by that additional row:

Student S1, named Anne, is enrolled on course C1. Student S1, named Anne, is enrolled on course C2. Student S2, named Boris, is enrolled on course C1. Student S3, named Cindy, is enrolled on course C3. Student S4, named Devinder, is enrolled on course C1.

Notice that in English we can join all these sentences together to form a single sentence, using conjunctions like “and”, “or”, “because” and so on. If we join them using “and” in particular, we get a single sentence that is logically equivalent to the given set of sentences in the sense that it is true if each one of them is true (and false if any one of them is false). A database, then, can be thought of as a representation of an account of the enterprise expressed as a single sentence! (But it’s more usual to think in terms of a collection of individual sentences.) We might also be able to conclude that the following sentences (for example) are false: Student S2, named Boris, is enrolled on course C2. Student S2, named Beth, is enrolled on course C1.

Whenever the variable is updated, the set of true sentences represented by its value changes in some way. Updates usually reflect perceived changes in the enterprise, affecting our beliefs about it and therefore our account of it.

### 4. The Relational Database

A relational database is one whose symbols are organized into a collection of relations. Figure 5.3 confirms that the examples we have already seen are in fact relations, depicted in tabular form. Indeed, according to Figure 5.2, the relation depicted in Figure 5.3 is the current value of the variable ENROLMENT.

**5.3. táblázat - Figure 5.3: A relation, shown in tabular form**

<table>
<thead>
<tr>
<th>StudentId</th>
<th>Name</th>
<th>CourseId</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Anne</td>
<td>C1</td>
</tr>
<tr>
<td>S1</td>
<td>Anne</td>
<td>C2</td>
</tr>
<tr>
<td>S2</td>
<td>Boris</td>
<td>C1</td>
</tr>
<tr>
<td>S3</td>
<td>Cindy</td>
<td>C3</td>
</tr>
</tbody>
</table>
Happily, the visual (tabular) representation we have been using thus far is suited particularly well to relational databases: so much so that many people use the word table as an alternative to relation. The language SQL in particular uses that term, so in the context of relational theory it is convenient and judicious to stick with relation for the theoretical construct, allowing SQL’s deviations from relational theory to be noted as differences between tables and relations.

Relation is a formal term in mathematics in particular, in the logical foundation of mathematics. It appeals to the notion of relationships between things. Most mathematical texts focus on relations involving things taken in pairs but our example shows a relation involving things taken three at a time and, as we shall see, relations in general can relate any number of things (and, as we shall see, the number in question can even be less than two, making the term relation seem somewhat inappropriate). Relational database theory is built around the concept of a relation. Our study of the theory will include.

The “anatomy” of a relation:-

- Relational algebra: a set of mathematical operators that operate on relations and yield relations as results.
- Relation variables: their creation and destruction, and operators for updating them.
- Relational comparison operators, allowing consistency rules to be expressed as constraints (commonly called integrity constraints) on the variables constituting the database.
- And we will see how these, and other constructs, can form the basis of a database language (specifically, a relational database language).

5. “Relation” Not Equal to “Table”

“Table”, here, refers to pictures of the kind shown in Figures 5.1, 5.2, and 5.3. The terms relation and table are not synonymous. For one thing, although every relation can be depicted as a table, not every table is a representation of (i.e., denotes) some relation. For another, several different tables can all represent the same relation. Consider Figure 5.4, for example.

5.4. táblázat - Figure 5.4: Same relation as Figure 5.3.

<table>
<thead>
<tr>
<th>Name</th>
<th>StudentId</th>
<th>CourseId</th>
</tr>
</thead>
<tbody>
<tr>
<td>Devinder</td>
<td>S4</td>
<td>C1</td>
</tr>
<tr>
<td>Cindy</td>
<td>S3</td>
<td>C3</td>
</tr>
<tr>
<td>Anne</td>
<td>S1</td>
<td>C1</td>
</tr>
<tr>
<td>Boris</td>
<td>S2</td>
<td>C1</td>
</tr>
<tr>
<td>Anne</td>
<td>S1</td>
<td>C2</td>
</tr>
</tbody>
</table>

The table in Figure 5.4 is different from the one in Figure 5.3, but it represents the same relation. I have changed the order of the columns and the order of the rows, each green row in Figure 5.4 has the same symbols for each column heading as some row in Figure 5.3 and each row in Figure 5.3 has a corresponding row, derived in that way, in Figure 5.4. What I am trying to illustrate is the principle that the relation represented by a table does not depend on the order in which we place the rows or the columns in that table. It follows that several different tables can all denote the same relation, because we can simply change the left-to-right order in which the columns are shown and/or the top-to-bottom order in which the rows are shown and yet still be depicting the same relation.
What does it mean to say that the order of columns and the order of rows doesn’t matter? We will find out the answer to this question when we later study the typical operators that are defined for operating on relations (e.g., to compute results of queries against the database) and relation variables (e.g., to update the database). None of these operators will depend on the notion of some row or some column being the first or last, or immediately before or after some other column or row.

We can also observe that not every table depicts a relation. Such tables can easily be obtained just by deleting the blue rows (the column headings) from each of Figures 5.1 to 5.4. Figure 5.5 shows another table that does not depict any relation.

5.5. táblázat - Figure 5.5: Not a relation

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>?</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

The various reasons why this table cannot be depicting a relation should become apparent to you by the time you reach the end of this chapter.

6. Anatomy of a Relation

Figure 5.6 shows the terminology we use to refer to parts of the structure of a relation.

5.1. ábra - Figure 5.6: Anatomy of a relation
Because of the distinction I have noted between the terms relation and table, we prefer not to use the terminology of tables for the anatomical parts of a relation. We use instead the terms proposed by E.F. Codd, the researcher who first proposed relational theory as a basis for database technology, in 1969.

The degree is the number of attributes. The cardinality is the number of tuples. The heading is the set of attributes (note set, because the attributes are not ordered in any way and no attribute appears more than once). The body is the set of tuples (again, note set — the tuples are not ordered and no tuple appears more than once). An attribute has an attribute name, and no two have the same name. Each attribute has an attribute value in each tuple.

7. The DBMS

A database management system (DBMS) is exactly what its name suggests a piece of software for managing databases and providing access to them. But be warned in the industry the term database is commonly used to refer to a DBMS, especially in promotional literature. You are strongly discouraged from adopting such sloppy practice (if such a system is a database, what are the things it manages?).

A DBMS responds to commands given by application programs, custom-written or general-purpose, executing on behalf of users. Commands are written in the database language of the DBMS (e.g., SQL). Responses include completion codes, messages and results of queries.

In order to support multiple concurrent users a DBMS normally operates as a server. Its immediate users are thus those application programs, running as clients of this server, typically (though not necessarily) on behalf of end users. Thus, some kind of communication protocol is needed for the transmission of commands and responses between client and server. Before submitting commands to the server a client application program must first establish a connection to it, thus initiating a session, which typically lasts until the client explicitly asks for it to be terminated. That is all you need to know about client-server architecture as far as this book is concerned.

This book is concerned with relational DBMSs and relational databases in particular, and soon we will be looking at the components we expect to find in a relational DBMS. Before that we need to briefly review what is expected of a DBMS in general.

8. Database Language

To repeat, the commands given to a DBMS by an application are written in the database language of the DBMS. The term data sublanguage is sometimes used instead of database language. The sub-prefix refers to the fact that application programs are sometimes written in some more general-purpose programming language (the “host” language), in which the database language commands are embedded in some prescribed style. Sometimes the embedding style is such that the embedded statements are unrecognized by the host language compiler or interpreter, and some special preprocessor is used to replace the embedded statements by, for example, CALL statements in the host language.

A query is an expression that, when evaluated, yields some result derived from the database. Queries are what make databases useful. Note that a query is not of itself a command. The DBMS might support some kind of command to evaluate a given query and make the result available for access, also using DBMS commands, by the application program. The application program might execute such commands in order to display a query result (usually in tabular form) in a window.

In response to requests from application programs, we expect a DBMS to be able, for example, to

• create and destroy variables in the database,
• take note of integrity rules (constraints),
• take note of authorisations (who is allowed to do what, to what),
• update variables (honouring constraints and authorisations),
• provide results of queries.
The requests take the form of commands written in the database language supported by the DBMS. The variables are the constituents of the database, like the ENROLMENT variable we looked at earlier. Such variables are both persistent and global. A persistent variable is one that ceases to exist only when its destruction is explicitly requested by some user. A global variable is one that exists independently of the application programs that use it, distinguishing it from a local variable, declared within the application program and automatically destroyed when the program unit ("block") in which it is declared finishes its execution.

Constraints (sometimes called integrity constraints) are rules governing permissible values, and permissible combinations of values, of the variables. For example, it might be possible to tell the DBMS that no student’s assessment score can be less than zero. A database that violates a constraint is, by definition, incorrect it represents an account that is in some respect false. A database that satisfies all its constraints is said to be consistent, even though it cannot in general be guaranteed to be correct.

In the sense that constraints are for integrity, authorisations are for security. Some of the data in a database might represent sensitive information whose accessibility is restricted to certain privileged users only. Similarly, it might be desired to allow some users to access certain parts of the database without also being able to update those parts.

Note the three parts of an authorisation: who, what, and to what. “Who” is a user of the database; “what” is one of the operations that are available for operating on the variables in the database; “to what” is one of those variables.

9. Relational Algebra – The Foundation

Sometimes that term, relational algebra, is used with the definite article: the relational algebra, even though several minor variations exist in the literature. Indeed, the term relational completeness is sometimes defined with reference to “the” relational algebra a language is deemed relationally complete if it supports, directly or indirectly, all of the operators of “that” algebra. The operators of the relational algebra are nearly all relational counterparts of the logical connectives AND, OR, and NOT, and existential quantification. Each relational operator, when invoked, yields a relation, which can be interpreted as the extension of some predicate. Because relations are used to represent predicates, it makes sense for relational operators to be counterparts of logical operators.

10. Access-2007

The primary function of Microsoft Office Access 2007 is an information management program. Information is stored in separate lists called tables, and information in one table may relate to information in one or more other tables. These groups of information, when considered together as a whole, become a database.

Access is designed to use the data in these databases to extract the information relevant to your situation. Access can also generate reports (such as quarterly sales by each employee) based on the data contained in the database. The Office 2007 package also features a lot of interconnectivity between the various programs, including a newly designed SharePoint service that lets users in your organization connect and share information using a special data centre via the Internet.

10.1. Basic Terminology

Let’s take a look at the terminology used in database-speak, starting with the basics. Let’s look at each piece of the database.

A field is the smallest piece of a database; that is, one specific piece of information like a number, a word, a date, a picture, or a reference for some other piece of data. Each column you see in the diagram would all be the same data type; that is, one column of data would all be numbers.

A record is a collection of one or more fields together in a row. (In a real database, you would not count the word ‘Record’ as depicted in the diagram – this is just to help visualize the concept.)

A table is comprised of one or more records. Each table also has a unique name.

A database is comprised of one or more tables. Each database is also given a unique name.
A form is a tool that is used to easily and accurately enter data into a table. A form presents one record of a database at a time to a user, or allows a user to enter data into the database one record at a time.

A query is just like a question you ask the database. There are two types of queries: select and action. A select query will extract and display data based on criteria you provide. An action query will find all data relevant to your query and perform some sort of operation on it. A query can be performed on one or more tables in a database.

Queries are primarily built from tables, but Access gives you the ability to construct a query based on the results of another query (Figure 5.7: Figure 5.8). Such ‘nested queries’ may require more computer memory and resources in order to execute but if constructed with care, can save a lot of time, especially when dealing with very large databases. For the purposes of this manual, we will keep things simple and stick to small and simple queries. Plus, the great thing about queries is that they are only questions asked about data that is already there. If you get query results that are completely off the mark, no problem! The data is untouched, so provided there is no design flaw in your database, only the query needs to be adjusted.

5.2. ábra - Figure 5.7: Creating Table

5.3. ábra - Figure 5.8: Creating criteria for Query
A report presents the data found by a query. A report can be formatted to show summaries, calculations, charts, and more based on the data returned by a query. Access takes the report one step further by letting you organize and format a report into a sleek, professional document suitable for printing, exporting, or e-mailing.

10.2. Interface Overview

In this section, we will learn about the Access starting screen and the view of a typical working database. We will introduce the views piece by piece in this section of the lesson. There are a large number of updates to learn about, but with time you will wonder how you ever managed without them!

If you have ever used Access before, the welcome screen for Microsoft Access 2007 has been completely redesigned. However, the layout is much easier to use, especially if you have never used Access before. Exploring the different parts of the Access 2007 Getting Started interface.

10.2.1. Template Pane

On the left-hand side of the Access window is the Template pane: Access has a number of templates built right into the program. To access the different categories of templates, simply click a category to see the available template files.

10.2.2. Recent Files

The right-hand side of the window lists any recently opened database files, just like the Office Menu. Click one of the database files to open the file. If you want to open a database file stored somewhere else on your computer or on another network, click the More ulink and browse to the file you want to open, and then click the Open button.

10.2.3. New Blank Database

In the centre of the window is a ulink to create a new Blank Database. Use this ulink to make your own database from scratch.
10.2.4. Microsoft Office Online

The centre of the Access window is a special page that extracts content from Microsoft Office Online (a service provided over the Internet). Microsoft Office Online provides quick ulinks to different templates, training material, and other downloads. It also provides ulinks directly to Office Online where you can read about updates to Office 2007 as they become available.

10.2.5. Status bar

Finally, at the very bottom of the Access window is the status bar. This bar will give information about the status of Access, if any particular lock keys are enabled on your keyboard, which view is currently active, and more.

10.2.6. Command Tabs

Along the top of the window are the command tabs. In the past, the Office package made use of menus that contained a listing of commands. At their core, the command tabs are essentially the same thing as menus but with a few big changes. For starters, the grouping of commands in tabs is much more intuitive. The commands listed under each tab are also the only commands that are applicable to your current view of the database.

Access 2007 takes this one step further with the addition of contextual tabs. The tab labelled Table Tools – Datasheet appears only when you have selected a table in Datasheet view. This tab will contain even more specific commands that can be used on a table being viewed in Datasheet view and will only be visible when a table is being viewed in Datasheet view (Figure 5.9).

5.4. ábra - Figure 5.9: Datasheet View

10.2.7. Ribbons

Consider the Home ribbon tab that is selected in the diagram above. Beneath the tab is a listing of all commands that are performed most often on the currently selected object, contained in what Microsoft refers to as the ‘ribbon’.
The ribbon was designed to allow access to all functionality of a tab at once. Also, the commands in the ribbon are only the commands that are available for use at the time.

### 10.3. Navigation Pane

On the right side of the Access window is the Navigation Pane. It is always visible on the left side of the screen, but can be expanded () or shrunk () by clicking the double arrows. The Navigation Pane allows quick and easy access to any of the database objects.

Click the pull-down arrow beside the Navigation Pane title () to show a list of all object categories.

### 10.4. Object Tabs

In previous versions of Access, any open database object was opened in its own window and designed to ‘float’ inside the Access Screen. When several database objects were open at once, it was difficult to navigate through all of the windows easily. Access 2007 has solved that problem by using tabs.

Simply click any of the tabs visible on the top to show the database object. Opening many database objects will create left and right facing arrows (and); click on the arrow to scroll that direction through the open database objects. If you want to close an object you are no longer using, click the Close button () located beside the tabs.

#### 10.4.1. Help Button

The Help button, located directly under the title bar, launches the Access help screen.

Click a topic to view help about that particular subject.

As we explore more of the features and functionality of Access, we will discover how to use the rest of the interface.

#### 10.4.2. Closing Microsoft Access

When you have finished using Access, click either Office Menu Exit Access or click the program’s close button () in the upper-right hand side of the Access window. If you have any unsaved work still open, Access will allow you to save any changes you have made before the program shuts down.

### 11. Making Database

Making a database might seem like a pretty big job, but taking the time to design one properly will save a lot of time down the road. You are exposed to databases everyday use them all the time probably without knowing it. In fact, you are likely in several, yourself!

The easiest method of identifying yourself in day to day life is a simple handshake and saying “Hello, my name is...” But you can’t really shake hands with a computer. Using your name, even your full name, isn’t a very good option either because there may be hundreds of people out there with exactly the same name as you. Therefore, you must be assigned some unique identifier, the most recognizable being your Social Security Number (SSN) or Social Insurance Number (SIN). No one else in the country has the same SSN as you.

This practice holds true for databases, too. Earlier in this manual you may recall seeing the term ‘primary key’. Every row in a table should have at least one field that is unique from every other record. That field is usually a number, and the unique field is referred to as the primary key. It is not imperative to have a primary key, but it makes the design of the database much easier and eliminates the possibility of duplicate data (which does nothing but confuse the issue)! It also allows a database program to (in most cases) search faster and more efficiently. Therefore, it is good practice to have a primary key for every table you make.

Let’s quickly review what we know about databases: they are made up of tables, and in each table are several records (or rows) of data. Every record is made up of one or more fields, and every record in a table is different from every other record because of the unique primary key. Knowing this, and with the knowledge of the commands we learned so far, we are ready to start making databases!
For the remainder of this manual, let’s pretend that you are Bugs Rabbit, CEO of an upstart animation company, Warner Cousins. You want to use Access 2007 to monitor the expenses made by you and your employees.

### 11.1. Planning a Database

Before you start using Access to create a database, take the time to answer a few questions:

- Why do you need a database? You want to keep track of the expenses made by you and your employees.
- Who will be using the database? Any employee of Warner Cousins will have access to this database.
- What kind of data would be extracted from the database? Total expenses of the company, total expenses by each employee, expenses by each category.

Once you have answered these questions, it is time to decide how to design the tables for your database. What fields of data do you need? What data types will the fields need to be? What tables would be important? Which fields will go in which table, and do the placements make sense?

Next comes the planning of relationships between the data. A big list of numbers doesn’t mean much by itself, but when constructed based on other data, it becomes meaningful. And finally, make sure that you talk to everyone who will be using the database will be able to get the data they need. Let’s examine some of the details.

You will obviously need an expense table that contains at least the following: who made the purchase, what did they purchase, how much was it, and when did they purchase it?

The payroll department already has a listing of the people who work for you:

- SIN (or company ID #),
- Name,
- Address,
- Phone Number,
- Company Position.

The database now should have two tables: an expenses table and an employee table. Now, there needs to be some sort of link between the two tables. You could use the name of each person, but that may become confusing, especially if your company grows into the hundreds. There is another option, however. You can use the SIN (or company ID) of each employee to tie their purchase to their personal information.

In database design, your most powerful tool is not the computer, but rather a piece of paper and a pencil (and a big eraser). Not only can you easily change the information you might need, but you can also visualize the information.

### 11.2. Tables

#### 11.2.1. The Record

A record is defined as one or more fields of data that create a single entry in a table. We have also learned that each record should have a primary key; that is, some unique identifier that sets it apart from every other record in a table.

You should be very familiar with the components of tables by now. We know a table is made up of several records each containing fields with data. Access also makes it easy to build and modify any component of a table using Design view, which we will cover later in this manual.

When designing a database, it is critical that you take the time to design your database carefully. Although it is not a difficult job to make some adjustments to a field, adding or removing fields in a large established database can be a real headache. It is important to communicate with everyone who will be using the database to make
sure that everyone has the information they will need. Don’t be afraid to build a database a little bigger than you think it needs to be; if you end up with unused fields they are much easier to take out than to put new fields in (Figure 5.10).

5.5. ábra - Figure 5.10: Creating Table

11.2.2. Empty Table

• Click the Table command to open a new empty table. A new tab will open, containing an empty table in Datasheet view.

• Click inside the Add New Field column and start entering data. Press Enter to keep adding fields to the record, then press Tab or click the command to make a new record.

11.2.3. Table from Template

• Click the small pull-down arrow beside the Table Templates command to see a short list of available templates.

• Click a template from the list; it will open a new empty table in the main part of the Access window.

• Press Enter on the keyboard to advance to the next field, then press Tab once you have reached the last field or to make a new record.

11.2.4. Table Design View

• Click the Table Design command to open a new blank table. A new section of Access we have yet to explore will appear: Design view for a table.

• Design view includes its own Design ribbon in a contextual tab. You have the ability to add a primary key, construct custom formulas, insert or delete different fields, and more.
• Using Design view is more in-depth than simply entering data into fields. You can specify the field name, its data type, and give the field some sort of description if you like.

• At the bottom of Design view is the Field Properties section. Here you can modify all of the properties of a particular field.

• For example, if you want to have a Price field in your database.

• Give the field a name, choose a data type for the field.

• A data type can be a word, number, currency, date, time, etc.

• The properties of the Price field (once defined as a number) include how large a price it can be, the number of decimal places, if the field should contain a default value (like $5.99), and more. As we use tables more we will explore more of the details regarding Field Properties.

11.2.5. Creating a Database from a Template

When you launch Access 2007, you will see the Getting Started page. From here you can choose from a number of different templates already built into Access. Choose a category on the left side of the screen. Then choose a template that best suits your needs from the centre of the window (Figure 5.11).

5.6. ábra - Figure 5.11: Creating Database from Template

Once you have chosen a template, choose a save location (default of My Documents) and then click Download/Open (depending on if you are opening from an online or offline template).

The template will open containing a number of pre-built database objects, including tables and relationships between the tables. Start entering data or modify the design of the objects as you see fit.

Although Access contains a number of templates already built in, it is important to understand how to create a database from scratch.

• From the Getting Started page, click the Blank Database ulink in the centre of the Access window.

• On the right side of the Access window, select a location (default of My Documents) to save the database and click Create.
• A new, blank database will appear in the Access window.

11.3. Using the Quick Access Toolbar

We introduced the new layout changes to Access 2007. In this section, we will learn a little bit more about each part of the new interface and how it works. This lesson will focus on features and customization options available with the Quick Access toolbar, located in the upper left-hand corner of the screen.

11.3.1. About the Default Buttons

Access features three default commands in the Quick Access toolbar:

Save

![Save button](image)

Saves the most recent changes to the current database file.

Print

![Print button](image)

Opens the Print dialogue box allowing you to adjust different print settings.

Undo

![Undo button](image)

The Undo command will revert most changes made in Access. For example, if you made a formatting change to a form that you were not happy with, click the Undo button to go back one command.

11.4. Using Database Objects

A database object is defined as some individual piece of a database that can be used on its own. We have discussed the major objects: tables, queries, forms, reports, and macros (Figure 5.12).

5.7. ábra - Figure 5.12: Access Objects

![Diagram of Access Objects](image)

The Navigation Pane is used to control and use the objects of an Access 2007 database.

• Expand the Navigation Pane () and click the pull-down arrow beside the title to show the full Navigation Pane toolbar.

• Then, click Object Type to display all objects currently in the database.
• All objects currently in the database are categorized by their object type.

• If you want to see the different objects in each category, click to expand that category. Each object contained in each category is listed in alphabetical order.

To open an object, simply double-click it. It will open in the main part of the Access window and will have its own identifying tab. Some objects, such as the report, include a time and date stamp right on the object. Access lets you rename or delete objects in your database. However in order to do so, the object must first be closed.

11.5. Report Wizard

11.5.1. Creating a Report with the Wizard

Many of the reports you create will simply be an exercise in displaying the data in a certain way. Since reports are made from queries, and most of the queries will have already been built, creating reports using the Wizard is easy (Figure 5.13).

5.8. ábra - Figure 5.13: Using Wizard in Creating Report

• The Report Wizard command can be found in the Create ribbon.

• The first page of the Report Wizard should be pretty familiar to you by now; it was used to create a form and a query.

• For this example, we will make a report based on the full results from the Customers Extended query. The next screen of the Report Wizard allows you to apply levels of grouping to the report.

• Grouping levels are useful in certain queries to help categorize the data returned from a query. For example, if you ran a query to list all the different times that a product was ordered, you could group based on the product. Each date the product was sold would then be categorized under each product name. For the purpose of this example, no grouping will be used. The next page of the Wizard lets you organize fields in the report in ascending or descending order.

• Select a field from the combo box. If you want to sort based on descending order, click the Ascending button to change the nature of the sort order.

• The Wizard then asks how you want to organize the items in your report.

11.5.2. Using Design View to Modify a Report
Like forms and queries, you can enter report Design view (Figure 5.14).

5.9. ábra - Figure 5.14: Modifying Report Layout

In the report layout options, select:
- Layout, such as Tabular
- Orientation, such as Landscape
- Verify that the Adjust field width to fit on a page option is checked.
- Preview the report and then close the Print Preview.
- Select Layout View to edit the report.

- Click the “Modify the report’s design” radio button before closing the wizard.
- Click the “Modify the report’s design” radio button before closing the wizard.
- Or Use the View menu after opening a report.
- Report Design view lets you drag and drop the various fields from the Field List pane. Reports use headers and footers like the Design view of a form. Reports also have three of their own contextual tabs.
- The Design and Layout tabs contain the same commands as the Design view of forms. In addition to listing only query results, you can add interactivity to the report to do things like show charts and calculate data values from user input.
- Design view for reports also features a Page Setup ribbon to customize how the report will look on a printed page.
- This section of the ribbon also contains a Page Layout button (the arrow in the bottom right hand corner). This button opens a dialogue that can be used to precisely modify page settings.

11.5.3. Using Reports

To view a report, simply double-click its object name in the Navigation Pane. The report will open in the main part of the Access window.

This Report View will let you scroll through all the details of the report. We will discuss how to print and further view a report later in this manual.

11.5.4. Common Report Tasks

As all the pieces of your report begin to come together, you can apply the formatting and ensure that the report gives you the information you need to know. Then your report will be ready to publish and print as handouts or catalogues. In the final lesson of this section, we will discuss how to give a report some extra flair to effectively present your product or data.

11.5.5. Adding a Photo

Adding a photo to a report is just like adding any other control to a report. To add a photo, click the Image command in the Report Tools-Design ribbon and then click and drag somewhere in the appropriate section you want the photo to appear. Navigate your computer to find the picture file you want to insert into the report, and then click OK. The image will be inserted as a best fit into the area you specified.
11.5.6. Adjusting Page Properties

Access 2007 features a number of page formatting options. Click the Report Tools-Page Setup tab to see the most common commands available for use. You can also click the Page Setup button to see extra commands.

11.5.7. Print Options Tab

Adjust the size of the margins for your page. If you would prefer to print only the data and not any logos or pictures, click the Print Data Only check box.

11.5.8. Page Tab

The Page Tab allows you to adjust the page orientation (portrait or landscape) as well as the size of paper you can print with using your current printer.

11.5.9. Columns Tab

The Columns tab is used if you want to print two or more pages of a report on one piece of paper. The number of columns, row spacing, and column spacing fields allow you to specify the dimensions between the multiple pages on your report.

The column size fields specify how large you would like each page of the report to be on the printed page. You can also check the Same as Detail checkbox to make the printed size the same as the current dimensions of the Detail section.

Lastly, you can choose how the layout of the report pages will be ordered by choosing one of the radio buttons. (The Column Layout control group is only active when you have two or more columns.)

12. Questions

1. What are the file organization concepts?
2. What is a relational database?
3. What is the structure of the relational database tables?
4. Give some example for operations on database tables!
5. What are the capabilities of a DBMS?
6. What are the main objects in the MS Access?
6. fejezet - Computer Networks

1. Introduction

Many companies have a substantial number of computers. For example, a company may have separate computers to monitor production, keep track of inventories, and do the payroll. Initially, each of these computers may have worked in isolation from the others, but at some point, management may have decided to connect them to be able to extract and correlate information about the entire company.

In the simplest of terms, one can imagine a company's information system as consisting of one or more databases and some number of employees who need to access them remotely. In this model, the data are stored on powerful computers called servers. Often these are centrally housed and maintained by a system administrator. In contrast, the employees have simpler machines, called clients, on their desks, with which they access remote data, for example, to include in spreadsheets they are constructing. (Sometimes we will refer to the human user of the client machine as the "client", but it should be clear from the context whether we mean the computer or its user.)

The client and server machines are connected by a network, as illustrated in Figure 6.1. Note that we have shown the network as a simple oval, without any detail. We will use this form when we mean a network in the abstract sense. When more detail is required, it will be provided.

6.1. ábra - Figure 6.1: A network with two clients and one server

This whole arrangement is called the client-server model. It is widely used and forms the basis of much network usage. It is applicable when the client and server are both in the same building (e.g., belong to the same company), but also when they are far apart. For example, when a person at home accesses a page on the World Wide Web, the same model is employed, with the remote Web server being the server and the user's personal computer being the client. Under most conditions, one server can handle a large number of clients.

If we look at the client-server model in detail, we see that two processes are involved, one on the client machine and one on the server machine. Communication takes the form of the client process sending a message over the network to the server process. The client process then waits for a reply message. When the server process gets the request, it performs the requested work or looks up the requested data and sends back a reply. These messages are shown in Figure 6.2.

6.2. ábra - Figure 6.2: The client-server model involves requests and replies
A second goal of setting up a computer network has to do with people rather than information or even computers. A computer network can provide a powerful communication medium among employees. Virtually every company that has two or more computers now has e-mail (electronic mail), which employees generally use for a great deal of daily communication. In fact, a common gripe around the water cooler is how much e-mail everyone has to deal with, much of it meaningless because bosses have discovered that they can send the same (often content-free) message to all their subordinates at the push of a button.

But e-mail is not the only form of improved communication made possible by computer networks. With a network, it is easy for two or more people who work far apart to write a report together. When one worker makes a change to an online document, the others can see the change immediately, instead of waiting several days for a letter. Such a speedup makes cooperation among far-flung groups of people easy where it previously had been impossible.

Yet another form of computer-assisted communication is videoconferencing. Using this technology, employees at distant locations can hold a meeting, seeing and hearing each other and even writing on a shared virtual blackboard. Videoconferencing is a powerful tool for eliminating the cost and time previously devoted to travel. It is sometimes said that communication and transportation are having a race, and whichever wins will make the other obsolete.

A third goal for increasingly many companies is doing business electronically with other companies, especially suppliers and customers. For example, manufacturers of automobiles, aircraft, and computers, among others, buy subsystems from a variety of suppliers and then assemble the parts. Using computer networks, manufacturers can place orders electronically as needed. Being able to place orders in real time (i.e., as needed) reduces the need for large inventories and enhances efficiency.

A fourth goal that is starting to become more important is doing business with consumers over the Internet. Airlines, bookstores, and music vendors have discovered that many customers like the convenience of shopping from home. Consequently, many companies provide catalogs of their goods and services online and take orders online. This sector is expected to grow quickly in the future. It is called e-commerce (electronic commerce).

Some of the more popular uses of the Internet for home users are as follows:

• Access to remote information;
• Person-to-person communication;
• Interactive entertainment;
• Electronic commerce.

Access to remote information comes in many forms. It can be surfing the World Wide Web for information or just for fun. Information available includes the arts, business, cooking, government, health, history, hobbies, recreation, science, sports, travel, and many others. Fun comes in too many ways to mention, plus some ways that are better left unmentioned.

Many newspapers have gone on-line and can be personalized. The next step beyond newspapers (plus magazines and scientific journals) is the on-line digital library. Depending on the cost, size, and weight of book-sized notebook computers, printed books may become obsolete. Skeptics should take note of the effect the printing press had on the medieval illuminated manuscript.

Another type of person-to-person communication often goes by the name of peer-to-peer communication, to distinguish it from the client-server model. In this form, individuals who form a loose group can communicate with others in the group, as shown in Figure 6.3. Every person can, in principle, communicate with one or more other people; there is no fixed division into clients and servers.

6.3. ábra - Figure 6.3: In a peer-to-peer system there are no fixed clients and servers
Other communication-oriented applications include using the Internet to carry telephone calls, video phone, and Internet radio, three rapidly growing areas. Another application is telelearning, meaning attending 8 A.M. classes without the inconvenience of having to get out of bed first. In the long run, the use of networks to enhance human-to-human communication may prove more important than any of the others.

Our third category is entertainment, which is a huge and growing industry. The killer application here (the one that may drive all the rest) is video on demand. A decade or so hence, it may be possible to select any movie or television program ever made, in any country, and have it displayed on your screen instantly. Already we have multiperson real-time simulation games, like hide-and-seek in a virtual dungeon, and flight simulators with the players on one team trying to shoot down the players on the opposing team. If games are played with goggles and three-dimensional real-time, photographic-quality moving images, we have a kind of worldwide shared virtual reality.

Our fourth category is electronic commerce in the broadest sense of the term. Home shopping is already popular and enables users to inspect the on-line catalogs of thousands of companies. Some of these catalogs will soon provide the ability to get an instant video on any product by just clicking on the product's name. After the customer buys a product electronically but cannot figure out how to use it, on-line technical support may be consulted.

Another area in which e-commerce is already happening is access to financial institutions. Many people already pay their bills, manage their bank accounts, and handle their investments electronically. This will surely grow as networks become more secure.

Computer networks may become hugely important to people who are geographically challenged, giving them the same access to services as people living in the middle of a big city. Telelearning may radically affect education; universities may go national or international. Telemedicine is only now starting to catch on (e.g., remote patient monitoring) but may become much more important. But the killer application may be something mundane, like using the webcam in your refrigerator to see if you have to buy milk on the way home from work.

Mobile computers, such as notebook computers and personal digital assistants (PDAs), are one of the fastest-growing segments of the computer industry. Many owners of these computers have desktop machines back at the office and want to be connected to their home base even when away from home or en route. Since having a wired connection is impossible in cars and airplanes, there is a lot of interest in wireless networks. In this section we will briefly look at some of the uses of wireless networks.

Wireless networks are of great value to fleets of trucks, taxis, delivery vehicles, and repairpersons for keeping in contact with home. For example, in many cities, taxi drivers are independent businessmen, rather than being employees of a taxi company. In some of these cities, the taxis have a display the driver can see. When a customer calls up, a central dispatcher types in the pickup and destination points. This information is displayed on the drivers' displays and a beep sounds. The first driver to hit a button on the display gets the call.

Wireless networks are also important to the military. If you have to be able to fight a war anywhere on earth on short notice, counting on using the local networking infrastructure is probably not a good idea. It is better to bring your own.

Although wireless networking and mobile computing are often related, they are not identical, as Figure 6.4 shows. Here we see a distinction between fixed wireless and mobile wireless. Even notebook computers are
sometimes wired. For example, if a traveller plugs a notebook computer into the telephone jack in a hotel room, he has mobility without a wireless network.

6.4. ábra - Figure 6.4: Combinations of wireless networks and mobile computing

<table>
<thead>
<tr>
<th>Wireless</th>
<th>Mobile</th>
<th>Applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>No</td>
<td>Desktop computers in offices</td>
</tr>
<tr>
<td>No</td>
<td>Yes</td>
<td>A notebook computer used in a hotel room</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>Networks in older, unwired buildings</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>Portable office; PDA for store inventory</td>
</tr>
</tbody>
</table>

On the other hand, some wireless computers are not mobile. An important example is a company that owns an older building lacking network cabling, and which wants to connect its computers. Installing a wireless network may require little more than buying a small box with some electronics, unpacking it, and plugging it in. This solution may be far cheaper than having workmen put in cable ducts to wire the building.

But of course, there are also the true mobile, wireless applications, ranging from the portable office to people walking around a store with a PDA doing inventory. At many busy airports, car rental return clerks work in the parking lot with wireless portable computers. They type in the license plate number of returning cars, and their portable, which has a built-in printer, calls the main computer, gets the rental information, and prints out the bill on the spot.

A whole different application area for wireless networks is the expected merger of cell phones and PDAs into tiny wireless computers. A first attempt was tiny wireless PDAs that could display stripped-down Web pages on their even tinier screens.

2. Network Hardware

It is now time to turn our attention from the applications and social aspects of networking (the fun stuff) to the technical issues involved in network design (the work stuff). There is no generally accepted taxonomy into which all computer networks fit, but two dimensions stand out as important: transmission technology and scale. We will now examine each of these in turn. Broadly speaking, there are two types of transmission technology that are in widespread use. They are as follows:

- Broadcast ulinks;
- Point-to-point ulinks.

Broadcast networks have a single communication channel that is shared by all the machines on the network. Short messages, called packets in certain contexts, sent by any machine are received by all the others. An address field within the packet specifies the intended recipient. Upon receiving a packet, a machine checks the address field. If the packet is intended for the receiving machine, that machine processes the packet; if the packet is intended for some other machine, it is just ignored.

As an analogy, consider someone standing at the end of a corridor with many rooms off it and shouting "Watson, come here. I want you." Although the packet may actually be received (heard) by many people, only Watson responds. The others just ignore it. Another analogy is an airport announcement asking all flight 644 passengers to report to gate 12 for immediate boarding.

Broadcast systems generally also allow the possibility of addressing a packet to all destinations by using a special code in the address field. When a packet with this code is transmitted, it is received and processed by every machine on the network. This mode of operation is called broadcasting. Some broadcast systems also support transmission to a subset of the machines, something known as multicasting. One possible scheme is to reserve one bit to indicate multicasting. The remaining n-1 address bits can hold a group number. Each machine can "subscribe" to any or all of the groups. When a packet is sent to a certain group, it is delivered to all machines subscribing to that group.

In contrast, point-to-point networks consist of many connections between individual pairs of machines. To go from the source to the destination, a packet on this type of network may have to first visit one or more
intermediate machines. Often multiple routes, of different lengths, are possible, so finding good ones is important in point-to-point networks. As a general rule (although there are many exceptions), smaller, geographically localized networks tend to use broadcasting, whereas larger networks usually are point-to-point. Point-to-point transmission with one sender and one receiver is sometimes called unicasting.

An alternative criterion for classifying networks is their scale. In Figure 6.5 we classify multiple processor systems by their physical size. At the top are the personal area networks, networks that are meant for one person. For example, a wireless network connecting a computer with its mouse, keyboard, and printer is a personal area network. Also, a PDA that controls the user's hearing aid or pacemaker fits in this category. Beyond the personal area networks come longer-range networks. These can be divided into local, metropolitan, and wide area networks. Finally, the connection of two or more networks is called an internetwork. The worldwide Internet is a well-known example of an internetwork. Distance is important as a classification metric because different techniques are used at different scales. In this book we will be concerned with networks at all these scales. Below we give a brief introduction to network hardware.

6.5. ábra - Figure 6.5: Classification of interconnected processors by scale

<table>
<thead>
<tr>
<th>Interprocessor distance</th>
<th>Processors located in same</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 m</td>
<td>Square meter</td>
<td>Personal area network</td>
</tr>
<tr>
<td>10 m</td>
<td>Room</td>
<td>Local area network</td>
</tr>
<tr>
<td>100 m</td>
<td>Building</td>
<td>Metropolitan area network</td>
</tr>
<tr>
<td>1 km</td>
<td>Campus</td>
<td>Wide area network</td>
</tr>
<tr>
<td>10 km</td>
<td>City</td>
<td>The Internet</td>
</tr>
<tr>
<td>100 km</td>
<td>Country</td>
<td></td>
</tr>
<tr>
<td>1000 km</td>
<td>Continent</td>
<td></td>
</tr>
<tr>
<td>10,000 km</td>
<td>Planet</td>
<td></td>
</tr>
</tbody>
</table>

2.1. Local Area Networks

Local area networks, generally called LANs, are privately-owned networks within a single building or campus of up to a few kilometers in size. They are widely used to connect personal computers and workstations in company offices and factories to share resources (e.g., printers) and exchange information. LANs are distinguished from other kinds of networks by three characteristics: (1) their size, (2) their transmission technology, and (3) their topology.

LANs are restricted in size, which means that the worst-case transmission time is bounded and known in advance. Knowing this bound makes it possible to use certain kinds of designs that would not otherwise be possible. It also simplifies network management.

LANs may use a transmission technology consisting of a cable to which all the machines are attached, like the telephone company party lines once used in rural areas. Traditional LANs run at speeds of 10 Mbps to 100 Mbps, have low delay (microseconds or nanoseconds), and make very few errors. Newer LANs operate at up to 10 Gbps. In this book, we will adhere to tradition and measure line speeds in megabits/sec (1 Mbps is 1,000,000 bits/sec) and gigabits/sec (1 Gbps is 1,000,000,000 bits/sec). Various topologies are possible for broadcast LANs. Figure 6.6 shows two of them. In a bus (i.e., a linear cable) network, at any instant at most one machine is the master and is allowed to transmit. All other machines are required to refrain from sending. An arbitration mechanism is needed to resolve conflicts when two or more machines want to transmit simultaneously. The arbitration mechanism may be centralized or distributed. IEEE 802.3, popularly called Ethernet, for example, is a bus-based broadcast network with decentralized control, usually operating at 10 Mbps to 10 Gbps. Computers on an Ethernet can transmit whenever they want to; if two or more packets collide, each computer just waits a random time and tries again later.

6.6. ábra - Figure 6.6: Two broadcast networks (a) Bus; (b) Ring
A second type of broadcast system is the ring. In a ring, each bit propagates around on its own, not waiting for the rest of the packet to which it belongs. Typically, each bit circumnavigates the entire ring in the time it takes to transmit a few bits, often before the complete packet has even been transmitted. As with all other broadcast systems, some rule is needed for arbitrating simultaneous accesses to the ring. Various methods, such as having the machines take turns, are in use. IEEE 802.5 (the IBM token ring), is a ring-based LAN operating at 4 and 16 Mbps. FDDI is another example of a ring network.

Broadcast networks can be further divided into static and dynamic, depending on how the channel is allocated. A typical static allocation would be to divide time into discrete intervals and use a round-robin algorithm, allowing each machine to broadcast only when its time slot comes up. Static allocation wastes channel capacity when a machine has nothing to say during its allocated slot, so most systems attempt to allocate the channel dynamically (i.e., on demand).

Dynamic allocation methods for a common channel are either centralized or decentralized. In the centralized channel allocation method, there is a single entity, for example, a bus arbitration unit, which determines who goes next. It might do this by accepting requests and making a decision according to some internal algorithm. In the decentralized channel allocation method, there is no central entity; each machine must decide for itself whether to transmit. You might think that this always leads to chaos, but it does not. Later we will study many algorithms designed to bring order out of the potential chaos.

### 2.2. Metropolitan Area Networks

A metropolitan area network, or MAN, covers a city. The best-known example of a MAN is the cable television network available in many cities. This system grew from earlier community antenna systems used in areas with poor over-the-air television reception. In these early systems, a large antenna was placed on top of a nearby hill and signal was then piped to the subscribers' houses.

At first, these were locally-designed, ad hoc systems. Then companies began jumping into the business, getting contracts from city governments to wire up an entire city. The next step was television programming and even entire channels designed for cable only. Often these channels were highly specialized, such as all news, all sports, all cooking, all gardening, and so on. But from their inception until the late 1990s, they were intended for television reception only.

Starting when the Internet attracted a mass audience, the cable TV network operators began to realize that with some changes to the system, they could provide two-way Internet service in unused parts of the spectrum. At that point, the cable TV system began to morph from a way to distribute television to a metropolitan area network. To a first approximation, a MAN might look something like the system shown in Figure 6.7. In this figure we see both television signals and Internet being fed into the centralized head end for subsequent distribution to people's homes.

6.7. ábra - Figure 6.7: A metropolitan area network based on cable TV
Cable television is not the only MAN. Recent developments in high-speed wireless Internet access resulted in another MAN, which has been standardized as IEEE 802.16.

2.3. Wide Area Networks

A wide area network, or WAN, spans a large geographical area, often a country or continent. It contains a collection of machines intended for running user (i.e., application) programs. We will follow traditional usage and call these machines hosts. The hosts are connected by a communication subnet, or just subnet for short. The hosts are owned by the customers (e.g., people's personal computers), whereas the communication subnet is typically owned and operated by a telephone company or Internet service provider. The job of the subnet is to carry messages from host to host, just as the telephone system carries words from speaker to listener. Separation of the pure communication aspects of the network (the subnet) from the application aspects (the hosts), greatly simplifies the complete network design.

In most wide area networks, the subnet consists of two distinct components: transmission lines and switching elements. Transmission lines move bits between machines. They can be made of copper wire, optical fiber, or even radio links. Switching elements are specialized computers that connect three or more transmission lines. When data arrive on an incoming line, the switching element must choose an outgoing line on which to forward them. These switching computers have been called by various names in the past; the name router is now most commonly used. Unfortunately, some people pronounce it "rooter" and others have it rhyme with "doubter". Determining the correct pronunciation will be left as an exercise for the reader. (Note: the perceived correct answer may depend on where you live.)

In this model, shown in Figure 6.8 each host is frequently connected to a LAN on which a router is present, although in some cases a host can be connected directly to a router. The collection of communication lines and routers (but not the hosts) form the subnet.

6.8. ábra - Figure 6.8: Relation between hosts on LANs and the subnet
A short comment about the term "subnet" is in order here. Originally, its only meaning was the collection of routers and communication lines that moved packets from the source host to the destination host. However, some years later, it also acquired a second meaning in conjunction with network addressing. Unfortunately, no widely-used alternative exists for its initial meaning, so with some hesitation we will use it in both senses. From the context, it will always be clear which is meant.

In most WANs, the network contains numerous transmission lines, each one connecting a pair of routers. If two routers that do not share a transmission line wish to communicate, they must do this indirectly, via other routers. When a packet is sent from one router to another via one or more intermediate routers, the packet is received at each intermediate router in its entirety, stored there until the required output line is free, and then forwarded. A subnet organized according to this principle is called a store-and-forward or packet-switched subnet. Nearly all wide area networks (except those using satellites) have store-and-forward subnets. When the packets are small and all the same size, they are often called cells.

The principle of a packet-switched WAN is so important that it is worth devoting a few more words to it. Generally, when a process on some host has a message to be sent to a process on some other host, the sending host first cuts the message into packets, each one bearing its number in the sequence. These packets are then injected into the network one at a time in quick succession. The packets are transported individually over the network and deposited at the receiving host, where they are reassembled into the original message and delivered to the receiving process. A stream of packets resulting from some initial message is illustrated in Figure 6.9.

6.9. ábra - Figure 6.9: A stream of packets from sender to receiver

In this figure, all the packets follow the route ACE, rather than ABDE or ACDE. In some networks all packets from a given message must follow the same route; in others each packet is routed separately. Of course, if ACE is the best route, all packets may be sent along it, even if each packet is individually routed.

Routing decisions are made locally. When a packet arrives at router A, it is up to A to decide if this packet should be sent on the line to B or the line to C. How A makes that decision is called the routing algorithm. Many of them exist.

Not all WANs are packet switched. A second possibility for a WAN is a satellite system. Each router has an antenna through which it can send and receive. All routers can hear the output from the satellite, and in some cases they can also hear the upward transmissions of their fellow routers to the satellite as well. Sometimes the routers are connected to a substantial point-to-point subnet, with only some of them having a satellite antenna. Satellite networks are inherently broadcast and are most useful when the broadcast property is important.

2.4. Wireless Networks

Digital wireless communication is not a new idea. As early as 1901, the Italian physicist Guglielmo Marconi demonstrated a ship-to-shore wireless telegraph, using Morse Code (dots and dashes are binary, after all). Modern digital wireless systems have better performance, but the basic idea is the same.

To a first approximation, wireless networks can be divided into three main categories:

- System interconnection;
- Wireless LANs;
- Wireless WANs.
System interconnection is all about interconnecting the components of a computer using short-range radio. Almost every computer has a monitor, keyboard, mouse, and printer connected to the main unit by cables. So many new users have a hard time plugging all the cables into the right little holes (even though they are usually color coded) that most computer vendors offer the option of sending a technician to the user's home to do it. Consequently, some companies got together to design a short-range wireless network called Bluetooth to connect these components without wires. Bluetooth also allows digital cameras, headsets, scanners, and other devices to connect to a computer by merely being brought within range. No cables, no driver installation, just put them down, turn them on, and they work. For many people, this ease of operation is a big plus.

In the simplest form, system interconnection networks use the master-slave paradigm of Figure 6.10. The system unit is normally the master, talking to the mouse, keyboard, etc., as slaves. The master tells the slaves what addresses to use, when they can broadcast, how long they can transmit, what frequencies they can use, and so on.

6.10. ábra - Figure 6.10: (a) Bluetooth configuration; (b) Wireless LAN

The next step up in wireless networking are the wireless LANs. These are systems in which every computer has a radio modem and antenna with which it can communicate with other systems. Often there is an antenna on the ceiling that the machines talk to. However, if the systems are close enough, they can communicate directly with one another in a peer-to-peer configuration. Wireless LANs are becoming increasingly common in small offices and homes, where installing Ethernet is considered too much trouble, as well as in older office buildings, company cafeterias, conference rooms, and other places. There is a standard for wireless LANs, called IEEE 802.11, which most systems implement and which is becoming very widespread.

The third kind of wireless network is used in wide area systems. The radio network used for cellular telephones is an example of a low-bandwidth wireless system. This system has already gone through three generations. The first generation was analog and for voice only. The second generation was digital and for voice only. The third generation is digital and is for both voice and data. In a certain sense, cellular wireless networks are like wireless LANs, except that the distances involved are much greater and the bit rates much lower. Wireless LANs can operate at rates up to about 50 Mbps over distances of tens of meters. Cellular systems operate below 1 Mbps, but the distance between the base station and the computer or telephone is measured in kilometers rather than in meters.

In addition to these low-speed networks, high-bandwidth wide area wireless networks are also being developed. The initial focus is high-speed wireless Internet access from homes and businesses, bypassing the telephone system. This service is often called local multipoint distribution service. We will study it later in the book. A standard for it, called IEEE 802.16, has also been developed.

Almost all wireless networks hook up to the wired network at some point to provide access to files, databases, and the Internet. There are many ways these connections can be realized, depending on the circumstances. For example, in Figure 6.11(a), we depict an airplane with a number of people using modems and seat-back telephones to call the office. Each call is independent of the other ones. A much more efficient option, however, is the flying LAN of Figure 6.11(b). Here each seat comes equipped with an Ethernet connector into which passengers can plug their computers. A single router on the aircraft maintains a radio ulink with some router on the ground, changing routers as it flies along. This configuration is just a traditional LAN, except that its connection to the outside world happens to be a radio ulink instead of a hardwired line.
6.11. ábra - Figure 6.11: (a) Individual mobile computers; (b) A flying LAN

2.5. Home Networks

Home networking is on the horizon. The fundamental idea is that in the future most homes will be set up for networking. Every device in the home will be capable of communicating with every other device, and all of them will be accessible over the Internet. This is one of those visionary concepts that nobody asked for (like TV remote controls or mobile phones), but once they arrived nobody can imagine how they lived without them. Many devices are capable of being networked. Some of the more obvious categories (with examples) are as follows:

- Computers (desktop PC, notebook PC, PDA, shared peripherals);
- Entertainment (TV, DVD, VCR, camcorder, camera, stereo, MP3);
- Telecommunications (telephone, mobile telephone, intercom, fax);
- Appliances (microwave, refrigerator, clock, furnace, airco, lights);
- Telemetry (utility meter, smoke/burglar alarm, thermostat, babycam).

Home computer networking is already here in a limited way. Many homes already have a device to connect multiple computers to a fast Internet connection. Networked entertainment is not quite here, but as more and more music and movies can be downloaded from the Internet, there will be a demand to connect stereos and televisions to it. Also, people will want to share their own videos with friends and family, so the connection will need to go both ways. Telecommunications gear is already connected to the outside world, but soon it will be digital and go over the Internet. The average home probably has a dozen clocks (e.g., in appliances), all of which have to be reset twice a year when daylight saving time (summer time) comes and goes. If all the clocks were on the Internet, that resetting could be done automatically. Finally, remote monitoring of the home and its contents is a likely winner. Probably many parents would be willing to spend some money to monitor their sleeping babies on their PDAs when they are eating out, even with a rented teenager in the house. While one can imagine a separate network for each application area, integrating all of them into a single network is probably a better idea.

Home networking has some fundamentally different properties than other network types. First, the network and devices have to be easy to install. The author has installed numerous pieces of hardware and software on various computers over the years, with mixed results. A series of phone calls to the vendor's helpdesk typically resulted in answers like (1) Read the manual, (2) Reboot the computer, (3) Remove all hardware and software except ours and try again, (4) Download the newest driver from our Web site, and if all else fails, (5) Reformat the hard disk and then reinstall Windows from the CD-ROM. Telling the purchaser of an Internet refrigerator to download and install a new version of the refrigerator's operating system is not going to lead to happy customers. Computer users are accustomed to putting up with products that do not work; the car-, television-, and refrigerator-buying public is far less tolerant. They expect products to work for 100% from the word go.

Third, low price is essential for success. People will not pay a $50 premium for an Internet thermostat because few people regard monitoring their home temperature from work that important. For $5 extra, it might sell, though.

Fourth, the main application is likely to involve multimedia, so the network needs sufficient capacity. There is no market for Internet-connected televisions that show shaky movies at 320 x 240 pixel resolution and 10 frames/sec. Fast Ethernet, the workhorse in most offices, is not good enough for multimedia. Consequently,
home networks will need better performance than that of existing office networks and at lower prices before they become mass market items.

Fifth, it must be possible to start out with one or two devices and expand the reach of the network gradually. This means no format wars. Telling consumers to buy peripherals with IEEE 1394 (FireWire) interfaces and a few years later retracting that and saying USB 2.0 is the interface-of-the-month is going to make consumers skittish. The network interface will have to remain stable for many years; the wiring (if any) will have to remain stable for decades.

Sixth, security and reliability will be very important. Losing a few files to an e-mail virus is one thing; having a burglar disarm your security system from his PDA and then plunder your house is something quite different.

An interesting question is whether home networks will be wired or wireless. Most homes already have six networks installed: electricity, telephone, cable television, water, gas, and sewer. Adding a seventh one during construction is not difficult, but retrofitting existing houses is expensive. Cost favours wireless networking, but security favours wired networking. The problem with wireless is that the radio waves they use are quite good at going through fences. Not everyone is overjoyed at the thought of having the neighbours piggybacking on their Internet connection and reading their e-mail on its way to the printer.

In short, home networking offers many opportunities and challenges. Most of them relate to the need to be easy to manage, dependable, and secure, especially in the hands of nontechnical users, while at the same time delivering high performance at low cost.

2.6. Internetworks

Many networks exist in the world, often with different hardware and software. People connected to one network often want to communicate with people attached to a different one. The fulfillment of this desire requires that different, and frequently incompatible networks, be connected, sometimes by means of machines called gateways to make the connection and provide the necessary translation, both in terms of hardware and software. A collection of interconnected networks is called an internetwork or internet. These terms will be used in a generic sense, in contrast to the worldwide Internet (which is one specific internet), which we will always capitalize.

A common form of internet is a collection of LANs connected by a WAN. In fact, if we were to replace the label "subnet" by "WAN", nothing else in the figure would have to change. The only real technical distinction between a subnet and a WAN in this case is whether hosts are present. If the system within the gray area contains only routers, it is a subnet; if it contains both routers and hosts, it is a WAN. The real differences relate to ownership and use.

Subnets, networks, and internetworks are often confused. Subnet makes the most sense in the context of a wide area network, where it refers to the collection of routers and communication lines owned by the network operator. As an analogy, the telephone system consists of telephone switching offices connected to one another by high-speed lines, and to houses and businesses by low-speed lines. These lines and equipment, owned and managed by the telephone company, form the subnet of the telephone system. The telephones themselves (the hosts in this analogy) are not part of the subnet. The combination of a subnet and its hosts forms a network. In the case of a LAN, the cable and the hosts form the network. There really is no subnet.

An internetwork is formed when distinct networks are interconnected. In our view, connecting a LAN and a WAN or connecting two LANs forms an internetwork, but there is little agreement in the industry over terminology in this area. One rule of thumb is that if different organizations paid to construct different parts of the network and each maintains its part, we have an internetwork rather than a single network. Also, if the underlying technology is different in different parts (e.g., broadcast versus point-to-point), we probably have two networks.

3. Network Software

The first computer networks were designed with the hardware as the main concern and the software as an afterthought. This strategy no longer works. Network software is now highly structured. In the following sections we examine the software structuring technique in some detail. The method described here forms the keystone of the entire book and will occur repeatedly later on.
3.1. Protocol Hierarchies

To reduce their design complexity, most networks are organized as a stack of layers or levels, each one built upon the one below it. The number of layers, the name of each layer, the contents of each layer, and the function of each layer differ from network to network. The purpose of each layer is to offer certain services to the higher layers, shielding those layers from the details of how the offered services are actually implemented. In a sense, each layer is a kind of virtual machine, offering certain services to the layer above it.

This concept is actually a familiar one and used throughout computer science, where it is variously known as information hiding, abstract data types, data encapsulation, and object-oriented programming. The fundamental idea is that a particular piece of software (or hardware) provides a service to its users but keeps the details of its internal state and algorithms hidden from them.

Layer n on one machine carries on a conversation with layer n on another machine. The rules and conventions used in this conversation are collectively known as the layer n protocol. Basically, a protocol is an agreement between the communicating parties on how communication is to proceed. As an analogy, when a woman is introduced to a man, she may choose to stick out her hand. He, in turn, may decide either to shake it or kiss it, depending, for example, on whether she is an American lawyer at a business meeting or a European princess at a formal ball. Violating the protocol will make communication more difficult, if not completely impossible.

A five-layer network is illustrated in Figure 6.12. The entities comprising the corresponding layers on different machines are called peers. The peers may be processes, hardware devices, or even human beings. In other words, it is the peers that communicate by using the protocol.

6.12. ábra - Figure 6.12: Layers, protocols, and interfaces

In reality, no data are directly transferred from layer n on one machine to layer n on another machine. Instead, each layer passes data and control information to the layer immediately below it, until the lowest layer is reached. Below layer 1 is the physical medium through which actual communication occurs. In Figure 6.12, virtual communication is shown by dotted lines and physical communication by solid lines.

Between each pair of adjacent layers is an interface. The interface defines which primitive operations and services the lower layer makes available to the upper one. When network designers decide how many layers to include in a network and what each one should do, one of the most important considerations is defining clean interfaces between the layers. Doing so, in turn, requires that each layer perform a specific collection of well-understood functions. In addition to minimizing the amount of information that must be passed between layers, clear-cut interfaces also make it simpler to replace the implementation of one layer with a completely different implementation (e.g., all the telephone lines are replaced by satellite channels) because all that is required of the new implementation is that it offer exactly the same set of services to its upstairs neighbor as the old implementation did. In fact, it is common that different hosts use different implementations.
A set of layers and protocols is called a network architecture. The specification of an architecture must contain enough information to allow an implementer to write the program or build the hardware for each layer so that it will correctly obey the appropriate protocol. Neither the details of the implementation nor the specification of the interfaces is part of the architecture because these are hidden away inside the machines and not visible from the outside. It is not even necessary that the interfaces on all machines in a network be the same, provided that each machine can correctly use all the protocols. A list of protocols used by a certain system, one protocol per layer, is called a protocol stack. The subjects of network architectures, protocol stacks, and the protocols themselves are the principal topics of this book.

An analogy may help explain the idea of multilayer communication. Imagine two philosophers (peer processes in layer 3), one of whom speaks Urdu and English and one of whom speaks Chinese and French. Since they have no common language, they each engage a translator (peer processes at layer 2), each of whom in turn contacts a secretary (peer processes in layer 1). Philosopher 1 wishes to convey his affection for oryctolagus cuniculus to his peer. To do so, he passes a message (in English) across the 2/3 interface to his translator, saying "I like rabbits", as illustrated in Figure 6.13. The translators have agreed on a neutral language known to both of them, Dutch, so the message is converted to "Ik vind konijnen leuk". The choice of language is the layer 2 protocol and is up to the layer 2 peer processes.

6.13. ábra - Figure 6.13: The philosopher-translator-secretary architecture

The translator then gives the message to a secretary for transmission, by, for example, fax (the layer 1 protocol). When the message arrives, it is translated into French and passed across the 2/3 interface to philosopher 2. Note that each protocol is completely independent of the other ones as long as the interfaces are not changed. The translators can switch from Dutch to say, Finnish, at will, provided that they both agree, and neither changes his interface with either layer 1 or layer 3. Similarly, the secretaries can switch from fax to e-mail or telephone without disturbing (or even informing) the other layers. Each process may add some information intended only for its peer. This information is not passed upward to the layer above.

Now consider a more technical example: how to provide communication to the top layer of the five-layer network in Figure 6.14. A message, M, is produced by an application process running in layer 5 and given to layer 4 for transmission. Layer 4 puts a header in front of the message to identify the message and passes the result to layer 3. The header includes control information, such as sequence numbers, to allow layer 4 on the
destination machine to deliver messages in the right order if the lower layers do not maintain sequence. In some layers, headers can also contain sizes, times, and other control fields.

6.14. ábra - Figure 6.14: Example information flow supporting virtual communication in layer 5

In many networks, there is no limit to the size of messages transmitted in the layer 4 protocol, but there is nearly always a limit imposed by the layer 3 protocol. Consequently, layer 3 must break up the incoming messages into smaller units, packets, prepending a layer 3 header to each packet. In this example, M is split into two parts, M1 and M2.

Layer 3 decides which of the outgoing lines to use and passes the packets to layer 2. Layer 2 adds not only a header to each piece, but also a trailer, and gives the resulting unit to layer 1 for physical transmission. At the receiving machine the message moves upward, from layer to layer, with headers being stripped off as it progresses. None of the headers for layers below n are passed up to layer n.

The important thing to understand about Figure 6.14 is the relation between the virtual and actual communication and the difference between protocols and interfaces. The peer processes in layer 4, for example, conceptually think of their communication as being "horizontal", using the layer 4 protocol. Each one is likely to have a procedure called something like SendToOtherSide and GetFromOtherSide, even though these procedures actually communicate with lower layers across the 3/4 interface, not with the other side.

The peer process abstraction is crucial to all network design. Using it, the unmanageable task of designing the complete network can be broken into several smaller, manageable design problems, namely, the design of the individual layers.

3.2. Reference Models

Now that we have discussed layered networks in the abstract, it is time to look at some examples. In the next two sections we will discuss two important network architectures, the OSI reference model and the TCP/IP reference model. Although the protocols associated with the OSI model are rarely used any more, the model itself is actually quite general and still valid, and the features discussed at each layer are still very important. The TCP/IP model has the opposite properties: the model itself is not of much use but the protocols are widely used. For this reason we will look at both of them in detail. Also, sometimes you can learn more from failures than from successes.

3.3. The OSI Reference Model

The OSI model (minus the physical medium) is shown in Figure 6.15. This model is based on a proposal developed by the International Standards Organization (ISO) as a first step toward international standardization.
of the protocols used in the various layers. The model is called the ISO OSI (Open Systems Interconnection) Reference Model because it deals with connecting open systems — that is, systems that are open for communication with other systems. We will just call it the OSI model for short.

6.15. ábra - Figure 6.15: The OSI reference model

The OSI model has seven layers. The principles that were applied to arrive at the seven layers can be briefly summarized as follows:

• A layer should be created where a different abstraction is needed.

• Each layer should perform a well-defined function.

• The function of each layer should be chosen with an eye toward defining internationally standardized protocols.

• The layer boundaries should be chosen to minimize the information flow across the interfaces.

• The number of layers should be large enough that distinct functions need not be thrown together in the same layer out of necessity and small enough that the architecture does not become unwieldy.

Below we will discuss each layer of the model in turn, starting at the bottom layer. Note that the OSI model itself is not a network architecture because it does not specify the exact services and protocols to be used in each layer. It just tells what each layer should do. However, ISO has also produced standards for all the layers, although these are not part of the reference model itself. Each one has been published as a separate international standard.

3.4. The Physical Layer

The physical layer is concerned with transmitting raw bits over a communication channel. The design issues have to do with making sure that when one side sends a 1 bit, it is received by the other side as a 1 bit, not as a 0
bit. Typical questions here are how many volts should be used to represent a 1 and how many for a 0, how many nanoseconds a bit lasts, whether transmission may proceed simultaneously in both directions, how the initial connection is established and how it is torn down when both sides are finished, and how many pins the network connector has and what each pin is used for. The design issues here largely deal with mechanical, electrical, and timing interfaces, and the physical transmission medium, which lies below the physical layer.

3.5. The Data Ulink Layer

The main task of the data ulink layer is to transform a raw transmission facility into a line that appears free of undetected transmission errors to the network layer. It accomplishes this task by having the sender break up the input data into data frames (typically a few hundred or a few thousand bytes) and transmits the frames sequentially. If the service is reliable, the receiver confirms correct receipt of each frame by sending back an acknowledgement frame.

Another issue that arises in the data ulink layer (and most of the higher layers as well) is how to keep a fast transmitter from drowning a slow receiver in data. Some traffic regulation mechanism is often needed to let the transmitter know how much buffer space the receiver has at the moment. Frequently, this flow regulation and the error handling are integrated.

Broadcast networks have an additional issue in the data ulink layer: how to control access to the shared channel. A special sublayer of the data ulink layer, the medium access control sublayer, deals with this problem.

3.6. The Network Layer

The network layer controls the operation of the subnet. A key design issue is determining how packets are routed from source to destination. Routes can be based on static tables that are “wired into” the network and rarely changed. They can also be determined at the start of each conversation, for example, a terminal session (e.g., a login to a remote machine). Finally, they can be highly dynamic, being determined anew for each packet, to reflect the current network load.

If too many packets are present in the subnet at the same time, they will get in one another's way, forming bottlenecks. The control of such congestion also belongs to the network layer. More generally, the quality of service provided (delay, transit time, jitter, etc.) is also a network layer issue.

When a packet has to travel from one network to another to get to its destination, many problems can arise. The addressing used by the second network may be different from the first one. The second one may not accept the packet at all because it is too large. The protocols may differ, and so on. It is up to the network layer to overcome all these problems to allow heterogeneous networks to be interconnected.

In broadcast networks, the routing problem is simple, so the network layer is often thin or even nonexistent.

3.7. The Transport Layer

The basic function of the transport layer is to accept data from above, split it up into smaller units if need be, pass these to the network layer, and ensure that the pieces all arrive correctly at the other end. Furthermore, all this must be done efficiently and in a way that isolates the upper layers from the inevitable changes in the hardware technology.

The transport layer also determines what type of service to provide to the session layer, and, ultimately, to the users of the network. The most popular type of transport connection is an error-free point-to-point channel that delivers messages or bytes in the order in which they were sent. However, other possible kinds of transport service are the transporting of isolated messages, with no guarantee about the order of delivery, and the broadcasting of messages to multiple destinations. The type of service is determined when the connection is established. (As an aside, an error-free channel is impossible to achieve; what people really mean by this term is that the error rate is low enough to ignore in practice.)

The transport layer is a true end-to-end layer, all the way from the source to the destination. In other words, a program on the source machine carries on a conversation with a similar program on the destination machine, using the message headers and control messages. In the lower layers, the protocols are between each machine and its immediate neighbors, and not between the ultimate source and destination machines, which may be separated by many routers.
3.8. The Session Layer

The session layer allows users on different machines to establish sessions between them. Sessions offer various services, including dialog control (keeping track of whose turn it is to transmit), token management (preventing two parties from attempting the same critical operation at the same time), and synchronization (checkpointing long transmissions to allow them to continue from where they were after a crash).

3.9. The Presentation Layer

Unlike lower layers, which are mostly concerned with moving bits around, the presentation layer is concerned with the syntax and semantics of the information transmitted. In order to make it possible for computers with different data representations to communicate, the data structures to be exchanged can be defined in an abstract way, along with a standard encoding to be used "on the wire". The presentation layer manages these abstract data structures and allows higher-level data structures (e.g., banking records), to be defined and exchanged.

3.10. The Application Layer

The application layer contains a variety of protocols that are commonly needed by users. One widely-used application protocol is HTTP (HyperText Transfer Protocol), which is the basis for the World Wide Web. When a browser wants a Web page, it sends the name of the page it wants to the server using HTTP. The server then sends the page back. Other application protocols are used for file transfer, electronic mail, and network news.

4. Questions

1. What is a computer network?
2. What kind of network topologies do you know?
3. What does the protocol mean?
4. Describe the OSI reference model!
5. What are the main types of physical media for data transmission?
6. What are the main types of network equipment?
7. What is the ad-hoc network?
7. fejezet - Internet Services

1. The TCP/IP Reference Model

Let us now turn from the OSI reference model to the reference model used in the grandparent of all wide area computer networks, the ARPANET, and its successor, the worldwide Internet. Although we will give a brief history of the ARPANET later, it is useful to mention a few key aspects of it now. The ARPANET was a research network sponsored by the DoD (U.S. Department of Defense). It eventually connected hundreds of universities and government installations, using leased telephone lines. When satellite and radio networks were added later, the existing protocols had trouble interworking with them, so a new reference architecture was needed. Thus, the ability to connect multiple networks in a seamless way was one of the major design goals from the very beginning. This architecture later became known as the TCP/IP Reference Model, after its two primary protocols.

Given the DoD's worry that some of its precious hosts, routers, and internetwork gateways might get blown to at a moment's notice, another major goal was that the network be able to survive loss of subnet hardware, with existing conversations not being broken off. In other words, DoD wanted connections to remain intact as long as the source and destination machines were functioning, even if some of the machines or transmission lines in between were suddenly put out of operation. Furthermore, a flexible architecture was needed since applications with divergent requirements were envisioned, ranging from transferring files to real-time speech transmission.

1.1. The Internet Layer

All these requirements led to the choice of a packet-switching network based on a connectionless internetwork layer. This layer, called the internet layer, is the linchpin that holds the whole architecture together. Its job is to permit hosts to inject packets into any network and have them travel independently to the destination (potentially on a different network). They may even arrive in a different order than they were sent, in which case it is the job of higher layers to rearrange them, if in-order delivery is desired. Note that "internet" is used here in a generic sense, even though this layer is present in the Internet.

The analogy here is with the (snail) mail system. A person can drop a sequence of international letters into a mail box in one country, and with a little luck, most of them will be delivered to the correct address in the destination country. Probably the letters will travel through one or more international mail gateways along the way, but this is transparent to the users. Furthermore, that each country (i.e., each network) has its own stamps, preferred envelope sizes, and delivery rules is hidden from the users.

The internet layer defines an official packet format and protocol called IP (Internet Protocol). The job of the internet layer is to deliver IP packets where they are supposed to go. Packet routing is clearly the major issue here, as is avoiding congestion. For these reasons, it is reasonable to say that the TCP/IP internet layer is similar in functionality to the OSI network layer. Figure 7.1 shows this correspondence.

7.1. ábra - Figure 7.1: The TCP/IP reference model
1.2. The Transport Layer

The layer above the internet layer in the TCP/IP model is now usually called the transport layer. It is designed to allow peer entities on the source and destination hosts to carry on a conversation, just as in the OSI transport layer. Two end-to-end transport protocols have been defined here. The first one, TCP (Transmission Control Protocol), is a reliable connection-oriented protocol that allows a byte stream originating on one machine to be delivered without error on any other machine in the internet. It fragments the incoming byte stream into discrete messages and passes each one on to the internet layer. At the destination, the receiving TCP process reassembles the received messages into the output stream. TCP also handles flow control to make sure a fast sender cannot swamp a slow receiver with more messages than it can handle.

The second protocol in this layer, UDP (User Datagram Protocol), is an unreliable, connectionless protocol for applications that do not want TCP’s sequencing or flow control and wish to provide their own. It is also widely used for one-shot, client-server-type request-reply queries and applications in which prompt delivery is more important than accurate delivery, such as transmitting speech or video. The relation of IP, TCP, and UDP is shown in Figure 7.2. Since the model was developed, IP has been implemented on many other networks.

7.2. ábra - Figure 7.2: Protocols and networks in the TCP/IP model initially

1.3. The Application Layer
The TCP/IP model does not have session or presentation layers. No need for them was perceived, so they were not included. Experience with the OSI model has proven this view correct: they are of little use to most applications.

On top of the transport layer is the application layer. It contains all the higher-level protocols. The early ones included virtual terminal (TELNET), file transfer (FTP), and electronic mail (SMTP), as shown in Figure 7.2. The virtual terminal protocol allows a user on one machine to log onto a distant machine and work there. The file transfer protocol provides a way to move data efficiently from one machine to another. Electronic mail was originally just a kind of file transfer, but later a specialized protocol (SMTP) was developed for it. Many other protocols have been added to these over the years: the Domain Name System (DNS) for mapping host names onto their network addresses, NNTP, the protocol for moving USENET news articles around, and HTTP, the protocol for fetching pages on the World Wide Web, and many others.

1.4. The Host-to-Network Layer

Below the internet layer is a great void. The TCP/IP reference model does not really say much about what happens here, except to point out that the host has to connect to the network using some protocol so it can send IP packets to it. This protocol is not defined and varies from host to host and network to network. Books and papers about the TCP/IP model rarely discuss it.

2. A Comparison of the OSI and TCP/IP Reference Models

The OSI and TCP/IP reference models have much in common. Both are based on the concept of a stack of independent protocols. Also, the functionality of the layers is roughly similar. For example, in both models the layers up through and including the transport layer are there to provide an end-to-end, network-independent transport service to processes wishing to communicate. These layers form the transport provider. Again in both models, the layers above transport are application-oriented users of the transport service.

Despite these fundamental similarities, the two models also have many differences. In this section we will focus on the key differences between the two reference models. It is important to note that we are comparing the reference models here, not the corresponding protocol stacks. The protocols themselves will be discussed later.

Three concepts are central to the OSI model:

• Services,
• Interfaces,
• Protocols.

Probably the biggest contribution of the OSI model is to make the distinction between these three concepts explicit. Each layer performs some services for the layer above it. The service definition tells what the layer does, not how entities above it access it or how the layer works. It defines the layer's semantics.

A layer's interface tells the processes above it how to access it. It specifies what the parameters are and what results to expect. It, too, says nothing about how the layer works inside.

Finally, the peer protocols used in a layer are the layer's own business. It can use any protocols it wants to, as long as it gets the job done (i.e., provides the offered services). It can also change them at will without affecting software in higher layers.

These ideas fit very nicely with modern ideas about object-oriented programming. An object, like a layer, has a set of methods (operations) that processes outside the object can invoke. The semantics of these methods define the set of services that the object offers. The methods' parameters and results form the object's interface. The code internal to the object is its protocol and is not visible or of any concern outside the object.

The TCP/IP model did not originally clearly distinguish between service, interface, and protocol, although people have tried to retrofit it after the fact to make it more OSI-like. For example, the only real services offered by the internet layer are SEND IP PACKET and RECEIVE IP PACKET.
As a consequence, the protocols in the OSI model are better hidden than in the TCP/IP model and can be replaced relatively easily as the technology changes. Being able to make such changes is one of the main purposes of having layered protocols in the first place.

The OSI reference model was devised before the corresponding protocols were invented. This ordering means that the model was not biased toward one particular set of protocols, a fact that made it quite general. The downside of this ordering is that the designers did not have much experience with the subject and did not have a good idea of which functionality to put in which layer.

For example, the data ulink layer originally dealt only with point-to-point networks. When broadcast networks came around, a new sublayer had to be hacked into the model. When people started to build real networks using the OSI model and existing protocols, it was discovered that these networks did not match the required service specifications (wonder of wonders), so convergence sublayers had to be grafted onto the model to provide a place for papering over the differences. Finally, when people started to build real networks using the OSI model and existing protocols, it was discovered that these networks did not match the required service specifications (wonder of wonders), so convergence sublayers had to be grafted onto the model to provide a place for papering over the differences. Finally, the committee originally expected that each country would have one network, run by the government and using the OSI protocols, so no thought was given to internetworking. To make a long story short, things did not turn out that way.

With TCP/IP the reverse was true: the protocols came first, and the model was really just a description of the existing protocols. There was no problem with the protocols fitting the model. They fit perfectly. The only trouble was that the model did not fit any other protocol stacks. Consequently, it was not especially useful for describing other, non-TCP/IP networks.

Turning from philosophical matters to more specific ones, an obvious difference between the two models is the number of layers: the OSI model has seven layers and the TCP/IP has four layers. Both have (inter)network, transport, and application layers, but the other layers are different.

Another difference is in the area of connectionless versus connection-oriented communication. The OSI model supports both connectionless and connection-oriented communication in the network layer, but only connection-oriented communication in the transport layer, where it counts (because the transport service is visible to the users). The TCP/IP model has only one mode in the network layer (connectionless) but supports both modes in the transport layer, giving the users a choice. This choice is especially important for simple request-response protocols.

Let us now leave the general principles and start looking at the details of the Internet's network layer. At the network layer, the Internet can be viewed as a collection of subnetworks or Autonomous Systems (ASes) that are interconnected. There is no real structure, but several major backbones exist. These are constructed from high-bandwidth lines and fast routers. Attached to the backbones are regional (midlevel) networks, and attached to these regional networks are the LANs at many universities, companies, and Internet service providers. A sketch of this quasi-hierarchical organization is given in Figure 7.3.

7.3. ábra - Figure 7.3: The Internet is an interconnected collection of many networks
The glue that holds the whole Internet together is the network layer protocol, IP (Internet Protocol). Unlike most older network layer protocols, it was designed from the beginning with internetworking in mind. A good way to think of the network layer is this. Its job is to provide a best-efforts (i.e., not guaranteed) way to transport datagrams from source to destination, without regard to whether these machines are on the same network or whether there are other networks in between them.

Communication in the Internet works as follows. The transport layer takes data streams and breaks them up into datagrams. In theory, datagrams can be up to 64 Kbytes each, but in practice they are usually not more than 1500 bytes (so they fit in one Ethernet frame). Each datagram is transmitted through the Internet, possibly being fragmented into smaller units as it goes. When all the pieces finally get to the destination machine, they are reassembled by the network layer into the original datagram. This datagram is then handed to the transport layer, which inserts it into the receiving process' input stream.

3. The IP Protocol

An appropriate place to start our study of the network layer in the Internet is the format of the IP datagrams themselves. An IP datagram consists of a header part and a text part. The header has a 20-byte fixed part and a variable length optional part. The header format is shown in Figure 7.4. It is transmitted in big-endian order: from left to right, with the high-order bit of the Version field going first. (The SPARC is big endian; the Pentium is little-endian.) On little endian machines, software conversion is required on both transmission and reception.

7.4. ábra - Figure 7.4: The IPv4 (Internet Protocol) header
The Version field keeps track of which version of the protocol the datagram belongs to. By including the version in each datagram, it becomes possible to have the transition between versions take years, with some machines running the old version and others running the new one. Currently a transition between IPv4 and IPv6 is going on, has already taken years, and is by no means close to being finished. Some people even think it will never happen. As an aside on numbering, IPv5 was an experimental real-time stream protocol that was never widely used.

Since the header length is not constant, a field in the header, IHL, is provided to tell how long the header is, in 32-bit words. The minimum value is 5, which applies when no options are present. The maximum value of this 4-bit field is 15, which limits the header to 60 bytes, and thus the Options field to 40 bytes. For some options, such as one that records the route a packet has taken, 40 bytes is far too small, making that option useless.

The Type of service field is one of the few fields that has changed its meaning (slightly) over the years. It was and is still intended to distinguish between different classes of service. Various combinations of reliability and speed are possible. For digitized voice, fast delivery beats accurate delivery. For file transfer, error-free transmission is more important than fast transmission.

### 3.1. IP Addresses

Every host and router on the Internet has an IP address, which encodes its network number and host number. The combination is unique: in principle, no two machines on the Internet have the same IP address. All IP addresses are 32 bits long and are used in the Source address and Destination address fields of IP packets. It is important to note that an IP address does not actually refer to a host. It really refers to a network interface, so if a host is on two networks, it must have two IP addresses. However, in practice, most hosts are on one network and thus have one IP address.

For several decades, IP addresses were divided into the five categories listed in Figure 7.5. This allocation has come to be called classful addressing. It is no longer used, but references to it in the literature are still common. We will discuss the replacement of classful addressing shortly.

#### 7.5. ábra - Figure 7.5: IP address formats
The class A, B, C, and D formats allow for up to 128 networks with 16 million hosts each, 16,384 networks with up to 64K hosts, and 2 million networks (e.g., LANs) with up to 256 hosts each (although a few of these are special). Also supported is multicast, in which a datagram is directed to multiple hosts. Addresses beginning with 1111 are reserved for future use. Over 500,000 networks are now connected to the Internet, and the number grows every year. Network numbers are managed by a nonprofit corporation called ICANN (Internet Corporation for Assigned Names and Numbers) to avoid conflicts. In turn, ICANN has delegated parts of the address space to various regional authorities, which then dole out IP addresses to ISPs and other companies.

Network addresses, which are 32-bit numbers, are usually written in dotted decimal notation. In this format, each of the 4 bytes is written in decimal, from 0 to 255. For example, the 32-bit hexadecimal address C0290614 is written as 192.41.6.20. The lowest IP address is 0.0.0.0 and the highest is 255.255.255.255.

### 3.2. The Application Layer

Having finished all the preliminaries, we now come to the layer where all the applications are found. The layers below the application layer are there to provide reliable transport, but they do not do real work for users. In this chapter we will study some real network applications.

However, even in the application layer there is a need for support protocols, to allow the applications to function. Accordingly, we will look at one of these before starting with the applications themselves. The item in question is DNS, which handles naming within the Internet. After that, we will examine three real applications: electronic mail, the World Wide Web, and finally, multimedia.

### 3.3. DNS – The Domain Name System

Although programs theoretically could refer to hosts, mailboxes, and other resources by their network (e.g., IP) addresses, these addresses are hard for people to remember. Also, sending e-mail to tana@128.111.24.41 means that if Tana's ISP or organization moves the mail server to a different machine with a different IP address, her e-mail address has to change. Consequently, ASCII names were introduced to decouple machine names from machine addresses. In this way, Tana's address might be something like tana@art.ucsb.edu. Nevertheless, the network itself understands only numerical addresses, so some mechanism is required to convert the ASCII strings to network addresses. In the following sections we will study how this mapping is accomplished in the Internet.

Way back in the ARPANET, there was simply a file, hosts.txt, that listed all the hosts and their IP addresses. Every night, all the hosts would fetch it from the site at which it was maintained. For a network of a few hundred large timesharing machines, this approach worked reasonably well.

However, when thousands of minicomputers and PCs were connected to the net, everyone realized that this approach could not continue to work forever. For one thing, the size of the file would become too large. However, even more important, host name conflicts would occur constantly unless names were centrally
managed, something unthinkable in a huge international network due to the load and latency. To solve these problems, DNS (the Domain Name System) was invented.

The essence of DNS is the invention of a hierarchical, domain-based naming scheme and a distributed database system for implementing this naming scheme. It is primarily used for mapping host names and e-mail destinations to IP addresses but can also be used for other purposes. DNS is defined in RFCs 1034 and 1035.

Very briefly, the way DNS is used is as follows. To map a name onto an IP address, an application program calls a library procedure called the resolver, passing it the name as a parameter. The resolver sends a UDP packet to a local DNS server, which then looks up the name and returns the IP address to the resolver, which then returns it to the caller. Armed with the IP address, the program can then establish a TCP connection with the destination or send it UDP packets.

4. The DNS Name Space

Managing a large and constantly changing set of names is a nontrivial problem. In the postal system, name management is done by requiring letters to specify (implicitly or explicitly) the country, state or province, city, and street address of the addressee. By using this kind of hierarchical addressing, there is no confusion between the Marvin Anderson on Main St. in White Plains, N.Y. and the Marvin Anderson on Main St. in Austin, Texas. DNS works the same way.

Conceptually, the Internet is divided into over 200 top-level domains, where each domain covers many hosts. Each domain is partitioned into subdomains, and these are further partitioned, and so on. All these domains can be represented by a tree, as shown in Figure 7.6. The leaves of the tree represent domains that have no subdomains (but do contain machines, of course). A leaf domain may contain a single host, or it may represent a company and contain thousands of hosts.

7.6. ábra - Figure 7.6: A portion of the Internet domain name space

The top-level domains come in two flavors: generic and countries. The original generic domains were com (commercial), edu (educational institutions), gov (the U.S. Federal Government), int (certain international organizations), mil (the U.S. armed forces), net (network providers), and org (nonprofit organizations). The country domains include one entry for every country, as defined in ISO 3166.

In November 2000, ICANN approved four new, general-purpose, top-level domains, namely, biz (businesses), info (information), name (people’s names), and pro (professions, such as doctors and lawyers). In addition, three more specialized top-level domains were introduced at the request of certain industries. These are aero (aerospace industry), coop (co-operatives), and museum (museums). Other top-level domains will be added in the future.

As an aside, as the Internet becomes more commercial, it also becomes more contentious. Take pro, for example. It was intended for certified professionals. But who is a professional? And certified by whom? Doctors and lawyers clearly are professionals. But what about freelance photographers, piano teachers, magicians, plumbers, barber, exterminators, tattoo artists, mercenaries, and prostitutes? Are these occupations professional and thus eligible for pro domains? And if so, who certifies the individual practitioners?
In general, getting a second-level domain, such as name-of-company.com, is easy. It merely requires going to a registrar for the corresponding top-level domain (com in this case) to check if the desired name is available and not somebody else's trademark. If there are no problems, the requester pays a small annual fee and gets the name. By now, virtually every common (English) word has been taken in the com domain. Try household articles, animals, plants, body parts, etc. Nearly all are taken.

Each domain is named by the path upward from it to the (unnamed) root. The components are separated by periods (pronounced "dot"). Thus, the engineering department at Sun Microsystems might be eng.sun.com., rather than a UNIX-style name such as /com/sun/eng. Notice that this hierarchical naming means that eng.sun.com., does not conflict with a potential use of eng in eng.yale.edu., which might be used by the Yale English department.

Domain names can be either absolute or relative. An absolute domain name always ends with a period (e.g., eng.sun.com.), whereas a relative one does not. Relative names have to be interpreted in some context to uniquely determine their true meaning. In both cases, a named domain refers to a specific node in the tree and all the nodes under it.

Domain names are case insensitive, so edu, Edu, and EDU mean the same thing. Component names can be up to 63 characters long, and full path names must not exceed 255 characters.

In principle, domains can be inserted into the tree in two different ways. For example, cs.yale.edu could equally well be listed under the us country domain as cs.yale.ct.us. In practice, however, most organizations in the United States are under a generic domain, and most outside the United States are under the domain of their country. There is no rule against registering under two top-level domains, but few organizations except multinationals do it (e.g., sony.com and sony.nl).

Each domain controls how it allocates the domains under it. For example, Japan has domains ac.jp and co.jp that mirror edu and com. The Netherlands does not make this distinction and puts all organizations directly under nl. Thus, all three of the following are university computer science departments:

• cs.yale.edu (Yale University, in the United States);
• cs.vu.nl (Vrije Universiteit, in The Netherlands);
• cs.keio.ac.jp (Keio University, in Japan).

To create a new domain, permission is required of the domain in which it will be included. For example, if a VLSI group is started at Yale and wants to be known as vlsi.cs.yale.edu, it has to get permission from whoever manages cs.yale.edu. Similarly, if a new university is chartered, say, the University of Northern South Dakota, it must ask the manager of the edu domain to assign it unsd.edu. In this way, name conflicts are avoided and each domain can keep track of all its subdomains. Once a new domain has been created and registered, it can create subdomains, such as cs.unsd.edu, without getting permission from anybody higher up the tree.

Naming follows organizational boundaries, not physical networks. For example, if the computer science and electrical engineering departments are located in the same building and share the same LAN, they can nevertheless have distinct domains. Similarly, even if computer science is split over Babbage Hall and Turing Hall, the hosts in both buildings will normally belong to the same domain.

5. Electronic Mail

E-mail, like most other forms of communication, has its own conventions and styles. In particular, it is very informal and has a low threshold of use. People who would never dream of calling up or even writing a letter to a Very Important Person do not hesitate for a second to send a sloppily-written e-mail.

5.1. Architecture and Services

In this section we will provide an overview of what e-mail systems can do and how they are organized. They normally consist of two subsystems: the user agents, which allow people to read and send e-mail, and the message transfer agents, which move the messages from the source to the destination. The user agents are local programs that provide a command- based, menu-based, or graphical method for interacting with the e-mail system. The message transfer agents are typically system daemons, that is, processes that run in the background.
Their job is to move e-mail through the system. Typically, e-mail systems support five basic functions. Let us take a look at them.

Composition refers to the process of creating messages and answers. Although any text editor can be used for the body of the message, the system itself can provide assistance with addressing and the numerous header fields attached to each message. For example, when answering a message, the e-mail system can extract the originator's address from the incoming e-mail and automatically insert it into the proper place in the reply.

Transfer refers to moving messages from the originator to the recipient. In large part, this requires establishing a connection to the destination or some intermediate machine, outputting the message, and releasing the connection. The e-mail system should do this automatically, without bothering the user.

Reporting has to do with telling the originator what happened to the message. Was it delivered? Was it rejected? Was it lost? Numerous applications exist in which confirmation of delivery is important and may even have legal significance (“Well, Your Honor, my e-mail system is not very reliable, so I guess the electronic subpoena just got lost somewhere”).

Displaying incoming messages is needed so people can read their e-mail. Sometimes conversion is required or a special viewer must be invoked, for example, if the message is a PostScript file or digitized voice. Simple conversions and formatting are sometimes attempted as well.

Disposition is the final step and concerns what the recipient does with the message after receiving it. Possibilities include throwing it away before reading, throwing it away after reading, saving it, and so on. It should also be possible to retrieve and reread saved messages, forward them, or process them in other ways.

In addition to these basic services, some e-mail systems, especially internal corporate ones, provide a variety of advanced features. Let us just briefly mention a few of these. When people move or when they are away for some period of time, they may want their e-mail forwarded, so the system should be able to do this automatically.

Most systems allow users to create mailboxes to store incoming e-mail. Commands are needed to create and destroy mailboxes, inspect the contents of mailboxes, insert and delete messages from mailboxes, and so on.

Corporate managers often need to send a message to each of their subordinates, customers, or suppliers. This gives rise to the idea of a mailing list, which is a list of e-mail addresses. When a message is sent to the mailing list, identical copies are delivered to everyone on the list.

Other advanced features are carbon copies, blind carbon copies, high-priority e-mail, secret (i.e., encrypted) e-mail, alternative recipients if the primary one is not currently available, and the ability for secretaries to read and answer their bosses' e-mail.

The message inside the envelope consists of two parts: the header and the body. The header contains control information for the user agents. The body is entirely for the human recipient. Envelopes and messages are illustrated in Figure 7.7.

7.7. ábra - Figure 7.7: Envelopes and messages (a) Paper mail; (b) Electronic mail
5.2. The User Agent

E-mail systems have two basic parts, as we have seen: the user agents and the message transfer agents. In this section we will look at the user agents. A user agent is normally a program (sometimes called a mail reader) that accepts a variety of commands for composing, receiving, and replying to messages, as well as for manipulating mailboxes. Some user agents have a fancy menu- or icon-driven interface that requires a mouse, whereas others expect 1-character commands from the keyboard. Functionally, these are the same. Some systems are menu- or icon-driven but also have keyboard shortcuts.

5.2.1. Sending E-mail

To send an e-mail message, a user must provide the message, the destination address, and possibly some other parameters. The message can be produced with a free-standing text editor, a word processing program, or possibly with a specialized text editor built into the user agent. The destination address must be in a format that the user agent can deal with. Many user agents expect addresses of the form user@dns-address. Since we have studied DNS earlier in this chapter, we will not repeat that material here.

5.2.2. Reading E-mail

Typically, when a user agent is started up, it looks at the user's mailbox for incoming e-mail before displaying anything on the screen. Then it may announce the number of messages in the mailbox or display a one-line summary of each one and wait for a command.

As an example of how a user agent works, let us take a look at a typical mail scenario. After starting up the user agent, the user asks for a summary of his e-mail. Each line refers to one message. In this example, the mailbox contains eight messages.

5.2.3. MIME – The Multipurpose Internet Mail Extensions
In the early days e-mail consisted exclusively of text messages written in English and expressed in ASCII. Nowadays, on the worldwide Internet, this approach is no longer adequate. The problems include sending and receiving:

- Messages in languages with accents (e.g., French and German);
- Messages in non-Latin alphabets (e.g., Hebrew and Russian);
- Messages in languages without alphabets (e.g., Chinese and Japanese);
- Messages not containing text at all (e.g., audio or images).

This solution, called MIME (Multipurpose Internet Mail Extensions) is now widely used. The basic idea of MIME is to continue to use the text format, but to add structure to the message body and define encoding rules for non-ASCII messages.

### 5.3. Final Delivery

Up until now, we have assumed that all users work on machines that are capable of sending and receiving e-mail. As we saw, e-mail is delivered by having the sender establish a TCP connection to the receiver and then ship the e-mail over it. This model worked fine for decades when all ARPANET (and later Internet) hosts were, in fact, on-line all the time to accept TCP connections. However, with the advent of people who access the Internet by calling their ISP over a modem, it breaks down. The problem is this: what happens when Elinor wants to send Carolyn e-mail and Carolyn is not currently on-line? Elinor cannot establish a TCP connection to Carolyn and thus cannot run the SMTP protocol.

One solution is to have a message transfer agent on an ISP machine accept e-mail for its customers and store it in their mailboxes on an ISP machine. Since this agent can be on-line all the time, e-mail can be sent to it 24 hours a day.

#### 5.3.1. POP3

Unfortunately, this solution creates another problem: how does the user get the e-mail from the ISP's message transfer agent? The solution to this problem is to create another protocol that allows user transfer agents (on client PCs) to contact the message transfer agent (on the ISP's machine) and allow e-mail to be copied from the ISP to the user. One such protocol is POP3 (Post Office Protocol Version 3). POP3 begins when the user starts the mail reader. The mail reader calls up the ISP (unless there is already a connection) and establishes a TCP connection with the message transfer agent. Once the connection has been established, the POP3 protocol goes through three states in sequence:

- Authorization,
- Transactions,
- Update.

The authorization state deals with having the user log in. The transaction state deals with the user collecting the e-mails and marking them for deletion from the mailbox. The update state actually causes the e-mails to be deleted.

#### 5.3.2. IMAP

For a user with one e-mail account at one ISP that is always accessed from one PC, POP3 works fine and is widely used due to its simplicity and robustness. However, it is a computer-industry truism that as soon as something works well, somebody will start demanding more features (and getting more bugs). That happened with e-mail, too. For example, many people have a single e-mail account at work or school and want to access it from work, from their home PC, from their laptop when on business trips, and from cybercafés when on so-called vacation. While POP3 allows this, since it normally downloads all stored messages at each contact, the result is that the user's e-mail quickly gets spread over multiple machines, more or less at random, some of them not even the user's.
This disadvantage gave rise to an alternative final delivery protocol, IMAP (Internet Message Access Protocol), which basically assumes that the user will clear out the mailbox on every contact and work off-line after that. IMAP assumes that all the e-mail will remain on the server indefinitely in multiple mailboxes. IMAP provides extensive mechanisms for reading messages or even parts of messages, a feature useful when using a slow modem to read the text part of a multipart message with large audio and video attachments. Since the working assumption is that messages will not be transferred to the user's computer for permanent storage, IMAP provides mechanisms for creating, destroying, and manipulating multiple mailboxes on the server. In this way a user can maintain a mailbox for each correspondent and move messages there from the inbox after they have been read.

5.3.3. Webmail

One final topic worth mentioning is Webmail. Some Web sites, for example, Google provide e-mail service to anyone who wants it. They work as follows. They have normal message transfer agents listening to the server for incoming SMTP connections. Basically, when the user goes to the e-mail Web page, a form is presented in which the user is asked for a login name and password. When the user clicks on Sign In, the login name and password are sent to the server, which then validates them.

6. The World Wide Web

The World Wide Web is an architectural framework for accessing unlinked documents spread out over millions of machines all over the Internet. The Web (also known as WWW) began in 1989 at CERN, the European center for nuclear research. CERN has several accelerators at which large teams of scientists from the participating European countries carry out research in particle physics. These teams often have members from half a dozen or more countries. Most experiments are highly complex and require years of advance planning and equipment construction. The Web grew out of the need to have these large teams of internationally dispersed researchers collaborate using a constantly changing collection of reports, blueprints, drawings, photos, and other documents.

6.1. Architectural Overview

From the users' point of view, the Web consists of a vast, worldwide collection of documents or Web pages, often just called pages for short. Each page may contain ulinks to other pages anywhere in the world. Users can follow a ulink by clicking on it, which then takes them to the page pointed to. This process can be repeated indefinitely. Pages are viewed with a program called a browser, of which Internet Explorer or Chrome ones. The browser fetches the page requested, interprets the text and formatting commands on it, and displays the page, properly formatted, on the screen.

Web pages, start with a title, contain some information, and end with the e-mail address of the page's maintainer. Strings of text that are ulinks to other pages, called hyperulinks, are often highlighted, by underlining, displaying them in a special color, or both. To follow a ulink, the user places the mouse cursor on the highlighted area, which causes the cursor to change, and clicks on it.

The basic model of how the Web works is shown in Figure 7.8. Here the browser is displaying a Web page on the client machine. When the user clicks on a line of text that is ulinked to a page on the abcd.com server, the browser follows the hyperulink by sending a message to the abcd.com server asking it for the page. When the page arrives, it is displayed. If this page contains a hyperulink to a page on the xyz.com server that is clicked on, the browser then sends a request to that machine for the page, and so on indefinitely.

7.8. ábra - Figure 7.8: The parts of the Web model
6.1.1. The Client Side

Let us now examine the client side of Figure 7.8 in more detail. In essence, a browser is a program that can display a Web page and catch mouse clicks to items on the displayed page. When an item is selected, the browser follows the hyperulink and fetches the page selected. Therefore, the embedded hyperulink needs a way to name any other page on the Web. Pages are named using URLs (Uniform Resource Locators). A typical URL is http://www.abcd.com/products.html.

We will explain URLs later in this chapter. For the moment, it is sufficient to know that a URL has three parts: the name of the protocol (http), the DNS name of the machine where the page is located (www.abcd.com), and (usually) the name of the file containing the page (products.html).

When a user clicks on a hyperulink, the browser carries out a series of steps in order to fetch the page pointed to. Suppose that a user is browsing the Web and finds a ulink on Internet telephony that point to ITU's home page, which is http://www.itu.org/home/index.html. Let us trace the steps that occur when this ulink is selected:

- The browser determines the URL (by seeing what was selected).
- DNS replies with 156.106.192.32.
- The browser makes a TCP connection to port 80 on 156.106.192.32.
- It then sends over a request asking for file /home/index.html.
- The www.itu.org server sends the file /home/index.html.
- The TCP connection is released.
- The browser displays all the text in /home/index.html.
- The browser fetches and displays all images in this file.

Many browsers display which step they are currently executing in a status line at the bottom of the screen. In this way, when the performance is poor, the user can see if it is due to DNS not responding, the server not responding, or simply network congestion during page transmission.

To be able to display the new page (or any page), the browser has to understand its format. To allow all browsers to understand all Web pages, Web pages are written in a standardized language called HTML, which describes Web pages. We will discuss it in detail later in this chapter.
Although a browser is basically an HTML interpreter, most browsers have numerous buttons and features to make it easier to navigate the Web. Most have a button for going back to the previous page, a button for going forward to the next page (only operative after the user has gone back from it), and a button for going straight to the user's own start page. Most browsers have a button or menu item to set a bookmark on a given page and another one to display the list of bookmarks, making it possible to revisit any of them with only a few mouse clicks. Pages can also be saved to disk or printed. Numerous options are generally available for controlling the screen layout and setting various user preferences.

In addition to having ordinary text (not underlined) and hypertext (underlined), Web pages can also contain icons, line drawings, maps, and photographs. Each of these can (optionally) be linked to another page. Clicking on one of these elements causes the browser to fetch the ulinked page and display it on the screen, the same as clicking on text. With images such as photos and maps, which page is fetched next may depend on what part of the image was clicked on.

Not all pages contain HTML. A page may consist of a formatted document in PDF format, an icon in GIF format, a photograph in JPEG format, a song in MP3 format, a video in MPEG format, or any one of hundreds of other file types. Since standard HTML pages may ulink to any of these, the browser has a problem when it encounters a page it cannot interpret.

Rather than making the browsers larger and larger by building in interpreters for a rapidly growing collection of file types, most browsers have chosen a more general solution. When a server returns a page, it also returns some additional information about the page. This information includes the MIME type of the page. Pages of type text/html are just displayed directly, as are pages in a few other built-in types. If the MIME type is not one of the built-in ones, the browser consults its table of MIME types to tell it how to display the page. This table associates a MIME type with a viewer.

There are two possibilities: plug-ins and helper applications. A plug-in is a code module that the browser fetches from a special directory on the disk and installs as an extension to itself. Because plug-ins run inside the browser, they have access to the current page and can modify its appearance. After the plug-in has done its job (usually after the user has moved to a different Web page), the plug-in is removed from the browser's memory.

Each browser has a set of procedures that all plug-ins must implement so the browser can call the plug-in. For example, there is typically a procedure the browser's base code calls to supply the plug-in with data to display. This set of procedures is the plug-in's interface and is browser specific.

In addition, the browser makes a set of its own procedures available to the plug-in, to provide services to plug-ins. Typical procedures in the browser interface are for allocating and freeing memory, displaying a message on the browser's status line, and querying the browser about parameters.

Before a plug-in can be used, it must be installed. The usual installation procedure is for the user to go to the plug-in's Web site and download an installation file. On Windows, this is typically a self-extracting zip file with extension .exe. When the zip file is double clicked, a little program attached to the front of the zip file is executed. This program unzips the plug-in and copies it to the browser's plug-in directory. Then it makes the appropriate calls to register the plug-in's MIME type and to associate the plug-in with it. On UNIX, the installer is often a shell script that handles the copying and registration.

The other way to extend a browser is to use a helper application. This is a complete program, running as a separate process. Since the helper is a separate program, it offers no interface to the browser and makes no use of browser services. Instead, it usually just accepts the name of a scratch file where the content file has been stored, opens the file, and displays the contents. Typically, helpers are large programs that exist independently of the browser, such as Adobe's Acrobat Reader for displaying PDF files or Microsoft Word. Some programs (such as Acrobat) have a plug-in that invokes the helper itself.

Many helper applications use the MIME type application. A considerable number of subtypes have been defined, for example, application/pdf for PDF files and application/msword for Word files. In this way, a URL can point directly to a PDF or Word file, and when the user clicks on it, Acrobat or Word is automatically started and handed the name of a scratch file containing the content to be displayed. Consequently, browsers can be configured to handle a virtually unlimited number of document types with no changes to the browser. Modern Web servers are often configured with hundreds of type/subtype combinations and new ones are often added every time a new program is installed.
Helper applications are not restricted to using the application MIME type. Adobe Photoshop uses image/x-photoshop and RealOne Player is capable of handling audio/mp3, for example.

On Windows, when a program is installed on the computer, it registers the MIME types it wants to handle. This mechanism leads to conflict when multiple viewers are available for some subtype, such as video/mpg. What happens is that the last program to register overwrites existing (MIME type, helper application) associations, capturing the type for itself. As a consequence, installing a new program may change the way a browser handles existing types.

### 6.1.2. The Server Side

So much for the client side. Now let us take a look at the server side. As we saw above, when the user types in a URL or clicks on a line of hypertext, the browser parses the URL and interprets the part between http:// and the next slash as a DNS name to look up. Armed with the IP address of the server, the browser establishes a TCP connection to port 80 on that server. Then it sends over a command containing the rest of the URL, which is the name of a file on that server. The server then returns the file for the browser to display.

To a first approximation, a Web server is similar to the server. That server, like a real Web server, is given the name of a file to look up and return. In both cases, the steps that the server performs in its main loop are:

- Accept a TCP connection from a client (a browser).
- Get the name of the file requested.
- Get the file (from disk).
- Return the file to the client.
- Release the TCP connection.

Modern Web servers have more features, but in essence, this is what a Web server does.

A problem with this design is that every request requires making a disk access to get the file. The result is that the Web server cannot serve more requests per second than it can make disk accesses. A high-end SCSI disk has an average access time of around 5 msec, which limits the server to at most 200 requests/sec, less if large files have to be read often.

One obvious improvement (used by all Web servers) is to maintain a cache in memory of the n most recently used files. Before going to disk to get a file, the server checks the cache. If the file is there, it can be served directly from memory, thus eliminating the disk access. Although effective caching requires a large amount of main memory and some extra processing time to check the cache and manage its contents, the savings in time are nearly always worth the overhead and expense.

### 6.2. URLs – Uniform Resource Locators

We have repeatedly said that Web pages may contain pointers to other Web pages. Now it is time to see in a bit more detail how these pointers are implemented. When the Web was first created, it was immediately apparent that having one page point to another Web page required mechanisms for naming and locating pages. In particular, three questions had to be answered before a selected page could be displayed:

- What is the page called?
- Where is the page located?
- How can the page be accessed?

If every page were somehow assigned a unique name, there would not be any ambiguity in identifying pages. Nevertheless, the problem would not be solved. Consider a parallel between people and pages. In the United States, almost everyone has a social security number, which is a unique identifier, as no two people are supposed to have the same one. Nevertheless, if you are armed only with a social security number, there is no way to find the owner's address, and certainly no way to tell whether you should write to the person in English, Spanish, or Chinese. The Web has basically the same problems.
The solution chosen identifies pages in a way that solves all three problems at once. Each page is assigned a URL (Uniform Resource Locator) that effectively serves as the page's worldwide name. URLs have three parts: the protocol (also known as the scheme), the DNS name of the machine on which the page is located, and a local name uniquely indicating the specific page (usually just a file name on the machine where it resides). As an example example, the Web site for the author's department contains several videos about the university and the city of Amsterdam. The URL for the video page is http://www.cs.vu.nl/video/index-en.html.

This URL consists of three parts: the protocol (http), the DNS name of the host (www.cs.vu.nl), and the file name (video/index-en.html), with certain punctuation separating the pieces. The file name is a path relative to the default Web directory at cs.vu.nl.

Many sites have built-in shortcuts for file names. At many sites, a null file name defaults to the organization's main home page. Typically, when the file named is a directory, this implies a file named index.html. Finally, ~user/ might be mapped onto user's WWW directory, and then onto the file index.html in that directory. Thus, the author's home page can be reached at http://www.cs.vu.nl/~ast/ even though the actual file name is index.html in a certain default directory.

Let us briefly go over the list. The http protocol is the Web's native language, the one spoken by Web servers. HTTP stands for HyperText Transfer Protocol. We will examine it in more detail later in this chapter.

The ftp protocol is used to access files by FTP, the Internet's file transfer protocol. FTP has been around more than two decades and is well entrenched. Numerous FTP servers all over the world allow people anywhere on the Internet to log in and download whatever files have been placed on the FTP server. The Web does not change this; it just makes obtaining files by FTP easier, as FTP has a somewhat arcane interface (but it is more powerful than HTTP, for example, it allows a user on machine A to transfer a file from machine B to machine C).

It is possible to access a local file as a Web page, either by using the file protocol, or more simply, by just naming it. This approach is similar to using FTP but does not require having a server. Of course, it works only for local files, not remote ones.

### 6.3. Static Web Documents

The basis of the Web is transferring Web pages from server to client. In the simplest form, Web pages are static, that is, are just files sitting on some server waiting to be retrieved. In this context, even a video is a static Web page because it is just a file. In this section we will look at static Web pages in detail. In the next one, we will examine dynamic content.

### 6.4. HTML – The HyperText Markup Language

Web pages are currently written in a language called HTML (HyperText Markup Language). HTML allows users to produce Web pages that include text, graphics, and pointers to other Web pages. HTML is a markup language, a language for describing how documents are to be formatted. The term "markup" comes from the old days when copyeditors actually marked up documents to tell the printer — in those day a human being — which fonts to use, and so on. Markup languages thus contain explicit commands for formatting. For example, in HTML, <b> means start boldface mode, and </b> means leave boldface mode. The advantage of a markup language over one with no explicit markup is that writing a browser for it is straightforward: the browser simply has to understand the markup commands. TeX and troff are other well-known examples of markup languages.

Below we will give a brief introduction to HTML, just to give an idea of what it is like. While it is certainly possible to write HTML documents with any standard editor, and many people do, it is also possible to use special HTML editors or word processors that do most of the work (but correspondingly give the user less control over all the details of the final result).

A Web page consists of a head and a body, each enclosed by <html> and </html> tags (Figure 7.9) (formatting commands), although most browsers do not complain if these tags are missing. The head is bracketed by the <head> and </head> tags and the body is bracketed by the <body> and </body> tags. The strings inside the tags are called directives. Most HTML tags have this format, that is they use, <something> to mark the beginning of something and </something> to mark its end. Most browsers have a menu item VIEW SOURCE or something like that. Selecting this item displays the current page's HTML source, instead of its formatted output.
### 7.9. ábra - Figure 7.9: A selection of common HTML tags. Some can have additional parameters

<table>
<thead>
<tr>
<th>Tag</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>&lt;html&gt;</code></td>
<td>Declares the Web page to be written in HTML</td>
</tr>
<tr>
<td><code>&lt;head&gt;</code></td>
<td>Delimits the page's head</td>
</tr>
<tr>
<td><code>&lt;title&gt;</code></td>
<td>Defines the title (not displayed on the page)</td>
</tr>
<tr>
<td><code>&lt;body&gt;</code></td>
<td>Delimits the page's body</td>
</tr>
<tr>
<td><code>&lt;h1&gt;</code></td>
<td>Delimits a level 1 heading</td>
</tr>
<tr>
<td><code>&lt;b&gt;</code></td>
<td>Set ... in boldface</td>
</tr>
<tr>
<td><code>&lt;i&gt;</code></td>
<td>Set ... in italics</td>
</tr>
<tr>
<td><code>&lt;center&gt;</code></td>
<td>Center ... on the page horizontally</td>
</tr>
<tr>
<td><code>&lt;ul&gt;</code></td>
<td>Brackets an unordered (bulleted) list</td>
</tr>
<tr>
<td><code>&lt;ol&gt;</code></td>
<td>Brackets a numbered list</td>
</tr>
<tr>
<td><code>&lt;li&gt;</code></td>
<td>Starts a list item (there is no &lt;/li&gt;)</td>
</tr>
<tr>
<td><code>&lt;br&gt;</code></td>
<td>Forces a line break here</td>
</tr>
<tr>
<td><code>&lt;p&gt;</code></td>
<td>Starts a paragraph</td>
</tr>
<tr>
<td><code>&lt;hr&gt;</code></td>
<td>Inserts a Horizontal rule</td>
</tr>
<tr>
<td><code>&lt;img src=&quot;...&quot;&gt;</code></td>
<td>Displays an image here</td>
</tr>
<tr>
<td><code>&lt;a href=&quot;...&quot;&gt;</code></td>
<td>Defines a hyperlink</td>
</tr>
</tbody>
</table>

Tags can be in either lower case or upper case. Thus, `<head>` and `<HEAD>` mean the same thing, but newer versions of the standard require lower case only. Actual layout of the HTML document is irrelevant. HTML parsers ignore extra spaces and carriage returns since they have to reformat the text to make it fit the current display area. Consequently, white space can be added at will to make HTML documents more readable, something most of them are badly in need of. As another consequence, blank lines cannot be used to separate paragraphs, as they are simply ignored. An explicit tag is required.

Some tags have (named) parameters, called attributes. For example, `<img src="abc" alt="foobar">` is a tag, `<img>`, with parameter src set equal to abc and parameter alt set equal to foobar. For each tag, the HTML standard gives a list of what the permitted parameters, if any, are, and what they mean. Because each parameter is named, the order in which the parameters are given is not significant.

Technically, HTML documents are written in the ISO 8859-1 Latin-1 character set, but for users whose keyboards support only ASCII, escape sequences are present for the special characters, such as è. The list of special characters is given in the standard. All of them begin with an ampersand and end with a semicolon. For example, `&nbsp;` produces a space, `&egrave;` produces è and `&eacute;` produces é. Since `<`, `>`, and `&` have special meanings, they can be expressed only with their escape sequences, `<`, `>`, and `&`; respectively.

The main item in the head is the title, delimited by `<title>` and `</title>`, but certain kinds of meta-information may also be present. The title itself is not displayed on the page. Some browsers use it to label the page's window.

The tags `<b>` and `<i>` are used to enter boldface and italics mode, respectively. If the browser is not capable of displaying boldface and italics, it must use some other method of rendering them, for example, using a different color for each or perhaps reverse video.

HTML provides various mechanisms for making lists, including nested lists. Lists are started with `<ul>` or `<ol>`, with `<li>` used to mark the start of the items in both cases. The `<ul>` tag starts an unordered list. The individual items, which are marked with the `<li>` tag in the source, appear with bullets (*) in front of them. A variant of this mechanism is `<ol>`, which is for ordered lists. When this tag is used, the `<li>` items are numbered by the browser. Other than the use of different starting and ending tags, `<ul>` and `<ol>` have the same syntax and similar results.
The `<br>`, `<p>`, and `<hr>` tags all indicate a boundary between sections of text. The precise format can be determined by the style sheet (see below) associated with the page. The `<br>` tag just forces a line break. Typically, browsers do not insert a blank line after `<br>`. In contrast, `<p>` starts a paragraph, which might, for example, insert a blank line and possibly some indentation. (Theoretically, `<p>` exists to mark the end of a paragraph, but it is rarely used; most HTML authors do not even know it exists.) Finally, `<hr>` forces a break and draws a horizontal line across the screen.

HTML allows images to be included in-line on a Web page. The `<img>` tag specifies that an image is to be displayed at the current position in the page. It can have several parameters. The `src` parameter gives the URL of the image. The HTML standard does not specify which graphic formats are permitted. In practice, all browsers support GIF and JPEG files. Browsers are free to support other formats, but this extension is a two-edged sword. If a user is accustomed to a browser that supports, say, BMP files, he may include these in his Web pages and later be surprised when other browsers just ignore all of his wonderful art.

Other parameters of `<img>` are `align`, which controls the alignment of the image with respect to the text baseline (top, middle, bottom). `alt`, which provides text to use instead of the image when the user has disabled images, and `ismap`, a flag indicating that the image is an active map (i.e., clickable picture).

Finally, we come to hyperlinks, which use the `<a>` (anchor) and `</a>` tags. Like `<img>`, `<a>` has various parameters, including `href` (the URL) and `name` (the hyperlink's name). The text between the `<a>` and `</a>` is displayed. If it is selected, the hyperlink is followed to a new page. It is also permitted to put an `<img>` image there, in which case clicking on the image also activates the hyperlink.

### 6.5. XML and XSL

HTML, with or without forms, does not provide any structure to Web pages. It also mixes the content with the formatting. As e-commerce and other applications become more common, there is an increasing need for structuring Web pages and separating the content from the formatting. For example, a program that searches the Web for the best price for some book or CD needs to analyze many Web pages looking for the item's title and price.

For this reason, the W3C has developed an enhancement to HTML to allow Web pages to be structured for automated processing. Two new languages have been developed for this purpose. First, XML (eXtensible Markup Language) describes Web content in a structured way and second, XSL (eXtensible Style Language) describes the formatting independently of the content. Both of these are large and complicated topics, so our brief introduction below just scratches the surface, but it should give an idea of how they work.

### 6.6. Dynamic Web Documents

So far, the model we have used is that the client sends a file name to the server, which then returns the file. In the early days of the Web, all content was, in fact, static like this (just files). However, in recent years, more and more content has become dynamic, that is, generated on demand, rather than stored on disk. Content generation can take place either on the server side or on the client side. Let us now examine each of these cases in turn.

#### 6.6.1. Server-Side Dynamic Web Page Generation

To see why server-side content generation is needed, consider the use of forms, as described earlier. When a user fills in a form and clicks on the submit button, a message is sent to the server indicating that it contains the contents of a form, along with the fields the user filled in. This message is not the name of a file to return. What is needed is that the message is given to a program or script to process. Usually, the processing involves using the user-supplied information to look up a record in a database on the server's disk and generate a custom HTML page to send back to the client. For example, in an e-commerce application, after the user clicks on `PROCEED TO CHECKOUT`, the browser returns the cookie containing the contents of the shopping cart, but some program or script on the server has to be invoked to process the cookie and generate an HTML page in response. The HTML page might display a form containing the list of items in the cart and the user's last-known shipping address along with a request to verify the information and to specify the method of payment.

#### 6.6.2. Client-Side Dynamic Web Page Generation
CGI, PHP, JSP, and ASP scripts solve the problem of handling forms and interactions with databases on the server. They can all accept incoming information from forms, look up information in one or more databases, and generate HTML pages with the results. What none of them can do is respond to mouse movements or interact with users directly. For this purpose, it is necessary to have scripts embedded in HTML pages that are executed on the client machine rather than the server machine. Starting with HTML 4.0, such scripts are permitted using the tag <script>. The most popular scripting language for the client side is JavaScript, so we will now take a quick look at it.

JavaScript is a scripting language, very loosely inspired by some ideas from the Java programming language. It is definitely not Java. Like other scripting languages, it is a very high level language. For example, in a single line of JavaScript it is possible to pop up a dialog box, wait for text input, and store the resulting string in a variable. High-level features like this make JavaScript ideal for designing interactive Web pages. On the other hand, the fact that it is not standardized and is mutating faster than a fruit fly trapped in an X-ray machine makes it extremely difficult to write JavaScript programs that work on all platforms, but maybe some day it will stabilize.

### 6.7. HTTP – The HyperText Transfer Protocol

The transfer protocol used throughout the World Wide Web is HTTP (HyperText Transfer Protocol). It specifies what messages clients may send to servers and what responses they get back in return. Each interaction consists of one ASCII request, followed by one RFC 822 MIME-like response. All clients and all servers must obey this protocol. It is defined in RFC 2616. In this section we will look at some of its more important properties.

### 6.8. The Wireless Web

There is considerable interest in small portable devices capable of accessing the Web via a wireless ulink. In fact, the first tentative steps in that direction have already been taken. No doubt there will be a great deal of change in this area in the coming years, but it is still worth examining some of the current ideas relating to the wireless Web to see where we are now and where we might be heading. We will focus on the first two wide area wireless Web systems to hit the market: WAP and i-mode.


Once the Internet and mobile phones had become commonplace, it did not take long before somebody got the idea to combine them into a mobile phone with a built-in screen for wireless access to e-mail and the Web. The “somebody” in this case was a consortium initially led by Nokia, Ericsson, Motorola, and phone.com (formerly Unwired Planet) and now boasting hundreds of members. The system is called WAP (Wireless Application Protocol).

A WAP device may be an enhanced mobile phone, PDA, or notebook computer without any voice capability. The specification allows all of them and more. The basic idea is to use the existing digital wireless infrastructure. Users can literally call up a WAP gateway over the wireless ulink and send Web page requests to it. The gateway then checks its cache for the page requested. If present, it sends it; if absent, it fetches it over the wired Internet. In essence, this means that WAP 1.0 is a circuit-switched system with a fairly high per-minute connect charge. To make a long story short, people did not like accessing the Internet on a tiny screen and paying by the minute, so WAP was something of a flop (although there were other problems as well). However, WAP and its competitor, i-mode (discussed below), appear to be converging on a similar technology, so WAP 2.0 may yet be a big success. Since WAP 1.0 was the first attempt at wireless Internet, it is worth describing it at least briefly.

WAP is essentially a protocol stack for accessing the Web, but optimized for low-bandwidth connections using wireless devices having a slow CPU, little memory, and a small screen. These requirements are obviously different from those of the standard desktop PC scenario, which leads to some protocol differences.

Besides cost, the other aspect that no doubt hurt WAP's acceptance is the fact that it does not use HTML. Instead, the WAE layer uses a markup language called WML (Wireless Markup Language), which is an application of XML. As a consequence, in principle, a WAP device can only access those pages that have been converted to WML. However, since this greatly restricts the value of WAP, the architecture calls for an on-the-fly filter from HTML to WML to increase the set of pages available. This architecture is illustrated in Figure 7.10.
7.10. ábra - Figure 7.10: The WAP architecture

In all fairness, WAP was probably a little ahead of its time. When WAP was first started, XML was hardly known outside W3C and so the press reported its launch as WAP DOES NOT USE HTML. A more accurate headline would have been: WAP ALREADY USES THE NEW HTML STANDARD. But once the damage was done, it was hard to repair and WAP 1.0 never caught on. We will revisit WAP after first looking at its major competitor.

While a multi-industry consortium of telecom vendors and computer companies was busy hammering out an open standard using the most advanced version of HTML available, other developments were going on in Japan. There, a Japanese woman, Mari Matsunaga, invented a different approach to the wireless Web called i-mode (information-mode). She convinced the wireless subsidiary of the former Japanese telephone monopoly that her approach was right, and in Feb. 1999 NTT DoCoMo (literally: Japanese Telephone and Telegraph Company everywhere you go) launched the service in Japan. Within 3 years it had over 35 million Japanese subscribers, who could access over 40,000 special i-mode Web sites. It also had most of the world's telecom companies drooling over its financial success, especially in light of the fact that WAP appeared to be going nowhere. Let us now take a look at what i-mode is and how it works.

The i-mode system has three major components: a new transmission system, a new handset, and a new language for Web page design. The transmission system consists of two separate networks: the existing circuit-switched mobile phone network (somewhat comparable to D-AMPS), and a new packet-switched network constructed specifically for i-mode service. Voice mode uses the circuit switched network and is billed per minute of connection time. I-mode uses the packet-switched network and is always on (like ADSL or cable), so there is no billing for connect time. Instead, there is a charge for each packet sent. It is not currently possible to use both networks at once.

7. Multimedia

The wireless Web is an exciting new development, but it is not the only one. For many people, multimedia is the holy grail of networking. When the word is mentioned, both the propeller heads and the suits begin salivating as if on cue. The former see immense technical challenges in providing (interactive) video on demand to every home. The latter see equally immense profits in it. Since multimedia requires high bandwidth, getting it to work over fixed connections is hard enough. Even VHS-quality video over wireless is a few years away, so our treatment will focus on wired systems.

Literally, multimedia is just two or more media. If the publisher of this book wanted to join the current hype about multimedia, it could advertise the book as using multimedia technology. After all, it contains two media: text and graphics (the figures). Nevertheless, when most people refer to multimedia, they generally mean the combination of two or more continuous media, that is, media that have to be played during some well-defined time interval, usually with some user interaction. In practice, the two media are normally audio and video, that is, sound plus moving pictures.

However, many people often refer to pure audio, such as Internet telephony or Internet radio as multimedia as well, which it is clearly not. Actually, a better term is streaming media, but we will follow the herd and consider
real-time audio to be multimedia as well. In the following sections we will examine how computers process audio and video, how they are compressed, and some network applications of these technologies. For a comprehensive (three volume) treatment on networked multimedia, see (Steinmetz and Nahrstedt, 2002; Steinmetz and Nahrstedt, 2003a; and Steinmetz and Nahrstedt, 2003b).

### 7.1. Introduction to Digital Audio

An audio (sound) wave is a one-dimensional acoustic (pressure) wave. When an acoustic wave enters the ear, the eardrum vibrates, causing the tiny bones of the inner ear to vibrate along with it, sending nerve pulses to the brain. These pulses are perceived as sound by the listener. In a similar way, when an acoustic wave strikes a microphone, the microphone generates an electrical signal, representing the sound amplitude as a function of time. The representation, processing, storage, and transmission of such audio signals are a major part of the study of multimedia systems.

### 7.2. Audio Compression

CD-quality audio requires a transmission bandwidth of 1.411 Mbps, as we just saw. Clearly, substantial compression is needed to make transmission over the Internet practical. For this reason, various audio compression algorithms have been developed. Probably the most popular one is MPEG audio, which has three layers (variants), of which MP3 (MPEG audio layer 3) is the most powerful and best known. Large amounts of music in MP3 format are available on the Internet, not all of it legal, which has resulted in numerous lawsuits from the artists and copyright owners. MP3 belongs to the audio portion of the MPEG video compression standard. We will discuss video compression later in this chapter; let us look at audio compression now.

### 7.3. Streaming Audio

Let us now move from the technology of digital audio to three of its network applications. Our first one is streaming audio, that is, listening to sound over the Internet. This is also called music on demand. In the next two, we will look at Internet radio and voice over IP, respectively. The Internet is full of music Web sites, many of which list song titles that users can click on to play the songs. Some of these are free sites (e.g., new bands looking for publicity); others require payment in return for music, although these often offer some free samples as well (e.g., the first 15 seconds of a song).

The process starts when the user clicks on a song. Then the browser goes into action. Step 1 is for it to establish a TCP connection to the Web server to which the song is hyperlinked. Step 2 is to send over a GET request in HTTP to request the song. Next (steps 3 and 4), the server fetches the song (which is just a file in MP3 or some other format) from the disk and sends it back to the browser. If the file is larger than the server's memory, it may fetch and send the music a block at a time.

Using the MIME type, for example, audio/mp3, (or the file extension), the browser looks up how it is supposed to display the file. Normally, there will be a helper application such as RealOne Player, Windows Media Player, or Winamp, associated with this type of file. Since the usual way for the browser to communicate with a helper is to write the content to a scratch file, it will save the entire music file as a scratch file on the disk (step 5) first. Then it will start the media player and pass it the name of the scratch file. In step 6, the media player starts fetching and playing the music, block by block.

### 7.4. Internet Radio

Once it became possible to stream audio over the Internet, commercial radio stations got the idea of broadcasting their content over the Internet as well as over the air. Not so long after that, college radio stations started putting their signal out over the Internet. Then college students started their own radio stations. With current technology, virtually anyone can start a radio station. The whole area of Internet radio is very new and in a state of flux, but it is worth saying a little bit about.

There are two general approaches to Internet radio. In the first one, the programs are pre-recorded and stored on disk. Listeners can connect to the radio station's archives and pull up any program and download it for listening. In fact, this is exactly the same as the streaming audio we just discussed. It is also possible to store each program just after it is broadcast live, so the archive is only running, say, half an hour, or less behind the live feed. The advantages of this approach are that it is easy to do, all the techniques we have discussed work here too, and listeners can pick and choose among all the programs in the archive.
The other approach is to broadcast live over the Internet. Some stations broadcast over the air and over the Internet simultaneously, but there are increasingly many radio stations that are Internet only. Some of the techniques that are applicable to streaming audio are also applicable to live Internet radio, but there are also some key differences.

One point that is the same is the need for buffering on the user side to smooth out jitter. By collecting 10 or 15 seconds worth of radio before starting the playback, the audio can be kept going smoothly even in the face of substantial jitter over the network. As long as all the packets arrive before they are needed, it does not matter when they arrived.

One key difference is that streaming audio can be pushed out at a rate greater than the playback rate since the receiver can stop it when the high-water mark is hit. Potentially, this gives it the time to retransmit lost packets, although this strategy is not commonly used. In contrast, live radio is always broadcast at exactly the rate it is generated and played back.

Another difference is that a live radio station usually has hundreds or thousands of simultaneous listeners whereas streaming audio is point to point. Under these circumstances, Internet radio should use multicasting with the RTP/RTSP protocols. This is clearly the most efficient way to operate.

In current practice, Internet radio does not work like this. What actually happens is that the user establishes a TCP connection to the station and the feed is sent over the TCP connection. Of course, this creates various problems, such as the flow stopping when the window is full, lost packets timing out and being retransmitted, and so on.

### 7.5. Voice over IP

Once upon a time, the public switched telephone system was primarily used for voice traffic with a little bit of data traffic here and there. But the data traffic grew and grew, and by 1999, the number of data bits moved equaled the number of voice bits (since voice is in PCM on the trunks, it can be measured in bits/sec). By 2002, the volume of data traffic was an order of magnitude more than the volume of voice traffic and still growing exponentially, with voice traffic being almost flat (5% growth per year).

As a consequence of these numbers, many packet-switching network operators suddenly became interested in carrying voice over their data networks. The amount of additional bandwidth required for voice is minuscule since the packet networks are dimensioned for the data traffic. However, the average person's phone bill is probably larger than his Internet bill, so the data network operators saw Internet telephony as a way to earn a large amount of additional money without having to put any new fiber in the ground. Thus Internet telephony (also known as voice over IP), was born.

### 7.6. Video on Demand

Video on demand is sometimes compared to an electronic video rental store. The user (customer) selects any one of a large number of available videos and takes it home to view. Only with video on demand, the selection is made at home using the television set's remote control, and the video starts immediately. No trip to the store is needed. Needless to say, implementing video on demand is a wee bit more complicated than describing it. In this section, we will give an overview of the basic ideas and their implementation.

Is video on demand really like renting a video, or is it more like picking a movie to watch from a 500-channel cable system? The answer has important technical implications. In particular, video rental users are used to the idea of being able to stop a video, make a quick trip to the kitchen or bathroom, and then resume from where the video stopped. Television viewers do not expect to put programs on pause.

If video on demand is going to compete successfully with video rental stores, it may be necessary to allow users to stop, start, and rewind videos at will. Giving users this ability virtually forces the video provider to transmit a separate copy to each one.

On the other hand, if video on demand is seen more as advanced television, then it may be sufficient to have the video provider start each popular video, say, every 10 minutes, and run these nonstop. A user wanting to see a popular video may have to wait up to 10 minutes for it to start. Although pause/resume is not possible here, a viewer returning to the living room after a short break can switch to another channel showing the same video but 10 minutes behind. Some material will be repeated, but nothing will be missed. This scheme is called near video
on demand. It offers the potential for much lower cost, because the same feed from the video server can go to many users at once. The difference between video on demand and near video on demand is similar to the difference between driving your own car and taking the bus.

Watching movies on (near) demand is but one of a vast array of potential new services possible once wideband networking is available. The general model that many people use is illustrated in Figure 7.11. Here we see a high-bandwidth (national or international) wide area backbone network at the center of the system. Connected to it are thousands of local distribution networks, such as cable TV or telephone company distribution systems. The local distribution systems reach into people's houses, where they terminate in set-top boxes, which are, in fact, powerful, specialized personal computers.

7.11. ábra - Figure 7.11: Overview of a video-on-demand system

Attached to the backbone by high-bandwidth optical fibers are numerous information providers. Some of these will offer pay-per-view video or pay-per-hear audio CDs. Others will offer specialized services, such as home shopping (letting viewers rotate a can of soup and zoom in on the list of ingredients or view a video clip on how to drive a gasoline-powered lawn mower). Sports, news, reruns of "I Love Lucy," WWW access, and innumerable other possibilities will no doubt quickly become available.

Also included in the system are local spooling servers that allow videos to be placed closer to the users (in advance), to save bandwidth during peak hours. How these pieces will fit together and who will own what are matters of vigorous debate within the industry. Below we will examine the design of the main pieces of the system: the video servers and the distribution network.

To have (near) video on demand, we need video servers capable of storing and outputting a large number of movies simultaneously. The distribution network is the set of switches and lines between the source and destination. Usually, the backbone is switched and the local distribution network is not. The main requirement imposed on the backbone is high bandwidth. It used to be that low jitter was also a requirement, but with even the smallest PC now able to buffer 10 sec of high-quality MPEG-2 video, low jitter is not a requirement anymore. Local distribution is highly chaotic, with different companies trying out different networks in different regions. Telephone companies, cable TV companies, and new entrants, such as power companies, are all convinced that whoever gets there first will be the big winner. The four main local distribution schemes for video on demand go by the acronyms ADSL, FTTC, FTTH, and HFC. We will now explain each of these in turn.
ADSL is the first telephone industry's entrant in the local distribution sweepstakes. The idea is that virtually every house in the United States, Europe, and Japan already has a copper twisted pair going into it (for analog telephone service). If these wires could be used for video on demand, the telephone companies could clean up.

The problem, of course, is that these wires cannot support even MPEG-1 over their typical 10-km length, let alone MPEG-2. High-resolution, full-color, full motion video needs 4-8 Mbps, depending on the quality desired. ADSL is not really fast enough except for very short local loops.

The second telephone company design is FTTC (Fiber To The Curb). In FTTC, the telephone company runs optical fiber from the end office into each residential neighborhood, terminating in a device called an ONU (Optical Network Unit). On the order of 16 copper local loops can terminate in an ONU. These loops are now so short that it is possible to run full-duplex T1 or T2 over them, allowing MPEG-1 and MPEG-2 movies, respectively. In addition, videoconferencing for home workers and small businesses is now possible because FTTC is symmetric.

The third telephone company solution is to run fiber into everyone's house. It is called FTTH (Fiber To The Home). In this scheme, everyone can have an OC-1, OC-3, or even higher carrier if that is required. FTTH is very expensive and will not happen for years but clearly will open a vast range of new possibilities when it finally happens. What do you think about each member of the family operating his or her own personal television station? ADSL, FTTC, and FTTH are all point-to-point local distribution networks, which is not surprising given how the current telephone system is organized.

A completely different approach is HFC (Hybrid Fiber Coax), which is the preferred solution currently being installed by cable TV providers. The story goes something like this. The current 300- to 450-MHz coax cables are being replaced by 750-MHz coax cables, upgrading the capacity from 50 to 75 6-MHz channels to 125 6-MHz channels. Seventy-five of the 125 channels will be used for transmitting analog television.

7.7. The MBone – The Multicast Backbone

While all these industries are making great — and highly publicized — plans for future (inter)national digital video on demand, the Internet community has been quietly implementing its own digital multimedia system, MBone (Multicast Backbone).

MBone can be thought of as Internet television. Unlike video on demand, where the emphasis is on calling up and viewing precompressed movies stored on a server, MBone is used for broadcasting live video in digital form all over the world via the Internet. It has been operational since early 1992. Many scientific conferences, especially IETF meetings, have been broadcast, as well as newsworthy scientific events, such as space shuttle launches. A Rolling Stones concert was once broadcast over MBone as were portions of the Cannes Film Festival. Whether this qualifies as a newsworthy scientific event is arguable.

Technically, MBone is a virtual overlay network on top of the Internet. It consists of multicast-capable islands connected by tunnels. In this figure, MBone consists of six islands, A through F, connected by seven tunnels. Each island (typically a LAN or group of interconnected LANs) supports hardware multicast to its hosts. The tunnels propagate MBone packets between the islands. Some day in the future, when all the routers are capable of handling multicast traffic directly, this superstructure will no longer be needed, but for the moment, it does the job.

Each island contains one or more special routers called mrouters (multicast routers). Some of these are actually normal routers, but most are just UNIX workstations running special user-level software (but as the root). The mrouters are logically connected by tunnels. MBone packets are encapsulated within IP packets and sent as regular unicast packets to the destinationmrouter's IP address.

Tunnels are configured manually. Usually, a tunnel runs above a path for which a physical connection exists, but this is not a requirement. If, by accident, the physical path underlying a tunnel goes down, the mrouters using the tunnel will not even notice it, since the Internet will automatically reroute all the IP traffic between them via other lines.

8. Questions

1. What is the TCP/IP reference model?
2. What types of IP service do you know?
3. Describe the IP datagram format!
4. What is the domain name system?
5. What are the main internet services?
6. What is the server/client model?
7. What is a Virtual Private Network?
8. What are the main features of the Web 3.0 solutions?
9. Describe the IP address structure!
8. fejezet - Information Systems in Agriculture

1. Introduction

There is a high demand for agricultural information at global, regional, national and community/sub-national levels. However, agricultural information/data is currently collected and kept/archived by different stakeholders leading to problems of overlapping and duplications, among others. Hence the need for an integrated information system.

Effective adoption of Information and Communication Technologies (ICT) now has a proven record in many parts of the world and a demonstrated potential to attain significant economic, social and environmental benefits at local, national and global levels. The past four decades have witnessed numerous attempts to understand the mechanisms of the Adoption of Technological Innovation.

2. Information systems and communication networks

Agricultural information can be seen as an important factor which interacts with the other production factors such as land, labour, capital and managerial ability. The productivity of these other factors can arguably be improved by the relevant, reliable and useful information and knowledge. Hence, the information supply from extension, research, education and others has become managed by agricultural organizations, and especially disseminated to farmers so that they can make better decisions to take advantage of market opportunities and to manage continuous changes in their production systems. Therefore, there is a need to understand the functioning of the particular agricultural information systems in order to manage and improve them. However, there have been limited studies about the agricultural information systems and especially communication networks for farmers. Thus, there is a need for substantial information about these issues, including the mechanisms of the information systems, interactions between components in the system, and their activity. Specifically, the information requirements of farmers, the structure of the organizations involved in these activities are issues that need to be explored.

The definitions of the terms used in this research, such as information, system, information system, agricultural information system and communication networks, are first presented and discussed. The rationale of the system theory and the agricultural information system approach is described.

Information is structured data within a context that gives it meaning. Information can be processed, generated, transformed and shared, through complex processes of coding and decoding, generally known as communication. The communication of information is a major concern for the agricultural extension services. A system is a group of interacting components, operating together for a common purpose. It is characterized in terms of its hierarchical structure, emergent properties, communication and control. The term subsystem is equivalent to system, but it is contained within a larger system. The system approach is a way of looking at an entity and dealing with problems in order to identify and improve the particular system. It can be applied to any subject. The system research has also shown a high potential for offering a conceptual framework to analyze, manage and improve the current system and to design a better one. The models of social systems can be used as a tool for analyzing the information requirements of actors involved in a system defines information systems as: “…deal with the deployment of information technology in organizations, institutions, and society at large”.

In the general system theory, an information system is accepted as a system, automated or manual, that comprises people, machines, and/or methods organized to collect, process, transmit, and disseminate data which represent information. Processed information becomes knowledge when an individual knows (understands) and evaluates it. Thus, the knowledge system is more an individual basis and emphasizes the personal cognition. However, groups of people may share a common knowledge system such as an indigenous knowledge system.

Information systems are also social systems whose behaviour is heavily influenced by the goals, values and beliefs of the individuals and groups, as well as the performance of the technology. Information behaviour as: “...the totality of human behaviour in relation to sources and channels of information, including both active and passive information seeking, and information use. Thus, it includes the face-to-face communication with others,
as well as the passive reception of information as in, for example, watching the TV advertisements, without any intention to act on the information given”.

Accordingly, the concept of agricultural information system reflects the components in the system, the information related processes (generation, transformation, storage, retrieval, integration, diffusion and utilization), system mechanisms (interfaces and networks) and system operations (control and management). Research, extension and farmer can be seen as the major components (subsystems) of an agricultural information system. However, various actors and organizations can be found in a system. It can be applied to any specific farming systems in order to analyze how the information system works. This approach is also useful to define the possible defaults and to improve the coordination between components (i.e. information management). In addition, the information exchange (communication) through networks among the system components is critically important for the successful technology generation and information transfers. A communication network consists of interconnected individuals who are unlinked by the patterned flows of information, and its analysis identifies the communication structure in the system, the exchange of information (communication) and its diffusion take place within a social system.

3. Management Information Systems

Management information systems encompass a broad and complex topic. To make this topic more manageable, boundaries will be defined. First, because of the vast number of activities relating to management information systems, a total review is not possible. Those discussed here is only a partial sampling of activities, reflecting the author's viewpoint of the more common and interesting developments. Likewise where there were multiple effects in a similar area of development, only selected ones will be used to illustrate concepts. This is not to imply one effort is more important than another. Also, the main focus of this paper will be on information systems for use at the farm level and to some lesser extent systems used to support researchers addressing farm level problems (e.g., simulation or optimization models, geographic information systems, etc.) and those used to support agribusiness firms that supply goods and services to agricultural producers and the supply chain beyond the production phase.

Secondly, there are several frameworks that can be used to define and describe management information systems. More than one will be used to discuss important concepts. Because more than one is used, it indicates the difficult of capturing the key concepts of what is a management information system. Indeed, what is viewed as an effective and useful management information system is one environment may not be of use or value in another.

Lastly, the historical perspective of management information systems cannot be ignored. This perspective gives a sense of how these systems have evolved, been refined and adapted as new technologies have emerged, and how changing economic conditions and other factors have influenced the use of information systems.

Before discussing management information systems, some time-tested concepts should be reviewed. Davis offers a commonly used concept in his distinction between data and information. Davis defines data as raw facts, figures, objects, etc. Information is used to make decisions. To transform data into information, processing is needed and it must be done while considering the context of a decision. We are often awash in data but lacking good information. However, the success achieved in supplying information to decision makers is highly variable. Barabba, expands this concept by also adding inference, knowledge and wisdom in his modification of Haechel's hierarchy which places wisdom at the highest level and data at the lowest. As one moves up the hierarchy, the value is increased and volume decreased. Thus, as one acquires knowledge and wisdom the decision making process is refined. Management information systems attempt to address all levels of Haechel's hierarchy as well as converting data into information for the decision maker. As both Barabba and Haechel argue, however, just supplying more data and information may actually be making the decision making process more difficult. Emphasis should be placed on increasing the value of information by moving up Haechel's hierarchy.

Another important concept from Davis and Olsen is the value if information. They note that “in general, the value of information is the value of the change in decision behavior caused by the information, less the cost of the information.” This statement implies that information is normally not a free good. Furthermore, if it does not change decisions to the better, it may have no value. Many assume that investing in a “better” management information system is a sound economic decision. Since it is possible that the better system may not change decisions or the cost of implementing the better system is high to the actual realized benefits, it could be a bad
investment. Also, since before the investment is made, it is hard to predict the benefits and costs of the better system, the investment should be viewed as one with risk associated with it.

Another approach for describing information systems is that proposed by Harsh and colleagues. They define information as one of four types and all these types are important component of a management information system. Furthermore, the various types build upon and interact with each other. A common starting level is Descriptive information (Figure 8.1) information portrays the “what is” condition of a business, and it describes the state of the business at a specified point in time. Descriptive information is very important to the business manager, because without it, many problems would not be identified. Descriptive information includes a variety of types of information including financial results, production records, test results, product marketing, and maintenance records.

8.1. ábra - Figure 8.1: Types of Information

Descriptive information can also be used as inputs to secure other needed types of information. For example, “what is” information is needed for supplying restraints in analyzing farm adjustment alternatives. It can also be used to identify problems other than the “what is” condition. Descriptive information is necessary but not completely sufficient in identifying and addressing farm management problems.

The second type of information is diagnostic information. This information portrays this “what is wrong” condition, where “what is wrong” is measured as the disparity between “what is” and “what ought to be”. This assessment of how things are versus how they should be (a fact-value conflict) is probably our most common management problem. Diagnostic information has two major uses. It can first be used to define problems that develop in the business. Are production levels too low? Is the rate earned on investment too low? These types of question cannot be answered with descriptive information alone (such as with financial and production records). A manager may often be well supplied with facts about his business, yet be unable to recognize this type of problem. The manager must provide norms or standards which, when compared with the facts for a particular business, will reveal an area of concern. Once a problem has been identified, a manager may choose an appropriate course of action for dealing with the problem (including doing nothing). Corrective measures may be taken so as to better achieve the manager’s goals. Several pitfalls are involved for managers in obtaining diagnostic information. Adequate, reliable, descriptive information must be available along with appropriate
norms or standards for particular business situations. Information is inadequate for problem solving if it does not fully describe both “what is” and “what ought to be”.

As description is concerned with “what is” and diagnostics with “what is wrong”, prediction is concerned with “what if...?” Predictive information is generated from an analysis of possible future events and is exceedingly valuable with “desirable” outcomes. With predictive information, one either defines problems or avoids problems in advance. Prediction also assists in analysis. When a problem is recognized, a manager will analyze the situation and specify at least one alternative (including doing nothing) to deal with it. Predictive information is needed by managers to reduce the risk and uncertainty concerning technology, prices, climate, institutions, and human relationships affecting the business. Such information is vital in formulating production plans and examining related financial impacts. Predictive information takes many forms. What are the expected prices next year? What yields are anticipated? How much capital will be required to upgrade production technologies? What would be the difference in expected returns in switching from a livestock farm to a cropping farm? Management has long used various budgeting techniques, simulation models, and other tools to evaluate expected changes in the business.

Without detracting from the importance of problem identification and analysis in management, the crux of management tasks is decision making. For every problem a manager faces, there is a “right” course of action. However, the rightness of a decision can seldom, if ever, be measured in absolute terms. The choice is conditionally right, depending upon a farm manager’s knowledge, assumptions, and conditions he wishes to impose on the decision. Prescriptive information is directed toward answering the “what should be done” question. Provision of this information requires the utilization of the predictive information. Predictive information by itself is not adequate for decision making. An evaluation of the predicted outcomes together with the goals and values of the manager provides that basis for making a decision. For example, suppose that a manager is considering a new marketing alternative. The new alternative being considered has higher “predicted” returns but also has higher risks and requires more management monitoring. The decision as to whether to change plans depends upon the managers evaluation of the worth of additional income versus the commitment of additional time and higher risk. Thus, the goals and values of a farm manager will ultimately enter into any decision.

The importance of management information systems to improve decision making has long been understood by farm management economists. Financial and production records have long been used by these economists as an instrument to measure and evaluate the success of a farm business. However, when computer technology became more widely available in the late 1950s and early 1960s, there was an increased enthusiasm for information systems to enhance management decision processes. At an IBM hosted conference, Ackerman, a respected farm management economist, stated that:

“The advances that have taken place in calculating equipment and methods make it possible to determine the relationship between ultimate yields, time of harvest and climatic conditions during the growing season. Relationship between the perspective and actual yields and changing prices can be established. With such information at hand the farmer should be in a position to make a decision on his prediction with a high degree of certainty at mid-season regarding his yield and income at harvest time.”

This statement, made in 1963, reflects the optimism that prevailed with respect to information systems. Even though there was much enthusiasm related to these early systems they basically concentrated on accounting activities and production records. Examples include the TelFarm electronic accounting system at Michigan State University and DHIA for dairy operations.

These early systems relieved on large mainframe computers with the data being sent to a central processing center and the reports send back to the cooperating businesses. To put these early efforts into a management information system framework, the one proposed by Alder (House, ed.) is useful (Figure 8.2). They would be defined as data oriented systems with limited data analysis capabilities beyond calculating typical ratios (e.g., return on assets, milk per cow, etc.).

8.2. ábra - Figure 8.2: Types of Information Systems
By the mid 1960s it became clear that the accounting systems were fairly effective in supplying descriptive and diagnostic information but they lacked the capacity to provide predictive and prescriptive information. Thus, a new approach was needed – a method of doing forward planning or a management information system that was more model oriented. Simulation models for improving management skills and testing system interaction were developed. As an example, Kuhlmann, Giessen University, developed a very robust and comprehensive whole farm simulation model (SIMPLAN) that executed on a mainframe computer. This model was based on systems modeling methods that could be used to analyze different production strategies of the farm business. To be used by managers, however, they often demanded that the model developer work closely with them in using the model.

Another important activity during this period was the “Top-Farmer Workshops” developed by Purdue University. They used a workshop setting to run large linear-programming models on mainframe computers (optimization models) to help crop producers find more efficient and effective ways to operate their business.

As mainframe timeshare computers emerged in the mid-1960's, it became possible to remotely access the computer with a terminal and execute software. Systems such TelPlan developed by Michigan State University made it possible for agricultural producers to run a farm related computer decision aids. Since this machine was shared by many users, the cost for executing an agriculturally related decision aid was relatively inexpensive and cost effective. These decision aids included optimization models (e.g., least cost animal rations) budgeting and simulation models, and other types of decision aids. These decision aids could be accessed by agricultural advisor with remote computer terminals (e.g., Teletype machine or a touch-tone telephone). These advisors used these computer models at the farm or at their own office to provide advice to farm producers.

These were exciting times with many people becoming involved in the development, testing, refining, and implementation of information systems for agriculture. Computer technology continued to advance at a rapid pace, new communication systems were evolving and the application of this technology to agriculture was very encouraging. Because of the rapid changes occurring, there were international conferences held where much of the knowledge learned in developing these systems was shared. One of the first of these was held in Germany in the mid-1980s.

It was also clear from these early efforts that the data oriented systems where not closely unlinked to the model oriented systems. Information for the data oriented systems often did not match the data needed for the model oriented systems. For example, a cash-flow projection model was not able to directly use financial data contained in the accounting system. In most cases, the data had to be manually extracted from the accounting system and re-entered into the planning model. This was both a time consuming and error prone process.

Because of the lack of integration capabilities of various systems, they were devoid of many of the desirable characteristics of an evolving concept describes as decision support systems (DSS). These systems are also
known as Executive Support Systems, and Management Support System, and Process Oriented Information Systems. The decision support system proposed by Sprague and Watson (House, ed.) has as its major components a database, a modelbase, a database/modelbase management system and a user interface (Figure 8.3). The database has information related to financial transactions, production information, marketing records, the resource base, research data, weather data and so forth. It includes data internally generated by the business (e.g., financial transactions and production data) and external data (e.g., market prices). These data are stored in a common structure such that it is easily accessible by other database packages as well as the modelbase.

The modelbase component of the system has decision models that relate to operational, tactical and strategic decisions (Figure 8.3). In addition, the modelbase is able to unlink models together in order to solve larger and more complex problems, particularly semi-structured problems. The database/modelbase management system is the bridge between database and modelbase components. It has the ability to extract data from the database and pass it to the modelbase and vice versa. The user interface, one of the more critical features of the system, is used to assist the decision maker in making more efficient and effective use of the system. Lastly, for these systems to be effective in supporting management decision, the decision maker must have the skills and knowledge on how to correctly use these systems to address the unique problem situation at hand.

8.3. ábra - Figure 8.3: Main Components of Decision Support System

Several follow-up international conferences were held to reflect these new advances in management information systems. The first of these conferences focused on decision support systems was held in Germany. This conference discussed the virtues of these systems and the approach used to support decisions. Several prototype systems being developed for agriculture were presented. From these presentations, it was clear that the decision support systems approach had many advantages but the implementation in agriculture was going to be somewhat involved and complex because of the diversity of agricultural production systems. Nevertheless, there was much optimism for the development of such systems.

A couple of years later, another conference was held in Germany that focused on knowledge-based systems with a major emphasis on expert systems and to a lesser extent optimum control methods and simulation models. Using Alter’s scheme to describe information systems, for the most part these would be described as suggestion models. It was interesting to note that the prototype knowledge-based systems for the most part did not utilize the concepts of decisions support systems which was the focus of the earlier conference. Perhaps this was related to the fact that many of the applications were prototypes.
The international conference that followed in France focused on the low adaption rate of management information systems. This was a topic of much discussion but there were few conclusions reached except the systems with the highest adaption rate were mainly data-oriented ones (e.g., accounting systems, field record systems, animal production and health records, etc.) which provide mainly descriptive and diagnostic information.

The international conferences that followed had varying themes. One of the major themes was precision agriculture with several conferences held. These conferences extolled the use of geographic information systems (GIS) in conjunction with geographic positioning systems (GPS) to record and display data regarding cropping operations (e.g., yields obtained) and to control production inputs (e.g., fertilizer levels). Other conference addressed the use of information systems to more tightly control agriculture production such as those developed for greenhouse businesses.

To briefly summarize the historical developments, there have been significant efforts devoted to improving the management information systems from the early computerized activities forty years earlier. The decision aids available have grown in number and they are more sophisticated. There has been some movement toward integration of the data oriented systems and the model oriented systems. An examination of our current usage of management information systems, however, suggests that we have not nearly harnessed the potential of the design concepts contained modern management information systems.

4. Internal Information Systems

The current status of management information systems is remains dynamic. Several adoption surveys and personal experiences lead to some interesting observations. These observations will be reviewed in the context of a decision support system as defined by Spraque and Watson.

4.1. On-Farm Information Systems – Computer Hardware

The percentage of farms owning a computer continues to grow. Most commercial farms now own a computer and have access to the Internet, many with high speed connections. Most of the computers are of recent vintage with large data storage and memory capacity. It is safe to state that the hardware is not the bottleneck with respect to management information systems.

4.2. On-Farm Database and Modelbase Applications

The decision support system literature stressed that the database and modelbase remain separate entities. They should be bridged by the database/modelbase management system. In examining much of the software developed for on-farm usage, it appears that most of it does not currently employ this design concept. Indeed most of the software is a stand-alone product with the database an integral part of the modelbase. However, some packages have the ability to export and import data, allowing for the sharing of data across the various packages, but these data sharing features are usually rather narrow in scope and flexibility.

The most common software packages used by agricultural producers are data oriented with the most common being one designed for financial accounting. Accounting packages explicitly designed for agricultural businesses and general business accounting packages are used for keeping the financial records. Because of their rather low cost relative to the agricultural specific packages, the general purpose packages are growing in market share. These financial accounting systems are used beyond completing tax documents. They are also important for providing information to creditors and for planning and control.

Production management also accounts for a significant proportion of computer usage. There are many software packages available that address livestock problems. Some are database programs to keep track of animal related data and/or feed inventories. There are models to address operational and tactical decisions such as ration balancing, culling decisions, alternative replacements options, etc.

However, many livestock producers also use off-farm production records processing such as using the DHIA service bureau for processing dairy records. These service bureaus provide a downloading feature so the data can be moved to the on-farm computer.
For cropping operations, there are similarities in software availability. Database systems are available for keeping track of information on fields and sub-fields, particularly fertilizers and pesticides applied, varieties planted and yields achieved.

Though there is increasing interest in geographic information systems by agricultural producers, the main usage is for yield monitoring and mapping. This approach is used to evaluate the effectiveness of alternative management practices employed in the production of the crop (e.g., comparison of varieties, seeding rates, pest control measures, tillage systems, etc.) and to identify field problems (e.g., soil compaction, drainage problems, etc.). This yield monitoring approach is finding the greatest acceptance and this may be in part because the yield monitoring and mapping systems are common option on grain harvesting equipment. One of the real concerns with using yield monitoring and mapping systems relates to the issue of arriving at the correct inference of what causes the variation in yields noted. The potential layers of data (e.g., pH, precious crops grown, soil structure, planting date, nutrients applied, variety grown, pesticides used, rainfall, etc.) has been suggested to exceed 100. To be able to handle the large number of data layers in an effective manner would suggest a full-feature geographic information system (GIS) might be needed. However, few agricultural producers have access to a full-feature GIS and/or training to utilize these systems, and there are substantial costs related to capturing and storing various data layers. Nevertheless, the more obvious observations originating from these systems (e.g., such as poor drainage and soil compaction) have resulted in sound investments being made in corrective measures.

To a limited extent, some agricultural producers are starting to make use of remote sensing data to identify problems related to the growing crop such as an outbreak of a disease. Those using remote sensing feel they are able to more quickly identify the problems and take corrective action, minimizing the damage done.

Precision agriculture applied to the animal industries is on a different scale. Information systems are playing a major role on the integrated mega-farms. When using information systems to carefully track genetic performance, balance rations, monitor health problems, facilities scheduling, control the housing environment and so forth, it is generally acknowledged that it is possible to achieve a fairly significant reduction in cost per unit of output (10-15%) over that of more traditional, smaller farming operations. These are proprietary information systems and the information from these systems are used to gain a strategic competitive advantage.

Lastly, the general purpose spreadsheet is the most common software used for planning purposes. Some of these applications are very sophisticated and address complex problems.

### 4.3. User Interface

The user interface has improved in greatly in quality. Most agricultural software now uses the windowing environment. This environment makes it easier for the user to use and access data and information, and to move data from one application to another or to ulink applications. However, this still remains a user-initiated task and in some cases can be complex. Also most of the data contained in the software package is unique to that package and not easily shared with other software packages. Thus, from a DSS viewpoint there are still significant shortcomings.

### 4.4. The Decision Maker

An often overlooked component of a decision support system is the decision maker. Prior surveys suggest that the primary user of the on-farm computer system is the farm operator. Operators that are younger and college educated were much more likely to routinely use the computer. Also large farms were more likely to utilize a computer in their farming operation. It is also observed that there is a fair amount of “learning cost” related to use of on-farm information systems. These cost can be large enough to hinder the adoption of management information systems.

### 5. External Information Systems

There is increased interest and excitement about the role external information systems available to agricultural producers, particularly Internet and satellite data transmission systems. Each of these technologies is a vast resource of data which can be used to enhance the various levels (e.g., information, intelligence, knowledge, wisdom) of Haechel's Hierarchy for an individual or organization.
Another information source is the outside advisor. As the complexity and breadth of the farm level decision process has increased, the use of consultants and advisors has grown. This is particularly true of the larger farming operations.

5.1. Internet

The growth in Internet is phenomenal. The growth in its use by agricultural producers is also phenomenal. Email is a common communication tool used by agricultural business. The same is true for the world-wide-web (WWW). They made extensive use of the web to find information that fit their unique requirement. Even though they find it a major source of information for their operation, it takes good skills to locate the information desired. One of the common complaints is the amount of time it takes to utilize the Internet effectively and the lack of depth of information. One of the critical questions relates to how effective Internet is in addressing the higher levels of Haechel’s hierarchy.

Other Internet resources available to agriculture include sites for downloading agricultural software. Much of the economic data compiled by the government is now available on-line. Lastly, in some cases it is being used as a marketing tool for products produced by the business.

5.2. Satellite Data Transmission Systems

The satellite data transmission systems are widely used by producers. These systems are passive data acquisition systems from the user’s viewpoint. Data is continuously broadcast to the leased data terminal from a satellite. The data is automatically stored in the data terminal and can be accessed by a menuing process. These systems provide current data/information on a number of topics. Amounts and types of data/information received depends upon the options purchased. The basic subsystem provides for the latest market prices and news, weather maps (e.g., rainfall, jet streams, severe weather, crop soil moisture index, soil temperature, air temperature, etc.), government reports on market developments, long- and short-term weather forecasts, political developments that pertain to agriculture, and product information. Premium service options add even more features.

5.3. Outside Advisors

Several recent studies suggest that use of outside advisory services by farmers to enhance and supplement their on-farm information systems was fairly prevalent. The tax preparer is the advisory most commonly used. Other important sources of information include the local Extension agents, veterinary consultants, accountants, crop/pest management consultants, and livestock management advisors (e.g., a nutritionist).

The outside advisors utilize many different software packages to help provide advice to producers. FINPAK developed by the University of Minnesota is an example of a software package widely used by outside advisors with farmers. This financial analysis and related projection package helps evaluate the financial process being made by the farm and compares alternative future business options. This package (an accounting type model) is widely used in the U.S.

6. The Future

Predicting the future is not an exact science. But with the structural changes occurring in agriculture today, the management problems are significantly different from the problems of yesterday. Earlier emphasis in information systems was on improving production management decisions. Today, major issues that are commonly faced in management relate to financial, human resource, and marketing management. These management areas and their importance are identified in the strategic management workshops I have conducted with agricultural producers. Thus, managers will have less time to address production issues because more time and effort are being focused in the other management areas. This will have an impact on information systems to address production management.

6.1. Structured Decisions

In the future information systems to address production management will likely be of five general types: 1) software for systems analysis, 2) theory testing, software for teaching purposes, 3) software for advisors, 4) software for use by producers, and 5) software to control and monitor the supply chain.
Software for systems analysis and theory testing will be developed with the primary objective of defining the structure and studying the dynamics and interaction of the various system components. Its main use is in research. These models are fairly complex and often have robust data requirements. Their utilization often depends upon availability of the developers to run the model or assist in the use of the model. This software is very useful in testing various hypotheses regarding system dynamics (e.g., would supplemental irrigation in the early growth stages greatly affect yields)?

These models play a vital role in generating a better understanding of the overall system and can give valuable insight on how to manage the system. They are also useful in identifying areas for further research. The results from these models are communicated in various ways (e.g., journal articles, trade journals, and advisory service publications and conferences) and these communicated results are often used by producers to adjust production practices. However, direct use by producers to evaluate their own unique situations is not common with these models. There are several reasons for this limited use including a poor user interface or lacking the data to drive the model. Also, it is generally unlikely that transformation of a model of this nature into one that is to be used by the producers will be successful.

Software developed for teaching purposes is likely to continue. Sometimes these software packages are referred to as simulation games. Because these models teach concepts and principles, they are often a simplification of reality. They tend to use the case analysis approach, making it difficult to use the model to analyze various options and alternatives utilizing actual business data. The models are often used in an interactive mode (e.g., in a classroom or workshop environment) where knowledge is gained by testing “what if” questions, then observing the results. These models can be very powerful teaching tools, but are rarely used to analyze actual business situations. Producers often lose interest in using this software because it is too simplistic, takes too much time and effort to extract knowledge for better decision-making, or it does not adequately reflect the reality of the business.

Software for advisors is a class of software that is used by agricultural advisors (e.g., Extension staff, consultants, and agribusiness firms) to assist producers in making decisions. The advisor is a necessary intermediary, because the software could demand a thorough understanding of a difficult set of concepts (e.g., long range planning) or it may be rather demanding of the user’s time and effort (e.g., a large amount of data has to be collected, entered and analyzed), or the time and effort to become proficient in the use of the model is considered excessive. This type of software will grow in importance as the use of outside consultants and advisory services by agricultural businesses grows. These outside advisors and consulting services will increasingly use many different software packages to help provide advice to the producer. The package they use depends upon their area of specialization. For instance, those that are offering production advise may use one of several production decision aid models.

Advisors also serve as an intermediary to extracting information from Internet (external data). They often subscribe to threaded discussion groups. They use these groups for posting problems and receiving back suggested solutions. They also learn from the exchange of ideas between others using the system. Also, advisors more readily see the merit of using a software program designed for systems analysis for enhancing their personal knowledge and skills and solving problems for their clients. This is particularly true if the software has a good user interface.

Software for use by producers is and will continue to be some of the most demanding software to develop. As indicated earlier, a large amount of software has been written, but much of it has fallen short of expected usage rate. One reason is the decision makers have found the software fails to address their problems. The software must be fairly easy to utilize, and the producer expects it to provide information that has a perceived value greater than the cost of attaining that information.

Software being used by producers can be grouped into two subcategories. The first subcategory is used to process transaction data and meet regulatory requirements. These are the software applications most used by the actual businesses. They must keep accounting, personnel and crop production records (e.g., pesticides used) because of government regulation. They also use software to reduce the time, effort and cost of processing the transaction records. This is why payroll packages, and shipping and billing systems are commonly employed on these operations. This usage will continue to grow in importance.

The other subcategory of software is used for management purposes. This currently accounts for a lesser portion of the computer usage. A large growth in this usage of this software is unlikely. The time and effort to master this software is major commitment. Since management time is being diverted to areas other than production management, they will have less and less time to become proficient in the use of this software. Thus, very
thorough and sophisticated systems (e.g., the SAP software system) currently being employed by large companies are not likely to be common on farm businesses because of their complexity and cost.

Software for process control is used to control and automate many of the structured-operational decisions of the business enterprises, such as controlling temperature, light, irrigation and fertility in greenhouses. These models are generally of a closed-loop optimal control design. The process control models are generally knowledge based systems and have been developed using knowledge from many sources including the systems analysis models discussed earlier. The use of process control systems will grow in importance and acceptance. This acceptance implies that the managers have confidence in the models and that they improve the efficiency and effectiveness of the business. These models also free them to concentrate on more complex decisions.

Software to control and monitor the supply chain will greatly grow in importance. The will be many factors driving this growth including concerns about food safety, country of origin labeling, organic foods, foods to meet special dietary requirements, and concerns about product liability suits. In will likely become commonplace that a food item purchased by the consumer at the retail level will have attached its entire history, including identity preservation and traceability, included with the purchase. The new advances in RFID chips and the requirements by certain major retailers to label all products with these chips will impact agricultural businesses including those engaged in producing farm products. The system imposed upon the entire supply chain will likely be designed by the retailers and the entire chain will need to adjust to the defined information structure. To adapt to the defined information structure may mean a major restructuring of the information system currently being used by the business with substantial costs associated with the conversion.

6.2. Structured and Unstructured Decisions

To address the management areas related to human resources, finances, and marketing, suggest information systems that can address ill-structured or unstructured problems. Some would state that we are in the process of moving from the “old economy” to “new economy”. With this paradigm shift, among the changes is a movement from resource based to idea based wealth creation, from a stable comparative advantage to a dynamic one, from investment in physical assets to investment in human capital, from protected to open markets, from subsidies to encouragement to adapt, from hierarchal organizations to strategies alliances and partnerships. In addition agriculture will move from commodity markets to product markets and it will become more environmentally friendly, concerned with food safety, and quality and supply coordination.

If this transition from the “old economy” to the “new economy” occurs for agriculture, then the information systems of the past will not be adequate for the future. They will need to be much broader and more comprehensive than the current systems. The future systems must:

• address the larger scope of financial management rather than financial record keeping, tax reporting, and analysis;
• help define marketing strategies and alliances;
• help identify potential niche markets rather than supplying data on current commodity market trends;
• support the creation of new ideas;
• nurture the growth of knowledge since this will become a major source of wealth creation;
• deal with the many dimensions and complexity of human resource management;
• signal needed production changes in an overall system of supply chain management;
• assist in negotiating contractual arrangements;
• help the producer adopt to an economic climate that has more risk and uncertainty because of less government intervention in markets;
• provide the capacity to track the identity of a product from its genetics to the consumer;
• assist in producing a product that meets customer desires rather than the production of a commodity.
Developing farm-level information systems to fulfill these needs will be a major challenge. It will take a major rethinking with regard to the role of management information systems. It will involve more than enhancing hardware, communications infrastructure, and software components of the information system. An equally important consideration will be the analytical skills, knowledge, wisdom, and interests of the agricultural decision maker.

The information system of the future will need to concentrate more on the upper levels of Haechel's hierarchy – knowledge and wisdom. As Honaka and Hirotaka observe, knowledge has two forms, tacit (subjective) and explicit (objective). Tacit knowledge is gained from experiences and practice, whereas explicit knowledge is based more on theory and rationality. As decision makers address problems, they convert knowledge between the two forms. An information system that focuses only on one form will have shortcomings. The information system of the future must have both forms of knowledge, and encourage the conversion of knowledge between the forms as a continuous process. Only by this process will the manager's knowledge base grow in size and function.

Information systems of the past have tended to concentrate on explicit knowledge (e.g., linear programming to balance a ration) and, to lesser extent tacit knowledge. Many of the problems of the future will involve tacit knowledge. The challenge will be designing information systems that will allow for an easier and more effective means of sharing tacit knowledge. The Internet will no doubt play a key role in meeting this challenge. Perhaps a system for documenting experiences (e.g., structured case studies) can be used to enhance the sharing of tacit knowledge.

6.3. Summary

Agriculture has a long and proud past history in applying information systems including farming operations. Although there have been significant strides forward in improving the decision making of farm managers there are still areas for improvement. The decisions of the future will be different from those of the past. There will be no quick and easy solutions on how to design the farm information system of the future. Indeed, each farm business will likely have its own unique system that has been tailored to meet the special informational requirement of the farm business and address the needs of the entire supply chain. Those that are able to build and effectively utilize the farm information systems of the future will have a strategic advantage over their competitors.

7. Farm Management Information System (FMIS)

Agriculture and farmers face a great challenge in effectively manage information both internally and externally in order to improve the economic and operational efficiency of operations, reduce environmental impact and comply with various documentation requirements. In order to meet this challenge, the flow of information between decision processes must be analyzed and modeled as a prerequisite for the subsequent design, construction and implementation of situated information systems.

A modern farm management system offers the potential to fundamentally alter agricultural decision-making. The use of large machinery and hired labour has caused many farmers to think of large fields as the basic management unit. Information technologies permit the modern grower to obtain detailed explicit information at a small scale common to farming practices of earlier times but with considerably more information, enabling them to efficiently manage the land at these finer scales.

The basic principle of modern knowledge management allows for more accurate production management. The whole process requires a big amount of data to be collected; this data enables control of the whole process and also introduces information about the situation outside the farm (economical information). For better understanding, to all this process is necessary to improve access to this data and make analysis of this data. Mathematical and statistical analysis and usage of spatial information retrieved from this data can bring new quality to the whole process of future farming. The real end-users of the technologies are not only farmers and agriculture managers, but also advisors and eventually service organizations. Part of knowledge could also be available for the food industry, market and direct consumers. The limitation of better utilization of knowledge is, unfortunately, their limited effective sharing of knowledge. It is necessary to improve access to these data and the possibility of using new information sources.

An integrated FMIS can support real-time management decisions and support reporting as well as application generation and improvement of the monitoring of compliance to management standards.
In the future, European farmers will have to effectively manage information on and off their farms to improve economic viability and to reduce environmental impact. All three levels, in which agricultural activities need to be harmonized with economical and environmental constraints, require integrated ICT adoption: (i) improvement of farm efficiency; (ii) integration of public goods provided by farming into management strategies; (iii) relating to the environmental and cultural diversity of Europe’s agriculture by addressing the region- farm interaction. In addition, the communication between agriculture and other sectors needs improvement. Crop products for the value added chains must show their provenance through a transparent and certified management strategy and farmers receiving subsidies are requested to respect the environment through compliance of standards.

To this end, an integration of information systems is needed to advise managers of formal instructions, recommended guidelines and implications resulting from different scenarios at the point of decision making during the crop cycle. This will help directly with making better decisions as the manager will be helped to be compliant at the point and time of decision making.

The potential of adoption of Precision Farming can be divided in four classes: very high, high, medium and low. The very high potential is focussing on the central parts of Western Europe, especially the north half of France, the east coast of England and Scotland, south Sweden, Denmark, the north and east of Germany and a few regions in the Netherlands and Belgium, one region in the Czech Republic and one in Spain. The high potential is also focussing on the countries in the north and west, while the medium and low potential is predominantly located in the periphery of Europe, like the Atlantic coast and the Mediterranean, and with a few exceptions in the new member states of central and east Europe.

7.1. Knowledge Management in Agriculture

A modern farm management system offers the potential to fundamentally alter agricultural decision- making. The use of large machinery and hired labour has caused many farmers to think of large fields as the basic management unit. Information technologies permit the modern grower to obtain detailed explicit information at a small scale common to farming practices of earlier times but with considerably more information, enabling them to efficiently manage the land at these finer scales.

The basic principle of modern knowledge management allows for more accurate production management. The whole process requires a big amount of data to be collected; this data enables control of the whole process and also introduces information about the situation outside the farm (economical information). For better understanding, to all this process is necessary to improve access to this data and make analysis of this data. Mathematical and statistical analysis and usage of spatial information retrieved from this data can bring new quality to the whole process of future farming. The real end-users of the technologies are not only farmers and agriculture managers, but also advisors and eventually service organizations. Part of knowledge could also be available for the food industry, market and direct consumers. The limitation of better utilization of knowledge is, unfortunately, their limited effective sharing of knowledge. It is necessary to improve access to these data and the possibility of using new information sources.

The FutureFarm project aims to build a model of a farm management information system (FMIS), which will introduce new knowledge management based on new system architecture. This architecture could be expressed by the formula:

Farm management is usually divided into the following steps:

- Data capturing (soil, crop, yield, costs)
- Data analysis and data processing
- Data access
- Management
- Decision and application
- Data archive

The Knowledge management methods analysis of existing software or services on the market demonstrated that existing solution could be divided from two different points of view:
On which part of farm management is focused

Platform

From point of view of farm management focus we can define:

- **Macro level** – systems including large area and covering more farms
  - Market analysis – crop forecast, markets needs
  - Traceability systems – production process control
  - Subsidies system – crop production support
  - Weather forecast – area management

- **Farm level** – systems managing knowledge inside of one farm
  - Economy – cost calculation, investment
  - Crop rotation
  - Decision supporting systems
  - Weather forecast – farm action plan
  - Traceability system

- **Micro level** – system managing knowledge inside of single fields
  - Data collection and data analysis – robotics or semi robotics system
  - Soil treatment system – field variable rate application (ploughing, other tillage)
  - Crop treatment system – field variable rate application (seedling, crop feeding, crop protection, harvest)
  - Water management – irrigation

From the platform point of view there exist few groups of solutions:

- Software packages, usually desktop or PDA running on machinery or eventually fields systems, which are focused mainly on supporting one rate application or variable rate application (precision farming). This software is usually produced by agricultural machinery producers as an added value of those machines for machines operator use.

- Large deployed software (Desktop), which cover all agriculture production, or important part of agriculture production for farmer use.

- Web based application, with focus on specific functionality, like support for LPIS, or system for traceability or central controlling system.

- Open Web Services based system combining Web based application with desktop system, data collection system and field machines management solution.

The Knowledge management user requirements analysis demonstrated needs for some way to share selected part of information, in specific conditions. So there is needed to build some collaborative environment, not necessary Web based. The Collaborative Environment (CE) is a new concept in agriculture. The current organisation of the work in agriculture is based on distinct services. At the moment there isn’t an open solution that is able to support the collaboration among actors along the whole value added chain. The services required are:

- To support cooperation between farmers, advisors (agriculture service organizations, market needs and trends), Universities (Research institutes).
• To build new concept of farming consultancy based on collaborative work and sharing of common workplace between farmers, advisors (agriculture service organizations, market needs and trends), Universities (Research institutes), when all organization could share selected part of data. Collaboration between advisors and Universities (Research institutes) is on different level of collaboration, but research is done on real farm data and result is delivered to farmer through advisors mainly.

The common workplace opens possibilities for service organisation to provide effective support and services. Data sharing between farmers and advisor provide better data treatment and analysis of data. Farmers will have opportunity to use results of analysis trough Internet (using Web services), but also trough mobile equipment directly in the field. Important question is to trust farmers, that the information, which will be shared will depend on them and will not be accessible by their competitors, but also by other organization, which could influence farm.

So far the biggest limitation of agriculture knowledge management systems is interoperability. To share information among macro, farm and micro level is very difficult. Also there is not optimal using of information from different knowledge management levels for decision making. It is clear, that changes on the market have and will have critical influences on the farming sector. To be possible to adapt farms on changing situation and to optimise any decision, we need both directional transfer of information and knowledge. We have consider, that from economical point of view, most important decision has to be done on farm level, but we will need for this a lot of information from macro level, which will have most important influence on economy of farm as a whole. Long time sustainability and environment protection will be influenced mainly by micro level. But for different strategies, we will need on micro level also input from farm level.

Due the grooving importance of cooperation and more collaboration and needs for knowledge sharing, goal of the FutureFarm knowledge management has to be based on virtual common workspace where selected part of knowledge, but also decision support tools could be shared and which has to be connected with knowledge management system. The solution has to be secure. There are too option. First will be, that the data has to be fully managed on farm side and only selected knowledge will be shared, or if part of data and information could be managed for example service organization. For sharing of parts of data and information, there is necessary guarantee security of transaction but also trust all information of single farms, which could have influence on competitiveness of the farm on the market.

The goal of the FutureFarm project is to build an open system, which will be able to manage all levels of incoming information on all levels and will be able to support optimal decision taking by farmers. Such kind of solution has to be able to cover all necessary manipulation with data, which will be able to present this data in the form of useful information and which will support effective methods of knowledge management, which could increase the effectiveness of precision farming.

This system will be universal; it will be able to support:

• Data access
• Data collection by farmers and services companies
• Data analysis
• Advisory services preparing recommendation for farmers
• Effective managing directly by farmers
• Using information by service companies
• Future research on the field of farm management
• Training.

Knowledge management offers a great opportunity for farmers to grow on the market, to be more flexible as far as market needs are concerned and to develop new products or services for their production. Information systems are a crucial part of Knowledge management referring to all the interactive technology needed to store and manage information. Farmers have a wide scale of possibilities to acquire knowledge in their daily decision making process within the farm. This also means that information can be processed in many different ways by farmers or computers, depending on the best suitable way for every farmer. Generally, differences exist...
regarding the tradition in each country, the structure of the support provided by the government to the farmers, the educational sector, the market services and others. In the near and far future there is a significant development and growing influence of the last two mentioned. These external drivers play and will play an important role in the adoption of Knowledge management in the farm. Educational sector, farm services and farms increasingly use computer and internet connection not only for data processing, but also for communication with other people through the internet. Basically, farmers need to be innovative in order to find the best path for growing their opportunities and improving their business processes in the market.

7.2. Adoption of Knowledge Management

Generally, adoption of Information technologies in KM for the majority of the farmers in EU in the future will bring changes in their traditions. An open and receptive culture must be nurtured, especially by managers and leaders. Farmers should be willing to share ideas and experiences, and to disseminate and accept new knowledge. Discussion and socialisation should be underlined.

The cultural change can be enhanced by incentives given to those adopting and practicing knowledge sharing. This can be either salary increase or some other facilities that the employees require. Furthermore, a culture which accepts mistakes due to experimentation must be enforced.

Individual farmers could produce knowledge through observation of their own farms. Farm managers can be asked to write reports about new projects, keep record of their progress, and create in this way a small and cost-effective database. It is necessary to make the knowledge acquisition possible through networking, access to journals, conference proceedings, and to databases. Employees at the farm have to be supported by training; this could include visits to competitors, job rotation and/or apprenticeship. Managers have to participate in discussion forums and interest groups. It is necessary to organize experience swapping sessions, conferences, exhibitions, and seminars with external speakers; distribute the results; and organize informal meetings or social activities of any type. For accomplishing adoption of new knowledge, it is important to involve advisors innovation centers and universities in the farm.

Adoption of knowledge management will not be homogeneous in Europe. Also, it is not expected that all farms will adopt knowledge management methods immediately. There are differences between countries, but also in different areas within the same countries. The crucial aspects for the adoption strategy are:

- economical aspects – given by the structure of farming sector (scale, products);
- social (age and education of farmers);
- cultural – different tradition in farming sectors in single countries.

All these aspects have to be included in the Future farm roadmap. Another crucial question would be which is going to be the target market for the Farm Management Information Systems. In principle there are two ways of implementing such a solution, and different strategies need to be used for each of them:

- deliver software as final product;
- offer not software but knowledge management service (Software as a Service – SaaS).

The experiences from different countries demonstrate, that both ways could be used and hence should be considered.

Knowledge management in agricultural production is adopted on many different levels by using Information technology and professional services supported by Universities and/or service organizations (of the public or the private sector). Precision farming plays important role on this adoption. Farmers involved to precision faming technologies are more flexible to work with computer or to use high technology level in crop production, animal production or farm management.

Every farmer wishes to increase his productivity, increase yield in crop production and keep a good quality of production. Still, the conditions on the field are not homogeneous! Precision farming systems are able to monitor the farm condition with a GPS system for crop production and, after a data analysis, to prepare a variable rate application for specific farm zones and specific time. So, farmers wish to monitor field conditions by data collection and analysis, and then prepare fertilizer in variable rate and apply fertilizer according to soil conditions, nutrient content of the soil and crop needs.
The description of Knowledge management – precision farming tools adoption on the farm start step by step as follows:

- The farmer, who expects to profit from the above mentioned ideas, must decide which one is the most suitable for his farm.
- Farmer does not need to make a big investment in the first year, and then depreciate it part by part every year and wait for result.
- Precision farming tools can be adopted on the farm in a timeframe of three to five years, but first results must be seen from the first year of adoption.

Today, many precision farming tools and information technology systems exist, which are capable of presenting more or less practical results. A central farm database must be established and the information needed for the decision making process should be stored there. Practically, a farmer will use a computer not only for calculations or for managing some work processes, but also for finding important information on the internet.

Workshop discussion revealed some still very strong farmers’ arguments against the adoption of KM on the farm. First of all, many farmers are skeptical of farm Knowledge Management with the use of information technology. From the farmer point of view, his first goal is the stability of the farm business by using common tools. New technology comes next, but they prefer to investigate how they could improve their business by using modern machinery than by using modern systems of Knowledge Management on the farm.

- Today, especially farmers from central and eastern Europe, struggle for surviving in the market. Their attention is mainly focused on market needs and their efforts on increasing their profits and profit margins.
- Concerning Knowledge Management, different approaches are used depending on the production type (food, energy and bio- fuel, sport- culture), definition and focus.
- The way a farm is run (whether it is run by a family or by the owner himself) influences the way that the new generation of farmers perceive the farm production, hence the new methods of work that they try to introduce.
- New generation of farmers in Europe and US have differences in their approach about profit and profit generation.
- Energy efficiency resulting from bio- mass utilization supports the development of new technologies with higher profit.
- Farming in different regions varies in the level and the strategies followed to reach development.
- A system of farm knowledge management based on information technologies will have a positive impact when adopted in almost any type of farm.
- Technology characteristics such as user- friendliness and simplicity in use and technologies such as touch screens and wireless data transfer are easy accepted by any farmer generation.

Farmers anticipate new developments in services either by farm consulting companies or by new robots. Small and medium sized companies supplying such services are expected to appear in the market.

Besides that, the agricultural community expects some measures from the state government concerning the importing and exporting regulations. Some kind of import barriers such as, for example, taxes on fertilizers and chemicals in different locations would stimulate the use of KM on the farms.

An open and receptive culture must be nurtured, especially inspired and motivated by managers and leaders. Farmers should be willing to share ideas, experiences, to share and embrace new knowledge. Discussions and socialisation should be a priority. The culture shift can be facilitated with incentives given to those adopting and implementing knowledge sharing. This can be an increase in salary or some other facilities that the employees require. Also, a culture which accepts the possibility of mistakes instead of the safe and ordinary way of doing things as it leads to innovation through experimentation must be enforced.

8. Farm Management Information System
Although most people can see the benefits of using a more precise approach to manage crops with additional information, the tools provided by precision farming and other information technologies have not yet moved into mainstream agricultural management. The increased complexity of the systems inhibits easy adoption and makes calculations as to the financial benefits uncertain. These issues can be resolved by improving the decision making process through better Management Information Systems, improved data interchange standards and clear management methods.

The FutureFarm project’s starting point has been the identification of the current and future data, information and knowledge management needs on the farms, as well as on the way that these needs will evolve in the future and that will influence farm data, farm information and farm knowledge management systems. Existing systems were categorized and evaluated through interviews with the project’s pilot farms.

Farm Management Information System (FMIS) specification was produced, by using a user-centric approach. The system boundaries were identified as well as the farmer’s personal management strategies. The integration in the FMIS of information coming from online soil sensors was used as an integration case study. The architecture of the proposed system is based on the Service Oriented Architecture. The main characteristic of such an architecture is that it allows to different publishers to develop components of the FMIS which can then be integrated to it with the use of a common vocabulary. The concept of the assisting services for the future FMIS was defined. Actors and information flows, usage processes and data elements for the FMIS have been modelled and analysed, and functional requirements of FMIS have been determined. The outlined system elements and requirements are very complex and diverse depending on the farm production type, level of automation and inherent business processes. When looking to the future, external services as decision making assisting features will become an important part of FMIS concept. At the moment, the utilisation of scientific models together with the large amounts of data in different formats produced by modern farm machinery, sensors located within the farm, remote sensing, etc. is still an open area of research and new methods are developed continuously. The seamless incorporation of new functionality and assisting features into an existing FMIS is of paramount importance.

An analysis of selected agricultural standards resulted in a methodology on how (and under which conditions) these standards could be stored in a machine readable format. Then, the software architecture as well as a prototype system for automated agricultural standards retrieval (and evaluation) was produced. Although specific problems still need to be solved, whether this system will be utilized or not is mainly a political question.

Further investigation is required in order to find out how automated retrieval of agricultural rules and standards can be adopted by the agricultural sector in Europe. Also, developing autonomous and visual crop detection and crop modeling in order to model nitrogen response and weed development in combination with the water response functions is now required in order to prove the advantages of the use of precision farming technologies. The use of semantics is inevitable for an open service oriented FMIS, but therefore the development of a common ontology language for the agricultural sector in Europe is required.

Precision farming was seen within the project as a technology that demands the development of information systems in agriculture. Therefore, the strategies in which farmers communicate and cooperate in the adoption of precision agriculture were identified as well as the precision farming potential of the EU areas. The most prominent precision farming technology to be used in the near future was found to be control traffic farming on the basis of its economic returns. The highest precision farming adoption potential have areas on the central parts of Western Europe.

The specifications of a farm’s portal from the external stakeholders point of view, revealed that the history of the farm, information about the producers in the form of curriculum vitae, farm location, climatic and soil conditions and, last but not least, farming practices, is the information that the consumers would like to see in it. Farmers would also like to be able to market other farm services through the portal, in the case of a multifunctional farm.

The consortium believes that further ICT developments in agriculture will include the development of agricultural robotics in collaboration with advanced FMIS systems.

8.1. FutureFarm project basic idea

Developing codes of good farming practice, diversifying markets and production systems as well as European standards of sustainable agricultural production systems require implementation of more elaborate management
strategies. These have to respect specific ecological conditions, demands from the rural regions and those from the value-added chains. On top of that, these strategies have to be simple, but flexible enough to be adapted easily to changing economic or environmental conditions and they need proof of their compliance. Beyond that, the demand for information about the production processes is growing, both from the perspective of the value-added chains (traceability) as well as from regional stakeholders in order to fulfill multifunctional objectives by farming. An important prerequisite for farmers to comply with all these different demands is to easily have sufficient and timely information available for decision making or providing documentary evidence. The rapid development of technologies for information and communication, new sensors as well as the vast potentials for providing geo-referenced data (remote-sensing, on-line sensors, public databases etc.) also allows farmers to access new and high quality data and use them as specific information in decision making or process documentation. With automated data acquisition and handling in an on-farm management information system the farmers can be seen to comply with a rapidly growing demand of standards in the management of the production processes.

Precision Farming (PF) in Europe uses new technologies in information handling and management as well as in managing the spatial and temporal variability found on all farms. Such explicit information use improves economic returns and reduces environmental impact. Precision farming is very data intensive and historically unlinked with site specific activities and management on the field. It has become very clear in recent years that PF is not limited to site-specific farming. The use of techniques and methods that form precision farming can provide a wealth of information and tools to handle and apply information properly for any type of farm in any region. This information-driven approach can be used to help improve crop management strategies and proof of compliance through documentation.

The introduction of advanced ICT technologies into agriculture will also be a significant progress in all efforts for measurements oriented payments within agro-environmental programs and related efforts to enforce environmentally sound systems in land use within the EU. This also includes the Best Management Practice according to the cross compliance scheme.

Crop products going into the food chain must show their certified provenance through a recognised management strategy and subsidy payments to farmers are now unlinked to respect of the environment through compliance to standards. To this end, an integration of information systems is needed to advise managers of formal advice, recommended guidelines and implications resulting from different scenarios at the point of decision making during the crop cycle. This can be achieved by integrating real-time modelling (a crop growth and development model unlinked to sensors within the growing canopy), with expert systems that have been configured with the guidelines from a recommended management strategy (e.g. organic, ICM, IPM, factored risk etc.) as well as legal guidance (such as health and safety and environmental protection). This will directly help the farm or crop manager to make better decisions. Expert knowledge in the form of models and expert systems can be published and made available in a machine readable form on the internet or made available as web-services to be dynamically bound into the end-user software. As the relevant farm data is already in the proposed information system, or may be automatically integrated using standardised services, documentation in the form of instructions to operators, certification of crop province and cross compliance of adopted standards can be generated more easily than with current systems.

Crop products can also stay on the farm – besides traditionally fodder this will be in the future the internal use of biofuels or bio energy. That would boost the possibility of moving towards a highly energy-efficient or even energy-neutral farm. This is supported by the significant reduction of energy required by small smart machines that can work by themselves while intelligently targeting inputs.

**8.2. A Farm Management Information System specification: The farm seen as a system**

The agricultural production sector faces increased pressure in terms of reduced margins of earnings. Farmers are constantly required to reduce production cost, maximise their physical output while maintaining the highest product quality. These requirements go hand in hand with adherence to strict environmental, social, health, and safety regulations (e.g., certification schemes such as International Food Standard (IFS) and GLOBALGAP). The use of information and communication technology (ICT) and farm management information systems (FMIS) and decision support has shown great promise for achieving the above goals, especially in the context of precision agriculture. The most important requirements for a Farm Management Information System (FMIS) include: a) a design aimed at the specific needs of the farmers, b) a simple user-interface, c) automated and simple-to-use methods for data processing, d) a user controlled interface allowing access to processing and
analysis functions, e) integration of expert knowledge and user preferences, f) improved integration of standardized computer systems, g) enhanced integration and interoperability, h) scalability, i) interchange-ability between applications, and j) low cost. A dedicated design of a FMIS complying with the above requirements involves an identification and specification of the scope and boundaries, an identification of system components (actors, decision processes, etc.) combined with information modelling, and finally, as part of the overall knowledge management, an identification of knowledge content in decision processes and functional requirements.

The development of a conceptual model for an effective FMIS based on information derived from pilot farms representing diverse conditions across the EU. The conceptual model was divided into four sections: internal data collection, external information collection, plan generation and report generation. The data collection and processing is an automated monitoring system, whereas the report and plan subsystems are to be initiated by the farm manager. The external repository contains information on standards, rules, all types of guidelines for farm activities etc., made available for the FMIS. This conceptual model is the first step towards the actual design of a novel FMIS.

The information flows and relevant input data were given for the strategic, tactical, and operational planning levels for field operations, together with the execution and evaluation phases. The information flow definitions included the actors and decision processes involved in the overall operation, and specified which.

8.3. The Farm Management Information System

Management information systems (MIS) is an integral part of the overall management system in an purposeful organisation comprising tools like enterprise resource planning (ERP), overall information systems (IS), etc. ERP is an industry notion for a wide set of management activities which support all essential business processes within the enterprise. As a part of the ERP, the information system (IS) refers to data records and activities that process the data and information in an organization, and it includes the organization's manual and automated processes supporting the business processes. Information systems are the software and hardware systems that support data-intensive applications. Especially, information systems provide the possibility to obtain more information in “real-time” enabling a close monitoring of the operations performance and enhance the connection between executed operations and the strategic targets of the enterprise.

Management Information Systems differ from regular information systems because the primary objectives of these systems are to analyse other systems dealing with the operational activities in the organization. In this way, MIS is a subset of the overall planning and control activities covering the application of humans, technologies, and procedures of the organisation. Within the field of scientific management, MIS is most often tailored to the automation or support of human decision making. In this way was also seen the Farm Management Information System within Futurefarm.

8.4. Conceptual Model of a Farm Management Information System (FMIS)

Increased amount of needed and used information in precision agriculture and increasing use of automation in farm machinery force farmers to utilize external services to support the decision making as well as to use machine readable communication between.

8.4.1. Assisting services within or outside the farm

The FMIS as defined earlier, is seen as a system of systems supporting farmer’s decision making. The concept of assisting services has already been defined in the previous chapter. In this chapter, work done within the project in order to develop prototypes for selected assisting services is presented.

In order for the FMIS to evolve as a decision support tool, which utilizes a wealth of information to assist the farmer in his decision-making, the issue of integration needs to be addressed. Within the FutureFarm project, except of specifying the FMIS requirements and architecture, there has been an attempt to specify an architecture, which will allow the import of data acquired through precision farming recording technologies into the FMIS. Also, another architecture is defined to facilitate the automated exchange of agricultural regulations and standards.

8.4.2. On-line data acquisition and the FMIS
As a specific assisting service for the proposed FMIS architecture, the automatic import to the FMIS of data coming from online soil sensors was studied. Whilst online instruments now exist to measure essential parameters on soil and crops, challenges remain in integrating these data into the FMIS in an efficient manner. These challenges stem from the fact that online collected data are with different formats e.g. images, spreadsheet xls, xlsx, proprietary binary, csv, etc.

Generally, online sensors consist of a sensing technology e.g. near infrared spectroscopy, a sensing unit e.g. optical probe or load cell, a digital global positioning system (DGPS) to record position, a software and laptop to record and store data in different formats. These are collected on a user specific designed platform, which is in most cases unlinked to the three point un linkage of a tractor to record data on soil, crop cover, crop disease, weed intensity, etc. However, online sensors can also be unlinked or installed on commercial agricultural machinery such as yield sensors on combine harvesters.

Current data communication standard for tractors and machinery in agriculture is ISO 11783, which is rather well established and has gained market acceptance. However, there is a significant number of non-ISO-11783 compliant online sensors in practice. With the ISO 11783 standard, data on parameters related to tractor and machine performance, e.g. speed, draught, moment, etc. are managed, whereas different formats of data collected with non-ISO 11783 sensors are discussed.

With the former case, process data from sensors with CAN interface is converted into ISO 11783 XML and then imported into relational database at FMIS using RelaXML tool. There is also the export function database to task controller (TC) to provide task management as described in ISO 11783:10. In the latter case, the import service is based on local or public sharing or semantic mapping outputting AgroXML format for FMIS. Import is best performed as close to the generation of sensor data as possible to maximise the availability of metadata.

Summarizing the recommendations of this case study:

- Where possible, any producer of a third party file should be upgraded to include agroXML output.
- A second choice is to provide a format exporter, which is incorporated in the workflow to be used by the original operator-integrated with the end of day or data transfer process.
- A next choice is to include the function as an import filter at the time data is taken into a farm-office system.
- Last choice is that the data is encapsulated in a raw format and incorporated later.

Automated unification will take place through a common data dictionary (semantics), an ontology based approach. It should also be possible to provide each translator with a manual interface to allow definition of the semantic mapping for the first import between systems. This mapping should be storable in the destination device to allow future translations to be entirely automated. The experience of other industries suggests such mappings can be productively shared in a community. As currently found in the industry, it is suggested that import processes are likely to remain to some extent manufacturer specific or based on de-facto industry standards. There is however the opportunity for third parties to produce translation layers for incorporation in machine controllers or FMIS PC environments where original manufacturers are unwilling or unable to provide a solution.

8.4.3. The new model of a Farm Management Information System

A new model and prototype of a new Farm Information Management System (FMIS) integrating information systems to advise managers of formal instructions, recommended guidelines and documentation requirements for various decision making processes has been designed. The proposed FMIS meets the requirements of future and current European farmers in terms of enhancing their managerial tasks as related to economic viability and the interaction with the surroundings. This research has shown the impact of using dedicated system analysis methodologies as a preliminary step to the actual design of a novel farm management information system compared with other more rigid and activity oriented system analysis methods. Also, it has been shown that the use of the soft system approach allows a fundamental analysis, incorporating the identification of required changes and most importantly, the unstructured analysis enables the identification of existing constraints, and possible solutions, which may not be apparent using more structured methods.

Complementary to the model identification of the proposed FMIS, the information flows for targeted field operations has been presented through a user-centric approach. The information models are centred on the farmer as the principal decision maker and involve external entities as well as mobile unit entities as the main
information producers. This is a detailed approach to information modelling that enables the specification and documentation for the generation and implementation of a Farm Management Information System in crop production.

### 8.4.4. The proposed Service Oriented Architecture for the future FMIS

Within the project, the prototypic implementation of concepts for a flexible and modern Farm Management Information System was made by using state of the art techniques. These techniques have the potential to serve as basis for the creation of distributed applications in the agricultural domain. One aspect during the development process was to choose the concept of a Service-Oriented Architecture, which consists of different parties with Service Providers, Service Brokers and Service Consumers.

In the use case “Integrating Knowledge for Compliance to Agricultural Management Standards” the Service Consumer is mainly seen in the user of an FMIS (e.g. the Farmer) but is not restricted to one party. Service Providers in this case are the Management Standard Setters, namely governmental institutions or publishers of private management standards. Agricultural Service Brokers help to find the Services by Service Consumers, e.g. agricultural organisations or similar could be appropriate for this part. The Service Providers agree on a common data exchange format for the specific service as it was proposed by FutureFarm. As the FutureFarm Partners expect various services which are based on knowledge and handled by rules in future, the machine-readable encoding of agricultural standards consists well known terms from domain agricultural and rules to describe what is allowed and forbidden by the management standards. The terms might be described in suitable formats (e.g. W3C OWL) and in freely accessible repositories in order to keep them reusable by various future applications. To test the proposed architecture, a prototype including example data sets was developed and disseminated via the FutureFarm website (http://test.futurefarm.eu).

All of these prototype services have a RESTful interface as this web service design paradigm seemed to be appropriate for this kind of services and makes it easy for developers of Farm Management Information Systems to integrate such services in their client software. Therefore, the necessary libraries for the .Net Framework have been developed to give developers of FMIS for the Microsoft Windows Platform easy access to the services.

The main impact was to attract developers and experts from the agricultural domain to use such a service in further applications. A first draft prototype of a Rules Interpreter Component was developed, enabling FMIS to automatically check compliance to the management standards served by the proposed architecture. Feedback has been received by showing a high strategic potential of such services, but there have to be taken a lot of efforts on the organizational and administrative side to practice such a generic service.

### 8.4.5. Energy on Farm

On-farm productions of bio-fuels offers a lot of potential advantages like the independence from oil shortages possible in the future, an important contribution of the farms in reducing CO2 emissions and contribute to the targets set by EU and an additional income to the farmers and rural areas an important factor to maintain farming communities in rural areas. Assuming that direct use of vegetable oils is technically and practically possible, an energy self sufficient farm can be developed. Based on the crop yields achieved and the oil extraction by cold pressing efficiency a part of the farm ranging from 6 to 12% devoted to energy crops can cover the liquid fuel requirements of a farm.

Besides on-farm biofuel production, reduction of on-farm energy use is a way to improve the sustainability of arable farming. Under conditions in Northern Europe, ploughing is amongst the main direct energy consuming activities. The use of artificial fertilizers ranks high on the list of indirect energy consumption. No-tillage farming seems a suitable alternative. Under south European conditions irrigation is besides ploughing the energy consuming activity. As water is a strongly limiting factor under these circumstances, suitable mitigation strategies will be hard to develop.

Fleet management in arable farming still relies strongly on human supervision and decision making. Yet, with growing farms, growing machine fleets and growing associated costs of mechanisation, effective use of these resources is required. Fleet management techniques based on mathematical optimization principles may lead to significant reduction in time and energy use. Energy savings in the order of 10% are expected. Automatic fleet management will require more and automated information exchange and data processing. As significant computing power is needed and this currently exceeds commonly available computing power of current board computers, computing for fleet management may require a centralized approach. Management strategies are
generated on a central computer and distributed amongst the individual vehicles in the fleet through the wireless network. Current developments in wireless networking and networking configurations support the developments in fleet management.

A next step in farming might be or, to put it more strongly, will be the replacement of human operated machines by robots. Though this technology is not yet mature, progress is considerable and will change arable farming in the coming decades.

9. Questions

1. What are the levels of farm management?
2. What does the Integrated Farm Management System mean?
3. What are the main modules of an IFMS?
4. What are the farm planning elements?
5. What is the precision agriculture?
9. fejezet - Geographical Information Systems

1. Introduction to GIS

A GIS a geographic information system as a computerized system that facilitates the phases of data entry, data analysis and data presentation especially in cases when we are dealing with georeferenced data. This means that a GIS user will expect support from the system to enter (georeferenced) data, to analyse it in various ways, and to produce presentations (maps and other) from the data. Many kinds of functionality should come with this: support for various kinds of coordinate systems and transformations between them, many different ways of ‘computing’ with the georeferenced data, and obviously a large degree of freedom of choice in presentation parameters such as colour scheme, symbol set, medium et cetera. Information system is specializing in the input, storage, manipulation, analysis and reporting of geographical (spatially related) information (Figure 9.1).

9.1. ábra - Figure 9.1: The GIS functions

1.1. Spatial data and geoinformation

Another subtle difference exists between the terms data and information. Most of the time, we use the two terms almost interchangeably, and without the risk of being ambiguous. Occasionally, however, we need to be precise and then their distinction matters. By data, we mean representations that can be operated upon by a computer. More specifically, by spatial data we mean data that contains positional values. Occasionally one will find in the literature the more precise phrase geospatial data as a further refinement, which then means spatial data that is georeferenced. (Strictly speaking, spatial data that is not georeferenced can have positional data unrelated to the Earth’s surface. Examples can be found in molecular chemistry, in which the position of atoms in molecules are defined relative to each other, and in industrial design engineering, in which the parts of a car engine are defined relative to each other.) In this book, we will use ‘spatial data’ as a synonym for ‘georeferenced data’.

By information, we mean data that has been interpreted by a human being. Humans work with and act upon information, not data. Human perception and mental processing leads to information, and hopefully understanding and knowledge. One cannot expect a machine like a computer to ‘understand’ or ‘have knowledge’. Geoinformation is a specific type of information that involves the interpretation of spatial data.

1.2. Applications of GIS

An important distinction between GIS applications is whether the geographic phenomena studied are man-made or natural. Clearly, setting up a cadastral information system, or using GIS for urban planning purposes involves a study of man-made things mostly: the parcels, roads, sidewalks, and at larger scale, sub-urbs and transportation routes are all man-made. These entities often have — or are assumed to have — clear-cut boundaries: we know, for instance, where one parcel ends and another begins.
On the other hand, geomorphologists, ecologists and soil scientists often have natural phenomena as their study objects. They may be looking at rock formations, plate tectonics, distribution of natural vegetation or soil units. Often, these entities do not have clear-cut boundaries, and there exist transition zones where one vegetation type, for instance, is gradually replaced by another.

It is not uncommon, of course, to find GIS applications that do a bit of both, i.e., they involve both natural and man-made entities. Examples are common in areas where we study the effect of human activity on the environment. Rail-road construction is such an area: it may involve parcels to be reclaimed by government, it deals with environmental impact assessment and will usually be influenced by many restrictions, such as not crossing seasonally flooded lands, and staying within inclination extremes in hilly terrain.

A second distinction in applications of GIS stems from the overall purposes of use of the system. A prototypical use of GIS is that of a research project with an explicitly defined project objective. Such projects usually have an a priori defined duration. Feasibility studies like site suitability, but also simulation studies, for instance in erosion modelling, are examples. We call all of these project-based GIS applications.

In contrast to these are what we call institutional GIS applications. They can be characterized in various ways. The life time (duration) of these applications is either indefinite, or at least not a priori defined. Their goal is usually to provide base data to others, not to address a single research issue (Figure 9.2).

9.2. ábra - Figure 9.2: The basic structure of GIS

1.3. The real world and representations of IT

When dealing with data and information we usually are trying to represent some part of the real world as it is, as it was, or perhaps as we think it will be. A computerized system can help to store such representations. We restrict our-selves to ‘some part’ of the real world simply because it cannot be represented completely. The question which part must be represented should be entirely answered through the notion of relevance to the purpose of the computerized system. The fact that we represent the real world only in part teaches us to be humble about the expectations that we can have about the system: all the data it can possibly generate for us in the future must in some way be made available to it first.

In general, a computer representation of some part of the real world, if set-up in a good way, will allow us to enter and store data, analyse the data and transfer it to humans or to other systems. We will now look at setting up real world representations.

1.4. Modelling

‘Modelling’ is a buzzword, used in many different ways and many different meanings. A representation of some part of the real world can be considered a model of that part. We call it such because the representation will have certain characteristics in common with the real world. This allows us to study the representation, i.e., the model, instead of the real world. The advantage of this is that we can ‘play around’ with the model and look at
different scenarios, for instance, to answer ‘what if’ questions. We can change the data in the model, and see what are the effects of the changes. Models — as representations — come in many different flavours. In the GIS environment, the most familiar model is that of a map. A map is a miniature representation of some part of the real world. Paper maps are the best known, but digital maps also exist, as we shall see in. We look more closely at maps below. Another important class of models are databases. A database stores a usually considerable amount of data, and provides various functions to operate on the stored data. Obviously, we will be especially interested in databases that store spatial data.

The phrase ‘data modelling’ is the common name for the design effort of structuring a database. This process involves the identification of the kinds of data that the database will store, as well as the relationships between these data kinds. In data modelling, the most important tool is the data model, and ‘Spatial data modelling’ is a specific type of data modelling.

Maps and databases can be considered static models. At any point in time, they represent a single state of affairs. Usually, developments or changes in the real world are not easily recognized in these models. Dynamic models or process models address precisely this issue. They emphasize changes that have taken place, are taking place or may take place. Dynamic models are inherently more complicated than static models, and usually require much more computation to obtain an intuitive presentation of the underlying processes. Simulation models are an important class of dynamic models that allow to simulate real world processes.

1.5. Maps

The best known (conventional) models of the real world are maps. Maps have been used for thousands of years to represent information about the real world. Their conception and design has developed into a science with a high degree of sophistication. Maps have proven to be extremely useful for many applications in various domains. A disadvantage of maps is that they are restricted to two-dimensional static representations, and that they always are displayed in a given scale. The map scale determines the spatial resolution of the graphic feature representation. The smaller the scale, the less detail a map can show. The accuracy of the base data, on the other hand, puts limits to the scale in which a map can be sensibly drawn. The selection of a proper map scale is one of the first and most important steps in map design. A map is always a graphic representation at a certain level of detail, which is determined by the scale. Map sheets have physical boundaries, and features spanning two map sheets have to be cut into pieces. Cartography as the science and art of map making functions as an interpreter translating real world phenomena (primary data) into correct, clear and understandable representations for our use. Maps also become a data source for other maps. With the advent of computer systems, analogue cartography became digital cartography. It is important to note that whenever we speak about cartography today, we implicitly assume digital cartography. The use of computers in map making is an integral part of modern cartography. The role of the map changed accordingly. Increasingly, maps lose their role as data storage. This role is taken over by (spatial) databases. What remains is the visualization function of maps.

1.6. Databases

A database is a repository capable of storing large amounts of data. It comes with a number of useful functions:

- the database can be used by multiple users at the same time – i.e., it allows concurrent use,
- the database offers a number of techniques for storing data and allows to use the most efficient one – i.e., it supports storage optimization,
- the database allows to impose rules on the stored data, which will be automatically checked after each update to the data – i.e., it supports data integrity,
- the database offers an easy to use data manipulation language, which allows to perform all sorts of data extraction and data updates – i.e., it has a query facility,
- the database will try to execute each query in the data manipulation language in the most efficient way – i.e., it offers query optimization.

Databases can store almost any sort of data. Modern database systems organize the stored data in tabular format. A database may have many tables, each of which stores data of a certain kind. It is not uncommon that a table has many thousands of data rows, sometimes even hundreds of thousands.
1.7. Spatial databases

Spatial databases are a specific type of database. They store representations of geographic phenomena in the real world to be used in a GIS. They are special in the sense that they use other techniques than tables to store these representations. This is because it is not easy to represent geographic phenomena using tables.

A spatial database is not the same thing as a GIS, though they have a number of common characteristics. A spatial database focuses on the functions we listed above for databases in general: concurrency, storage, integrity, and querying, especially, but not only, spatial data. A GIS, on the other hand, focuses on operating on spatial data with what we might call a ‘deeper understanding’ of geographic space. It knows about spatial reference systems, and functionality like distance and area computations, spatial interpolations, digital elevation models et cetera. Obviously, a GIS must also store its data, and for this it provided relatively rudimentary facilities. More and more, we see GIS applications that use the GIS for the spatial analysis, and a separate spatial database for the data storage.

The assumption for the design of a spatial database schema is that the relevant spatial phenomena exist in a two- or three-dimensional Euclidean space. Euclidean space can be informally defined as a model of space in which locations are represented as coordinates – (x, y) in 2D; (x, y, z) in 3D – and notions like distance and direction have been defined, with the usual formulas. In 2D, we also talk about the Euclidean plane.

The phenomena that we want to store representations for in a spatial database may have point, line, area or image characteristics. Different storage techniques exist for each of them. An important choice in the design of a spatial database application is whether some geographic phenomenon is better represented as a point, as a line or as an area. Currently, the support for image data exists but is not impressive. Some GIS applications may even be more demanding and require point representations in certain cases, and area representation in other cases. Cities on a map may have to be represented as points or as areas, depending on the scale of the map.

To support this, the database must store representations of geographic phenomena (spatial features) in a scaleless and seamless manner. Scaleless means that all coordinates are world coordinates given in units that are normally used to reference features in the real world (using a spatial reference system). From such values, calculations can be easily performed and any (useful) scale can be chosen for visualization. A seamless database does not show map sheet boundaries or other partitions of the geographic space other than imposed by the spatial features themselves. This may seem a trivial remark, but early GIS applications had map production as their prime purpose, and considered map sheet boundaries as important spatial features.

All geographic phenomena have various relationships among each other and possess spatial (geometric), thematic and temporal attributes (they exist in space and time). Phenomena are classified into thematic data layers depending on the purpose of the database. This is usually described by a qualification of the database as, for example, a cadastral, topographic, land use, or soil database. A spatial database not only serves to store the data and manipulate it, as it should also allow the users to carry out simple forms of spatial analysis.

Spatial analysis involves questions about the data that relate topological and other relationships. Such questions may involve neighbourhood, distance, direction, incidence, disjointness and a few more characteristics that may exist among geographic phenomena.

1.8. GIS and databases

A database, like a GIS, is a software package capable of storing and manipulating data. This begs the question when to use which, or possibly when to use both. Historically, these systems have different strengths, and the distinction remains until this day.

Databases are good at storing large quantities of data, they can deal with multiple users at the same time, they support data integrity and system crash recovery, and they have a high-level, easy to use data manipulation language. GISs are not very good at any of this.

GIS, however, is tailored to operate on spatial data, and allows all sorts of analysis that are inherently geographic in nature. This is probably GIS’s main stronghold: combining in various ways the representations of geographic phenomena. GIS packages, moreover, nowadays have wonderful, highly flexible tools for map production, of the paper and the digital type. GIS have an embedded ‘understanding’ of geographic space. Databases mostly lack this type of understanding.
The two, however, are growing towards each other. All good GIS packages allow to store the base data in a database, and to extract it from there when needed for GIS operation. This can be achieved with some simple settings and/or program statements inside the GIS. Databases, likewise, have moved towards GIS and many of them nowadays allow to store spatial data also in different ways. Previously, they in principle were capable of storing such data, but the techniques were fairly inefficient.

In summary, one might conclude that small research projects can probably be carried out without the use of a real database. GIS have rudimentary database facilities on board; the user should be aware they are really rudimentary. Midsized projects use a database/GIS tandem for data storage and manipulation. Larger projects, long-term projects and institutional projects organize their spatial data processing around a spatial database, not around a GIS. They use the GIS mostly for spatial analysis and output presentation.

2. Geographic information and Spatial data types

Geographic phenomena exist in the real world: for true examples, one has to look outside the window. In using GIS software, we first obtain some computer representations of these phenomena – stored in memory, in bits and bytes – as faithfully as possible. This is where we speak of spatial data. We continue to manipulate the data with techniques usually specific to the application domain, for instance, in geology, to obtain a geological classification. This may result in additional computer representations, again stored in bits and bytes. For true examples of these representations, one would have to look into the files in which they are stored. One would see the bits and bytes, but very exciting this would not be. Therefore, we can also use the GIS to create visualizations from the computer representation, either on-screen, printed on paper, or otherwise.

It is crucial to understand the fundamental differences between these three notions. The real world, after all, is a completely different domain than the GIS/computer world, in which we simulate the real world (Figure 9.3). Our simulations, we know for sure, will never be perfect, so some facts may not be found.

9.3. ábra - Figure 9.3: Taxonomy of GIS

Crossing the barrier between the real world and a computer representation of it is a domain of expertise by itself. Mostly, it is done by direct observations using sensors and digitizing the sensor output for computer usage. This is the domain of remote sensing. Other techniques for obtaining computer representations are more indirect: we can take a visualization result of a previous project, for instance a paper map, and re-digitize it.

2.1. Geographic phenomena

In the previous chapter, we discussed the reasons for taking GIS as a topic of study: they are the software packages that allow us to analyse geographic phenomena and understand them better. Now it is time to make a more prolonged excursion along these geographic phenomena and to look at how a GIS can be used to represent each of them. There is of course a wide range of geographic phenomena as a short walk through the ITC building easily demonstrates. In the corridors, one will find poster presentations of many different uses of GIS. All of them are based on one or more notions of geographic phenomenon.

2.2. Geographic phenomenon defined

We might define a geographic phenomenon (Figure 9.4) as something of interest that:
• can be named or described,
• can be georeferenced, and
• can be assigned a time (interval) at which it is/was present.

9.4. ábra - Figure 9.4: Spatialy relation, Scale, Time

What the relevant phenomena are for one’s current use of GIS depends entirely on the objectives that one has. For instance, in water management, the objects of study can be river basins, agro-ecologic units, measurements of actual evapotranspiration, meteorological data, ground water levels, irrigation levels, water budgets and measurements of total water use. Observe that all of these can be named/described, georeferenced and provided with a time interval at which each exists.

In multipurpose cadastral administration, the objects of study are different: houses, barns, parcels, streets of various types, land use forms, sewage canals and other forms of urban infrastructure may all play a role. Again, these can be named or described, georeferenced and assigned a time interval of existence.

Observe that we do not claim that all relevant phenomena come as triplets (description, georeference, time interval), though many do. If the georeference is missing, we seem to have something of interest that is not positioned in space: an example is a legal document in a cadastral system. It is obviously somewhere, but its position in space is considered irrelevant.

If the time interval is missing, we seem to have a phenomenon of interest that is considered to be always there, i.e., the time interval is (likely to be considered) infinite. If the description is missing..... we have something funny that exists in space and time, yet cannot be described. (We do not think such things can be interesting in GIS usage.)

2.3. Different types of geographic phenomena

Our discussion above of what are geographic phenomena was necessarily abstract, and therefore perhaps somewhat difficult to grasp. The main reason for this is that geographic phenomena come in so many different ‘flavours’. We will now try to categorize the different ‘flavours’ of geographic phenomena.

To this end, first make the observation that the representation of a phenomenon in a GIS requires us to state what it is, and where it is. We must provide a description – or at least a name – on the one hand, and a georeference on the other hand.
A second fundamental observation is that some phenomena manifest themselves essentially everywhere in the study area, while others only occur in certain localities. If we define our study area as the equatorial Pacific Ocean, for instance, we can say that Sea Surface Temperature can be measured anywhere in the study area. Therefore, it is a typical example of a (geographic) field.

A (geographic) field is a geographic phenomenon for which, for every point in the study area, a value can be determined.

The usual examples of geographic fields are temperature, barometric pressure and elevation. These fields are actually continuous in nature. Examples of discrete fields are land use and soil classifications. Again, any location is attributed a single land use class or soil class.

Many other phenomena do not manifest themselves everywhere in the study area, but only in certain localities. The array of buoys of the previous chapter is a good example: there is a fixed number of buoys, and for each we know exactly where it is located. The buoys are typical examples of (geographic) objects.

(Geographic) objects populate the study area, and are usually well-distinguishable, discrete, bounded entities. The space between them is potentially empty. A general rule-of-thumb is that natural geographic phenomena are more often fields, and man-made phenomena are more often objects. Many exceptions to this rule actually exist, so one must be careful in applying it.

2.4. Geographic fields

A field is a geographic phenomenon that has a value ‘everywhere’ in the study space. We can therefore think of a field $f$ as a function from any position in the study space to the domain of values of the field. If $(x, y)$ is a position in the study area then $f(x, y)$ stands for the value of the field $f$ at locality $(x, y)$. Fields can be discrete or continuous, and if they are continuous, they can even be differentiable.

In a continuous field, the underlying function is assumed to be continuous, such as is the case for temperature, barometric pressure or elevation. Continuity means that all changes in field values are gradual. A continuous field can even be differentiable. In a differentiable field we can determine a measure of change (in the field value) per unit of distance anywhere and in any direction. If the field is elevation, this measure would be slope, i.e., the change of elevation per metre distance; if the field is soil salinity, it would be salinity gradient, i.e., the change of salinity per metre distance.

There are many variations of non-continuous fields, the simplest example being elevation in a study area with perfectly vertical cliffs. At the cliffs there is a sudden change in elevation values. An important class of non-continuous fields are the discrete fields. Discrete fields cut up the study space in mutually exclusive, bounded parts, with all locations in one part having the same field value. Typical examples are land classifications, for instance, using geological classes, soil type, land use type, crop type or natural vegetation type.

A field-based model consists of a finite collection of geographic fields: we may be interested in elevation, barometric pressure, mean annual rainfall, and maximum daily evapotranspiration, and thus use four different fields.

3. Kinds of data values

Since we have now discriminated between continuous and discrete fields, we may also look at different kinds of data values. Nominal data values are values that provide a name or identifier so that we can discriminate between different values, but that is about all we can do. Specifically, we cannot do true computations with these values. An example is the names of geological units. This kind of data value is sometimes also called categorical data.

Ordinal data values are data values that can be put in some natural sequence but that do not allow any other type of computation. Household income, for instance, could be classified as being either ‘low’, ‘average’ or ‘high’. Clearly this is their natural sequence, but this is all we can say — we can not say that a high income is twice as high as an average income.

Interval data values and ratio data values do allow computation. The first differs from the second in that it knows no arithmetic zero value, and does not support multiplication or division. For instance, a temperature of 20 oC is not twice as warm as 10 oC, and thus centigrade temperatures are interval data values, not ratio data values.
values. Rational data have a natural zero value, and multiplication and division of values are sensible operators: distances measured in metres are an example. Observe that continuous fields can be expected to have ratio data values, simply because we must be able to interpolate them. The attributes as features are stored in a database along with information describing them. The descriptive information stored with a feature (Figure 9.5). For example attributes of a street might include its name, street type, length, street code, number of lanes, and pavement type. The attributes of a park may be its name, area, hours of operation, and maintenance schedule.

9.5. ábra - Figure 9.5: Attributes

3.1. Geographic objects

When the geographic phenomenon is not present everywhere in the study area, but somehow ‘sparsely’ populates it, we look at it in terms of geographic objects. Such objects are usually easily distinguished and named. Their position in space is determined by a combination of one or more of the following parameters:

* location (where is it?),
* shape (what form is it?),
* size (how big is it?), and
* orientation (in which direction is it facing?).

Several attempts have been made to define a taxonomy of geographic object types. Dimension is an important aspect of the shape parameter. It answers the question whether an object is perceived as a point feature, a linear, area or volume feature.

How we want to use the information about a geographic object determines which of the four above parameters is required to represent it. For instance, in a car navigation system, all that matters about geographic objects like petrol stations is where they are, and thus, location suffices. Shape, size and orientation seem to be irrelevant. In the same system, however, roads are important objects, and for these some notion of location (where does it begin and end), shape (how many lanes does it have), size (how far can one travel on it) and orientation (in which direction can one travel on it) seem to be relevant information components.

Shape is usually important because one of its factors is dimension: are the objects inherently considered to be zero-, one-, two- or three-dimensional? The petrol stations mentioned above apparently are zero-dimensional, i.e., they are perceived as points in space; roads are one-dimensional, as they are considered to be lines in space. In another use of road information – for instance, in multi-purpose cadastre systems where precise location of sewers and manhole covers matters – roads might well be considered to be two-dimensional entities, i.e., areas within which a manhole cover may fall.

We usually do not study geographic objects in isolation, but whole collections of objects viewed as a unit. These object collections may also have specific geographic characteristics. Most of the more interesting collections of geographic objects obey certain natural laws. The most common (and obvious) of these is that different objects do not occupy the same location. This, for instance, holds for
• the collection of petrol stations in a car navigation system,
• the collection of roads in that system,
• the collection of parcels in a cadastral system,
• and in many more cases.

3.2. Boundaries

Where shape and/or size of contiguous areas matter, the notion of boundary comes into play. This is true for geographic objects but also for the constituents of a discrete geographic field. Location, shape and size are fully determined if we know an area’s boundary, so the boundary is a good candidate for representing it. This is especially true for areas that have naturally crisp boundaries. A crisp boundary is one that can be determined with almost arbitrary precision, dependent only on the data acquisition technique applied. Fuzzy boundaries contrast with crisp boundaries in that the boundary is not a precise line, but rather itself an area of transition.

As a general rule-of-thumb — again — crisp boundaries are more common in man-made phenomena, whereas fuzzy boundaries are more common with natural phenomena. In recent years, various research efforts have addressed the issue of explicit treatment of fuzzy boundaries, but in day-to-day GIS use these techniques are neither often supported, nor often needed.

3.3. Computer representations of geographic information

Up to this point, we have not discussed at all how geoinformation, like fields and objects, is represented in a computer. One needs to understand at least a little bit about the computer representations to understand better what the system does with the data, and also what it cannot do with it. In the above, we have seen that various geographic phenomena have the characteristics of continuous functions over geometrically bounded, yet infinite domains of space (Figure 9.6).

Elevation, for instance, can be measured at arbitrarily many locations, even within one’s backyard, and each location may give a different value. When we want to represent such a phenomenon faithfully in computer memory, we could either:

• try to store as many (location, elevation) pairs as possible, or
• try to find a symbolic representation of the elevation function, as a formula in x and y — like \((3.0678x^2 + 20.08x - 7.34y)\) or so — which after evaluation will give us the elevation value at a given \((x, y)\).

Both approaches have their drawbacks. The first suffers from the fact that we will never be able to store all elevation values for all locations; after all, there are infinitely many locations. The second approach suffers from the fact that we have no clue what such a function should be, or how to derive it, and it is likely that for larger areas it will be an extremely complicated function.
In GISs, typically a combination of both approaches is taken. We store a finite, but intelligently chosen set of locations with their elevation. This gives us the elevation for those stored locations, but not for others. Therefore, the stored values are paired with an interpolation function that allows to infer a reasonable elevation value for locations that are not stored. The underlying principle is called spatial autocorrelation: locations that are close are more likely to have similar values than locations that are far apart.

Line objects, either by themselves or in their role of region object boundaries, are another common example of continuous phenomena that must be finitely represented. In real life, these objects are usually not straight, and often erratically curved. A famous paradoxical question is whether one can actually measure the length of Great Britain’s coastline... can one measure around rocks, pebbles or even grains of sand? In a computer, such random, curvilinear features can never be fully represented.

One must, thus, observe that phenomena with intrinsic continuous and/or infinite characteristics have to be represented with finite means (computer memory) for computer manipulation, and that any finite representation scheme that forces a discrete look on the continuum that it represents is open to errors of interpretation.

In GIS, fields are usually implemented with a tessellation approach, and objects with a (topological) vector approach. This, however, is not a hard and fast rule, as practice sometimes demands otherwise (Figure 9.7).

9.7. ábra - Figure 9.7: Raster – Vector representation

3.4. Regular tessellations

A tessellation (or tiling) is a partition of space into mutually exclusive cells that together make up the complete study space. With each cell, some (thematic) value is associated to characterize that part of space. In a regular tessellation, the cells are the same shape and size.

All regular tessellations have in common that the cells are of the same shape and size, and that the field attribute value assigned to a cell is associated with the entire area occupied by the cell. The square cell tessellation is by far the most commonly used, mainly because georeferencing a cell is so straightforward. Square, regular tessellations are known under various names in different GIS packages: raster or raster map. The size of the area that a raster cell represents is called the raster ‘s resolution. Sometimes, the word grid is also used, but strictly speaking, a grid is an equally spaced collection of points, which all have some attribute value assigned. They are often used for discrete measurements that occur at regular intervals. Grid points are often considered synonymous with raster cells.

Our finite approximation of the study space leads to some forms of interpolation that must be dealt with. The field value of a cell can be interpreted as one for the complete tessellation cell, in which case the field is discrete, not continuous or even differentiable. Some convention is needed to state which value prevails on cell boundaries; with square cells, this convention often says that lower and left boundaries belong to the cell. To improve on this continuity issue, we can do two things:
• make the cell size smaller, so as to make the ‘continuity gaps’ between the cells smaller, and/or

• assume that a cell value only represents elevation for one specific location in the cell, and to provide a good interpolation function for all other locations that has the continuity characteristic.

Usually, if one wants to use rasters for continuous field representation, one does the first but not the second. The second technique is usually considered too computationally costly for large rasters.

The location associated with a raster cell is fixed by convention, and may be the cell centroid (mid-point) or, for instance, its left lower corner. Values for other positions than these must be computed through some form of interpolation function, which will use one or more nearby field values to compute the value at the requested position. This allows to represent continuous, even differentiable, functions.

An important advantage of regular tessellations is that we a priori know how they partition space, and we can make our computations specific to this partitioning. This leads to fast algorithms. An obvious disadvantage is that they are not adaptive to the spatial phenomenon we want to represent. The cell boundaries are both artificial and fixed: they may or may not coincide with the boundaries of the phenomenon of interest.

Adaptivity to the phenomenon to represent can pay off. Suppose we use any of the above regular tessellations to represent elevation in a perfectly flat area. Then, clearly we need as many cells as in a strongly undulating terrain: the data structure does not adapt to the lack of relief. We would, for instance, still use the $m \times n$ cells for the raster, although the elevation might be 1500 m above sea level everywhere.

### 3.5. Irregular tessellations

Above, we discussed that regular tessellations provide simple structures with straightforward algorithms, which are, however, not adaptive to the phenomena they represent. This is why substantial effort has also been put into irregular tessellations. Again, these are partitions of space into mutually disjoint cells, but now the cells may vary in size and shape, allowing them to adapt to the spatial phenomena that they represent. We discuss here only one type, namely the region quadtree, but we point out that many more structures have been proposed in the literature and have been implemented as well.

Irregular tessellations are more complex than the regular ones, but they are also more adaptive, which typically leads to a reduction in the amount of memory used to store the data.

A well-known data structure in this family – upon which many more variations have been based – is the region quadtree. It is based on a regular tessellation of square cells, but takes advantage of cases where neighbouring cells have the same field value, so that they can together be represented as one bigger cell. The quadtree that represents this raster is constructed by repeatedly splitting up the area into four quadrants, which are called NW, NE, SE, SW for obvious reasons. This procedure stops when all the cells in a quadrant have the same field value. The procedure produces an upside-down, tree-like structure, known as a quadtree. In main memory, the nodes of a quadtree are represented as records. The ulinks between them are pointers, a programming technique to address (i.e., to point to) other records.

### 3.6. Vector representations

In summary of the above, we can say that tessellations cut up the study space into cells, and assign a value to each cell. A raster is a regular tessellation with square cells, and this is by far the most commonly used. How the study space is cut up is (to some degree) arbitrary, and this means that cell boundaries usually have no bearing to the real world phenomena that are represented.

In vector representations, an attempt is made to associate georeferences with the geographic phenomena explicitly. A georeference is a coordinate pair from some geographic space, and is also known as a vector. This explains the name. We will see a number of examples below.

Observe that tessellations do not explicitly store georeferences of the phenomena they represent. Instead, they might provide a georeference of the lower left corner of the raster, for instance, plus an indicator of the raster’s resolution, thereby implicitly providing georeferences for all cells in the raster.

Below, we discuss various vector representations. We start with our discussion with the TIN, a representation for geographic fields that can be considered a hybrid between tessellations and vector representations.
Points are defined as single coordinate pairs \((x, y)\) when we work in 2D or coordinate triplets \((x, y, z)\) when we work in 3D. Points are used to represent objects that are best described as shape- and sizeless, single-locality features. Whether this is the case really depends on the purposes of the spatial application and also on the spatial extent of the objects compared to the scale applied in the application. For a tourist city map, parks will not usually be considered as point features, but perhaps museums will be, and certainly public phone booths could be represented as point features.

Besides the georeference, usually extra data is stored for each point object. This so-called administrative or thematic data, can capture anything that is considered relevant about the object. For phone booth objects, this may include the owning telephone company, the phone number, the data last serviced et cetera.

Line data are used to represent one-dimensional objects such as roads, railroads, canals, rivers and power lines. Again, there is an issue of relevance for the application and the scale that the application requires. For the example application of mapping tourist information, bus, subway and streetcar routes are likely to be relevant line features. Some cadastral systems, on the other hand, may consider roads to be two-dimensional features, i.e., having a width as well. Arbitrary, continuous curvilinear features are equally difficult to represent as continuous fields. GISs therefore approximate such features (finitely!) as lists of nodes. The two end nodes and zero or more internal nodes define a line. Another word for internal node is vertex (plural: vertices); another phrase for line that is used in some GISs is polyline, arc or edge. A node or vertex is like a point (as discussed above) but it only serves to define the line; it has no special meaning to the application other than that.

The vertices of a line help to shape it, and to obtain a better approximation of the actual feature. The straight parts of a line between two consecutive vertices or end nodes are called line segments. Many GISs store a line as a simple sequence of coordinates of its end nodes and vertices, assuming that all its segments are straight. This is usually good enough, as cases in which a single straight line segment is considered an unsatisfactory representation can be dealt with by using multiple (smaller) line segments instead of only one.

Still, there are cases in which we would like to have the opportunity to use arbitrary curvilinear features as representation of real-world phenomena. Think of garden design with perfect circular or elliptical lawns, or of detailed topographic maps representing roundabouts and the annex sidewalks.

When area objects are stored using a vector approach, the usual technique is to apply a boundary model. This means that each area feature is represented by some arc/node structure that determines a polygon as the area’s boundary. Common sense dictates that area features of the same kind are best stored in a single data layer, represented by mutually non-overlapping polygons. In essence, what we then get is an application-determined (i.e., adaptive) partition of space, similar to, but not quite like an irregular tessellation of the raster approach.

Observe that a polygon representation for an area object is yet another example of a finite approximation of a phenomenon that inherently may have a curvilinear boundary. In the case that the object can be perceived as having a fuzzy boundary, a polygon is an even worse approximation, though potentially the only one possible.

A simple but native representation of area features would be to list for each polygon simply the list of lines that describes its boundary. Each line in the list would, as before, be a sequence that starts with a node and ends with one, possibly with vertices in between. But this is far from optimal.

### 4. General spatial topology

Topology deals with spatial properties that do not change under certain transformations. A simple example will illustrate what we mean. Assume you have some features that are drawn on a sheet of rubber. Now, take the sheet and pull on its edges, but do not tear or break it. The features will change in shape and size. There are a number of advantages when our computer representations of geographic phenomena have built-in sensitivity of topological issues. Questions related to the ‘neighbourhood’ of an area are a point in case. To obtain some ‘topological sensitivity’ simple building blocks have been proposed with which more complicated representations can be constructed.

#### 4.1. The topology of two dimensions

We can use the topological properties of interior and boundary to define relationships between spatial features. Since the properties of interior and boundary do not change under topological mappings, we can investigate their possible relations between spatial features. We can define the interior of a region \(R\) as the maximal set of
points in R for which we can construct a disk-like environment around it (no matter how small) that also falls completely inside R. The boundary of R is the set of those points belonging to R but that do not belong to the interior of R, i.e., one cannot construct a disk-like environment around such points that still belongs to R completely.

It turns out that the rules of how simplices and simplicial complexes can be embedded in space are quite different for two-dimensional space than they are for three-dimensional space. Such a set of rules defines the topological consistency of that space. It can be proven that if the rules below are satisfied for all features in a two-dimensional space, the features define a topologically consistent configuration in 2D space.

### 4.2. Scale and resolution

In the practice of spatial data handling, one often comes across questions like “what is the resolution of the data?” or “at what scale is your data set?” Now that we have moved firmly into the digital age, these questions defy an easy answer sometimes.

Map scale can be defined as the ratio between distance on a paper map and distance of the same stretch in the terrain. A 1:50,000 scale map means that 1 cm on the map represents 50,000 cm, i.e., 500 m, in the terrain. ‘Large-scale’ means that the ratio is large, so typically it means there is much detail; ‘small-scale’ in contrast means a small ratio, hence fewer detail. When applied to spatial data, the term resolution is commonly associated with the cell width of the tessellation applied.

Digital spatial data, as stored in a GIS, is essentially without scale: scale is a ratio notion associated with visual output, like a map, not with the data that was used to produce the map.

### 4.3. Representations of geographic fields

In the above we have looked at various representation techniques. Now we can study which of them can be used to represent a geographic field. A geographic field can be represented through a tessellation, through a TIN or through a vector representation. The choice between them is determined by the requirements of the application at hand. It is more common to use tessellations, notably rasters, for field representation, but vector representations are in use too. We have already looked at TINs. We provide an example how a raster represents a continuous field (Figure 9.8). Different shades of blue indicate different elevation values, with darker blues indicating higher elevations. The choice of a blue colour spectrum is only to make the illustration aesthetically pleasing; real elevation values are stored in the raster, so instead we could have printed a real number value in each cell. This would not have made the figure very legibile, however.

9.8. ábra - Figure 9.8: Raster Representation

![Raster Representation](image)

### 4.4. Vector representation of a field

We shortly mention a final representation for fields like elevation, but with a vector flavour. This technique uses isolines of the field. An isoline is a linear feature that connects the points with equal field value. When the field is elevation, we also speak of contour lines. Both TINs and isoline representations use vectors.

### 5. Representation of geographic objects

The representation of geographic objects is most naturally supported with vectors. After all, objects are identified by the parameters of location, shape, size and many of these parameters can be expressed in terms of vectors.
Tessellations are not entirely out of the picture, though, and are commonly used for representing geographic objects as well.

### 5.1. Tessellations to represent geographic objects

Remotely sensed images are an important data source for GIS applications. Un-processed digital images contain pixels, with each pixel carrying a reflectance value. Various techniques exist to process digital images into classified images that can be stored in a GIS as a raster. Image classification attempts to characterise each pixel into one of a finite list of classes, thereby obtaining an interpretation of the contents of the image.

Nonetheless, we must make a few observations regarding the representation of geographic objects in rasters. Area objects are conveniently represented in raster, albeit that area boundaries may appear ragged. This is a typical byproduct of raster resolution versus area size, and artificial cell boundaries. One must be aware, for instance, of the consequences for area size computations: what is the precision with which the raster defines the object's size?

Line and point objects are more awkward to represent using rasters. After all, we could say that rasters are area-based, and geographic objects that are perceived as lines or points are perceived to have zero area size. Standard classification techniques, moreover, may fail to recognise these objects as points or lines.

Many GIS do offer support for line representations in raster, and operations on them. Lines can be represented as strings of neighbouring raster cells with equal value. Supported operations are connectivity operations and distance computations. There is again an issue of precision of such computations.

### 5.2. Vector representations for geographic objects

The somehow more natural way to represent geographic objects is by vector representations (Figure 9.10).

![Figure 9.10: Vector representation](image)

### 6. Organizing one’s spatial data

In the previous sections, we have discussed various types of geographic information and ways of representing them. We have looked at case-by-case examples, however, without looking much at how various sorts of spatial data are combined in a single system.

The main principle of data organization applied in GIS systems is that of a spatial data layer. A spatial data layer is either a representation of a continuous or discrete field, or a collection of objects of the same kind. The intuition is that the data is organized by kind: all telephone booth point objects would be in a single data layer, all road line objects in another one. A data layer contains spatial data — of any of the types discussed above — as well as attribute (or thematic) data, which further describes the field or objects in the layer. Attribute data is quite often arranged in tabular form.
Data layers can be overlaid with each other, inside the GIS package, so as to study combinations of geographic phenomena. We shall see later that a GIS can be used to study the spatial correlation between different phenomena: in what way are occurrences/events occurring in the same location? To that end, a computation is performed that overlays one data layer with another. But GIS software also allows to overlay field layers, or even a field with an object layer.

6.1. Spatiotemporal data

Beside having geometric, thematic and topological properties, geographic phenomena change over time; we say that they have temporal characteristics. And for many applications, it is change over time that is quite often the most interesting aspect of the phenomenon to study. This area of work is commonly known as change detection. It is, for instance, interesting to know who were the owners of a land parcel in 1980, or how land cover changed from the original primary forest to pastures over time. Change detection addresses such questions as:

- Where and when did change take place?
- What kind of change occurred?
- With what speed did change occur?
- What else can be understood about the pattern of change?

The support that GISs offer for change detection is at present not very impressive. Most studies require substantial efforts from the GIS user in data preparation and data manipulation. Spatiotemporal data structures are representations of geographic phenomena changing over time. Several representation techniques have been proposed in the literature. The most important ones will be discussed briefly below. Observe that besides 2D or 3D space, the extra dimension of time is again inherently of a continuous nature, and that again, if we want to represent this in a computer, we will have to ‘discretize’ this dimension.

Before we describe the major characteristics of various techniques, we need a framework to describe the nature of time itself. The time dimension can be characterized with the following properties:

Time density Time can be measured along a discrete or continuous scale. Discrete time is composed of discrete elements (seconds, minutes, hours, days, months, or years). In continuous time, no such discrete elements exist, and for any two different points in time, there is always another point in between. We can also structure time by events (points in time) or periods (time intervals). When we represent time periods by a start and end event, we can derive temporal relationships between events and periods such as ‘before’, ‘overlap’, ‘after ’, et cetera.

Dimensions of time Valid time (or world time) is the time when an event really happened, or a string of events took place. Transaction time (or database time) is the time when the event was stored in the database or GIS. Observe that the time at which we store something in the database/GIS typically is (much) later than when the related event took place.

Often, what we record in a computer system is a ‘snapshot state’ that represents a single point in time of an ongoing natural or man-made process. We may store a string of ‘snapshot states’ but must be aware that this is still only a feeble representation of that process.

Time order Time can be considered to be linear, extending from the past to the present (‘now’), and into the future. For some types of temporal analysis, branching time — in which different time lines from a certain point in time onwards are possible — and cyclic time — in which repeating cycles such as seasons or days of a week are recognized, make more sense and can be useful.

Measures of time When measuring time, we speak of a chronon as the shortest non-decomposable unit of time that is supported by a GIS or database (e.g., this could be a millisecond). The life span of an object is measured by a (finite) number of chronons. Granularity is the precision of a time value in a GIS or database (e.g., year, month, day, second, etc.). Different applications require different granularity. In cadastral applications, time granularity could well be a day, as the law requires deeds to be date-marked; in geological mapping applications, time granularity is more likely in the order of thousands or millions of years.

Time reference Time can be represented as absolute (fixed time) or relative (implied time). Absolute time marks a point on the time line where events happen (e.g., ‘6 July 1999 at 11:15 p.m.’). Relative time is indicated
relative to other points in time (e.g., ‘yesterday’, ‘last year’, ‘tomorrow’, which are all relative to ‘now’, or ‘two weeks later’, which may be relative to an arbitrary point in time.).

7. Geographic information systems

The handling of spatial data usually involves processes of data acquisition, storage and maintenance, analysis and output. For many years, this has been done using analogue data sources, manual processing and the production of paper maps. The introduction of modern technologies has led to an increased use of computers and digital information in all aspects of spatial data handling. The software technology used in this domain is geographic information systems.

9.10. ábra - Figure 9.11: Information system

Typical planning projects require data sources, both spatial and non-spatial, from different institutes, like mapping agency, geological survey, soil survey, forest survey, or the census bureau. These data sources may have different time stamps, and the spatial data may be in different scales and projections. With the help of a GIS, the maps can be stored in digital form in a database in world coordinates (metres or feet). This makes scale transformations unnecessary, and the conversion between map projections can be done easily with the software. The spatial analysis functions of the GIS are then applied to perform the planning tasks. This can speed up the process and allows for easy modifications to the analysis approach.

7.1. The context of GIS usage

Spatial data handling involves many disciplines. We can distinguish disciplines that develop spatial concepts, provide means for capturing and processing of spatial data, provide a formal and theoretical foundation, are application-oriented, and support spatial data handling in legal and management aspects.

The discipline that deals with all aspects of spatial data handling is called geoinformatics. It is defined as:

Geoinformatics is the integration of different disciplines dealing with spatial information.

Geoinformatics has also been described as the science and technology dealing with the structure and character of spatial information, its capture, its classification and qualification, its storage, processing, portrayal and dissemination, including the infrastructure necessary to secure optimal use of this information.

7.2. GIS software

The main characteristics of a GIS software package are its analytical functions that provide means for deriving new geoinformation from existing spatial and attribute data. A GIS can be defined as follows:
A GIS is a computer-based system that provides the following four sets of capabilities to handle georeferenced data:

- input,
- data management (data storage and retrieval),
- manipulation and analysis, and
- output.

Depending on the interest of a particular application, a GIS can be considered to be a data store (i.e., a database that stores spatial data), a toolbox, a technology, an information source or a field of science (as part of spatial information science).

Like in any other discipline, the use of tools for problem solving is one thing, to produce these tools is something different. Not all tools are equally well-suited for a particular application. Tools can be improved and perfected to better serve a particular need or application. The discipline that provides the background for the production of the tools in spatial data handling is spatial information theory.

All GIS packages available on the market have their strengths and weaknesses, resulting typically from the package’s development history and/or intended application domain(s). Some GIS have traditionally focused more on support for raster manipulation, others more on (vector-based) spatial objects. We can safely state that any package that provides support for only raster or only objects, is not a full-fledged, generic GIS.

One cannot say that one GIS package is ‘better’ than another one: it all depends what one wants to use the package for. ILWIS’s traditional strengths have been in raster processing and scientific spatial data analysis, especially suitable in what we called project-based GIS applications. ArcInfo has been renowned more for its support of vector-based spatial data and their operations, user interface and map production, a bit more typical of institutional GIS applications. Any such brief characterization, however, does not do justice to these packages, and it is only after extended use that preferences become clear.

### 7.3. Software architecture and functionality of a GIS

A geographic information system in the wider sense consists of software, data, people, and an organization in which it functions. In the narrow sense, we consider a GIS as a software system for which we discuss its architecture and functional components.

According to the definition, a GIS always consists of modules for input, storage, analysis, display and output of spatial data. For a particular GIS, each of these modules may provide many or only few functions. However, if one of these functions would be completely missing, the system should not be called a geographic information system.

An explanation of the various functions of the four components for data input, storage, analysis, and output can provide a functional description of a GIS. Here, we only briefly describe them. A more detailed treatment can be found in follow-up chapters.

Beside data input (data capture), storage and maintenance, analysis and output, geoinformation processes involve also dissemination, transfer and exchange as well as organizational issues. The latter define the context and rules according to which geoinformation is acquired and processed.

### 7.4. Querying, maintenance and spatial analysis

The most distinguishing part of a GIS are its functions for spatial analysis, i.e., operators that use spatial data to derive new geoinformation. Spatial queries and process models play an important role in satisfying user needs. The combination of a database, GIS software, rules, and a reasoning mechanism (in- plemented as a so-called inference engine) leads to what is sometimes called a spatial decision support system (SDSS).
In a GIS, data are stored in layers (or themes). Usually, several themes are part of a project. The analysis functions of a GIS use the spatial and non-spatial attributes of the data in a spatial database to answer questions about the real world.

In spatial analysis, various kinds of question may arise. The following three classes are the most important query and analysis functions of a GIS:

- Maintenance and analysis of spatial data,
- Maintenance and analysis of attribute data, and
- Integrated analysis of spatial and attribute data.

7.5. Maintenance and analysis of spatial data

Maintenance of (spatial) data can best be defined as the combined activities to keep the data set up-to-date and as supportive as possible to the user community. It deals with obtaining new data, and entering them into the system, possibly replacing outdated data. The purpose is have available an up-to-date, stored data set. After a major earthquake, for instance, we may have to update our digital elevation model to reflect the current elevations better so as to improve our hazard analysis.

Operators of this kind operate on the spatial properties of GIS data, and provide a user with functions as described below.

Format transformation functions convert between data formats of different systems or representations, e.g., reading a DXF file into a GIS.

Geometric transformations help to obtain data from an original hard copy source through digitizing the correct world geometry. These operators transform device coordinates (coordinates from digitizing tablets or screen coordinates) into world coordinates (geographic coordinates, metres, etc.).

Map projections provide means to map geographic coordinates onto a flat surface (for map production), and vice versa.

Edge matching is the process of joining two or more map sheets. At the map sheet edges, feature representations have to be matched so as to be combined.

Graphic element editing allows to change digitized features so as to correct errors, and to prepare a clean data set for topology building.

7.6. Integrated analysis of spatial and attribute data

Analysis of (spatial) data can be defined as computing from the existing, stored data set new information that provides insights we possibly did not have before. It really depends on the application requirements, and the examples are manifold. Road construction in mountainous areas is a complex engineering task with many cost factors such as the amount of tunnels and bridges to be constructed, the total length of the tarmac, and the volume of rock and soil to be moved. GIS can help to compute such costs on the basis of an up-to-date digital elevation model and soil map.

Functions of this kind operate on both spatial and non-spatial attributes of data, and can be grouped into the following types.

Retrieval, classification, and measurement functions:

- Retrieval functions allow the selective search and manipulation of data without the need to create new entities.
- Classification allows assigning features to a class on the basis of attribute values or attribute ranges (definition of data patterns).
• Generalization is a function that joins different classes of objects with common characteristics to a higher level (generalized) class.

The term generalization has different meanings in different contexts. In geography the term ‘aggregation’ is often used to indicate the process that we call generalization. In cartography, generalization means either the process of producing a graphic representation of smaller scale from a larger scale original (cartographic generalization), or the process of deriving a coarser resolution representation from a more detailed representation within a database (model generalization). Finally, in computer science generalization is one of the abstraction mechanisms in object-orientation.

Overlay functions belong to the most frequently used functions in a GIS application. They allow combining two spatial data layers by applying the set-theoretic operations of intersection, union, difference, and complement using sets of positions (geometric attribute values) as their arguments. Thus we can find:

• the potato fields on clay soils (intersection),
• the fields where potato or maize is the crop (union),
• the potato fields not on clay soils (difference),
• the fields that do not have potato as crop (complement).

Neighbourhood functions operate on the neighbouring features of a given feature or set of features:

• Search functions allow the retrieval of features that fall within a given search window (which may be a rectangle, circle, or polygon).
• Line-in-polygon and point-in-polygon functions determine whether a given linear or point feature is located within a given polygon, or
• They report the polygons that a given point or line are contained in.
• The best known example of proximity functions is the buffer zone generation (or buffering). This function determines a fixed-width (or variable-width) environment surrounding a given feature.
• Topographic functions compute the slope or aspect from a given digital representation of the terrain (digital terrain model or DTM).
• Interpolation functions predict unknown values using the known values at nearby locations.
• Contour generation functions calculate contours as a set of lines that connect points with the same attribute value. Examples are points with the same elevation (contours), same depth (bathymetric contours), same barometric pressure (isobars), or same temperature (isothermal lines).

Connectivity functions accumulate values as they traverse over a feature or over a set of features:

• Contiguity measures evaluate characteristics of spatial units that are contiguous are connected with unbroken adjacency. Think of the search for a contiguous area of forest of certain size and shape.
• Network analysis is used to compute the shortest path (in terms of distance or travel time) between two points in a network (routing).
• Alternatively, it finds all points that can be reached within a given distance or duration from a centre (allocation).
• Visibility functions are used to compute the points that are visible from a given location (viewshed modelling or viewshed mapping) using a digital terrain model.

8. entry and preparation

The first step of using a GIS is to provide it with data. The acquisition and pre-processing of spatial data is an expensive and time-consuming process. Much of the success of a GIS project, however, depends on the quality
of the data that is entered into the system, and thus this phase of a GIS project is critical and must be taken seriously.

8.1. Spatial data input

Spatial data can be obtained from scratch, using direct spatial data acquisition techniques, or indirectly, by making use of spatial data collected earlier, possibly by others. Under the first heading fall field survey data and remotely sensed images. Under the second fall paper maps and available digital data sets.

8.2. Direct spatial data acquisition

The primary, and sometimes ideal, way to obtain spatial data is by direct observation of the relevant geographic phenomena. This can be done through ground-based field surveys in situ, or by using remote sensors in satellites or airplanes. An important aspect of ground-based surveying is that some of the data can be interpreted immediately by the surveyor. Many Earth sciences have developed their own survey techniques, and where these are relevant for the student, they will be taught in subsequent modules, as ground-based techniques remain the most important source for reliable data in many cases.

For remotely sensed imagery, obtained from satellites or aerial reconnaissance, this is not the case. These data are usually not fit for immediate use, as various sources of error and distortion may have been present at the time of sensing, and the imagery must first be freed from these as much as possible.

An important distinction that we must make is that between ‘image’ and ‘raster’. By the first term, we mean a picture with pixels that represent measured local reflectance values in some designated part of the electromagnetic spectrum. No value has yet been added in terms of interpreting such values as thematic or geographic characteristics. When we use the term ‘raster’, we assume this value-adding interpretation has been carried out. With an image, we talk of its constituent pixels; with a raster we talk of its cells.

In practice, it is not always feasible to obtain spatial data using these techniques. Factors of cost and available time may be a hindrance, and moreover, previous projects sometimes have acquired data that may fit the current project’s purpose. We look at some of the ‘indirect’ techniques of using existing sources below.

8.3. Digitizing paper maps

A cost-effective, though indirect, method of obtaining spatial data is by digitizing existing maps. This can be done through a number of techniques, all of which obtain a digital version of the original (analog) map. Before adopting this approach, one must be aware that, due to the indirect process, positional errors already in the paper map will further accumulate, and that one is willing to accept these errors.

In manual digitizing, a human operator follows the map’s features (mostly lines) with a mouse device, and thereby traces the lines, storing location coordinates relative to a number of previously defined control points. Control points are sometimes also called ‘tie points’. Their function is to ‘lock’ a coordinate system onto the digitized data: the control points on the map have known coordinates, and by digitizing them we tell the system implicitly where all other digitized locations are. At least three control points are needed, but preferably more should be digitized to allow a check on the positional errors made. There are two forms of digitizing: on-tablet and on-screen manual digitizing.

In on-tablet digitizing, the original map is fitted on a special tablet and the operator moves a special tablet mouse over the map, selecting important points. In on-screen digitizing, a scanned image of the map — or in fact, some other image — is shown on the computer screen, and the operator moves an ordinary mouse cursor over the screen, again selecting important points. In both cases, the GIS works as a point recorder, and from this recorded data, line features are later constructed. There are usually two modes in which the GIS can record: in point mode, the system only records a mouse location when the operator says so; in stream mode, the system almost continuously records locations. The first is the more useful technique because it can be better controlled, as it is less prone to shaky hand movements.

Another set of techniques also works from a scanned image of the original map, but uses the GIS to find features in the image. These techniques are known as semi-automatic or automatic digitizing, depending on how much operator interaction is required. If vector data is to be distilled from this procedure, a process known as
vectorization follows the scanning process. This procedure is less labour-intensive, but can only be applied on relatively simple sources.

### 8.4. The scanning process

A digital scanner illuminates a to-be-scanned document and measures with a sensor the intensity of the reflected light. The result of the scanning process is an image as a matrix of pixels, each of which holds a reflectance value. Before scanning, one has to decide whether to scan the document in line art, grey-scale or colour mode. The first results in either ‘white’ or ‘black’ pixel values; the second in one of 256 ‘grey’ values per pixel, with white and black as extremes.

### 8.5. The vectorization process

Vectorization is the process that attempts to distill points, lines and polygons from a scanned image. As scanned lines may be several pixels wide, they are often first ‘thinned’, to retain only the centreline. This thinning process is also known as skeletonizing, as it removes all pixels that make the line wider than just one pixel. The remaining centreline pixels are converted to series of (x, y) coordinate pairs, which define the found polyline. Afterwards, features are formed and attributes are attached to them. This process may be entirely automated or performed semi-automatically, with the assistance of an operator.

Semi-automatic vectorization proceeds by placing the mouse pointer at the start of a line to be vectorized. The system automatically performs line-following with the image as input. At junctions, a default direction is followed, or the operator may indicate the preferred direction.

Pattern recognition methods — like Optical Character Recognition (OCR) for text — can be used for the automatic detection of graphic symbols and text. Once symbols are recognized as image patterns, they can be replaced by symbols in vector format, or better, by attribute data. For example, the numeric values placed on contour lines can be detected automatically to attach elevation values to these vectorized contour lines.

### 8.6. Obtaining spatial data elsewhere

Various spatial data sources are available from elsewhere, though sometimes at a price. It all depends on the nature, scale, and date of production that one requires. Topographic base data is easier to obtain than elevation data, which is in turn easier to get than natural resource or census data. Obtaining large-scale data is more problematic than small-scale, of course, while recent data is more difficult to obtain than older data. Some of this data is only available commercially, as usually is satellite imagery.

National mapping organizations (NMOs) historically are the most important spatial data providers, though their role in many parts of the world is changing. Many governments seem to be less willing to maintain large institutes like NMOs, and are looking for alternatives to the nation’s spatial data production. Private companies are probably going to enter this market, and for the GIS application people this will mean they no longer have a single provider.

Statistical, thematic data always was the domain of national census or statistics bureaus, but they too are affected by changing policies. Various commercial research institutes also are starting to function as provider for this type of information.

Clearinghouses as digital data provision is an expertise by itself, many of the above-mentioned organizations dispatch their data via centralized places, essentially creating a marketplace where potential data users can ‘shop’. It will be no surprise that such markets for digital data have an entrance through the world-wide web. They are sometimes called spatial data clearinghouses. The added value that they provide is to-the-point metadata: searchable descriptions of the data sets that are available.

Data formats are an important problem in any environment involved in digital data exchange is that of data formats and data standards. Different formats were implemented by different GIS vendors; different standards come about with different standardization committees.

The good news about both formats and standards is that there are so many to choose from; the bad news is that this causes all sorts of conversion problems. We will skip the technicalities — as they are not interesting, and little can be learnt from them — but warn the reader that conversions from one format to another may mean trouble. The reason is that not all formats can capture the same information, and therefore conversions often
mean loss of information. If one obtains a spatial data set in format F, but wants it in format G, for instance because the locally preferred GIS package requires it, then usually a conversion function can be found, likely in that same GIS. The proof of the pudding is to also find an inverse conversion, back from G to F, and to ascertain whether the double conversion back to F results in the same data set as the original. If this is the case, both conversions are not causing information loss, and can safely be applied.

9. Spatial referencing

In the early days of GIS, users were handling spatially referenced data from a single country. The data was derived from paper maps published by the country’s mapping organization. Nowadays, GIS users are combining spatial data from a certain country with global spatial data sets, reconciling spatial data from a published map with coordinates established with satellite positioning techniques and integrating spatial data from neighbouring countries. To perform these tasks successfully, GIS users need a certain level of appreciation for a few basic spatial referencing concepts pertinent to published maps and spatial data.

Spatial referencing encompasses the definitions, the physical/geometric constructs and the tools required to describe the geometry and motion of objects near and on the Earth’s surface. Some of these constructs and tools are usually itemized in the legend of a published map. For instance, a GIS user may en-counter the following items in the map legend of a conventional published large-scale topographic map: the name of the local vertical datum (e.g., Tide-gauge Amsterdam), the name of the local horizontal datum (e.g., Potsdam Datum), the name of the reference ellipsoid and the fundamental point (e.g., Bessel Ellipsoid and Rauenberg), the type of coordinates associated with the map grid lines (e.g., geographic coordinates, plane coordinates), the map projection (e.g., Universal Transverse Mercator projection), the map scale (e.g., 1 : 25, 000), and the transformation parameters from a global datum to the local horizontal datum.

In the following subsections we shall explain the meaning of these items. An appreciation of basic spatial referencing concepts will help the reader identify potential problems associated with incompatible spatially referenced data.

10. Questions

1. What is a GIS?
2. History of GIS
3. Basic Structure of a GIS
4. Applications
5. Remote Sensing
6. General Purpose of GIS
7. Examples of Applied GIS
10. fejezet - ICT in Quality Management

1. Introduction

With the rising liberalisation of agro-industrial markets and thus the world-wide integration of food supply chains, the assurance of food quality and safety has become a major concern. Global trading needs standardised products. Following serious and repeated incidents such as mad cow disease (Bovine Spongiform Encephalitis – BSE), Dioxin, Aflatoxin, and most recently, Sudan Red, consumer protection has become a priority in policy making in the large consumer markets of the United States and the European Union. Hence, legal requirements for quality assurance systems and food control along the entire food chain, from seed and agricultural production, through food processing and the distribution system, up to the consumers’ table, are increasing considerably. In parallel, many retailers introduced private labels thus becoming more vulnerable and responsive to consumer concerns. As a reaction to widespread protest following food scandals, retailers and their respective business associations took the initiative to develop common good practices for food quality and safety from farm to fork by integrating the whole supply chain into their quality concepts.

The consequent pressure on fresh and raw material producers, processors, forwarding companies as well as control institutions goes beyond boundaries and implies high requirements on quality assurance systems abroad. As a consequence, farmers and companies, legislative and control bodies, accreditation, certification and advice giving organisations in export countries need to develop and implement respective institutional capacities, guidelines and knowledge transfer systems aimed at assuring food quality and safety.

1.1. Food quality and safety

Major prerequisite for ensuring food quality and safety is that all stakeholders in the food supply chain recognise that primary responsibility lies with those who produce, process and trade food and that public control should be based on (scientific) risk assessment. Operators’ responsibilities cover the whole food supply and marketing chain from primary production to final consumption and encompass all actors in exporting and importing countries, such as.

1.2. Actors in the food supply chain

Food Supply Chain Operators:

- farmers (variety and seed selection, soil preparation, crop and pest management, harvesting methods, sorting, grading, packing, etc. including documentation of all work routines);

- processors (raw material handling and control, product and process management and control, etc. including documentation of all work routines, handling and control of raw material of local and imported origin, etc. including documentation of all work routines);

- consumers (product selection, transport, storage, preparation, consumption and disposal of household waste, fresh and processed produce handling, storage and preparation in appropriate conditions);

- wholesalers and retailers (fresh and processed produce storage and distribution in appropriate conditions, etc. including documentation).

Food Supply Chain Supporters:

- operational service providers such as forwarding agents and the like subcontracted by operators (produce handling, transport and storage, laboratory services, etc. including documentation);

- support service providers such as publicly or collectively financed institutions incl. branch associations (research and development, education, training, trade promotion, advocacy, consumer associations and environmental organisations etc.).
Food Supply Chain Enablers:

- public entities such as policy makers and regulatory bodies (political, economic and legislative framework conditions);
- food control agents at boundaries (risk-based [phyto]-sanitary control, certificate of origin, monitoring and reporting, etc.).

The list illustrates that in the case of exports of agro-industrial products, an important part of the responsibility for food quality and safety assurance stays with public and private stakeholders in the exporting countries. A safe and good-quality product should be the result of adequate control at all stages of the supply chain rather than corrective action taken late in the process. Stakeholders report that investments into Good Practice and Quality Assurance Systems (compliance costs) are justifiable and in many cases result in a more than reasonable return on investment, namely:

- reduced input costs through implementation of integrated crop/pest management;
- higher labour productivity through improved work-flow;
- improved market access through communication of the Good Practices applied;
- improved long-term supplier-customer relationships through reliable and continuous food quality.

### 1.3. World Trade Organization (WTO)

As the second large negotiation round of the WTO, the ‘Doha Development Round’ was explicitly devoted to assist less and least developing countries to achieve their development goals by focusing on the triangle of market access and domestic support, both in agriculture, and non-agricultural market access. With the negotiation groups failing to submit the modalities even after several extensions, the Secretary General of the WTO recommended to the General Council to suspend the Doha Development Agenda negotiations on 24 July 2006 by leaving the opportunity to members to resume negotiations any time.

The main blockage is in the two agriculture related fields of market access and domestic support (the third angle, non-agricultural market access, has not even been discussed yet). The EU and the United States (US) hold one each other responsible for the current impasse in the Doha negotiations. The US is considered to be inflexible with regard to domestic support whereas the EU is regarded as focusing mainly on market access. At the same time, developing countries consider their agriculture to be unprotected against subsidised exports from industrialised countries.

### 1.4. International Organization for Standardization (ISO)

Published on 1 September 2005, the “Food Safety Management System” ISO 22000:2005 is a new certification standard, which combines the generic management system of the ISO 9001:2000 family with the hygiene requirements for the food industry (HACCP – Hazard Analysis Critical Control Points). ISO 22000:2005 specifies requirements for a food safety management system for all operators along the food supply chain.

### 1.5. European Union (EU)

In 2002, the Directorate General Health and Consumer Protection (DG SANCO) of the European Commission adopted much tougher measures to strengthen official food and feed controls. The aim was to streamline the previously weak and scattered controls and to strengthen consumer protection by giving both Member States and the Commission stricter enforcement tools. While the controls remain the responsibility of the Member States, performance criteria for competent authorities and the design and development of control systems have been harmonised EU-wide.

The EU looks back at many years of tracking and tracing research while precision agriculture is dealt with in global conference series since many years as well. It is not the least due to deficiencies in the capability of ICT that these initiatives have not reached widespread adoption irrespective of needs. Especially tracking and tracing as a baseline information and communication activity for transparency in the food sector has not reached a level of operation that matches the global network activity of the food sector in sourcing and sales.
2. RFID (Radio Frequency Identification): Principles and Applications

Radio frequency identification (RFID) is a rapidly growing technology that has the potential to make great economic impacts on many industries. While RFID is a relatively old technology, more recent advancements in chip manufacturing technology are making RFID practical for new applications and settings, particularly consumer item level tagging. These advancements have the potential to revolutionize supply-chain management, inventory control, and logistics.

At its most basic, RFID systems consist of small transponders, or tags, attached to physical objects. RFID tags may soon become the most pervasive microchip in history. When wirelessly interrogated by RFID transceivers, or readers, tags respond with some identifying information that may be associated with arbitrary data records. Thus, RFID systems are one type of automatic identification system, similar to optical bar codes (Figure 10.1, 10.2).

10.1. ábra - Figure 10.1: BarCode

10.2. ábra - Figure 10.2: BarCodes Family
There are many kinds of RFID systems used in different applications and settings. These systems have different power sources, operating frequencies, and functionalities. The properties and regulatory restrictions of a particular RFID system will determine its manufacturing costs, physical specifications, and performance. Some of the most familiar RFID applications are item-level tagging with electronic product codes, proximity cards for physical access control, and contact-less payment systems. Many more applications will become economical in the coming years.

While RFID adoption yields many efficiency benefits, it still faces several hurdles. Besides the typical implementation challenges faced in any information technology system and economic barriers, there are major concerns over security and privacy in RFID systems. Without proper protection, RFID systems could create new threats to both corporate security and personal privacy.

2.1. Auto-Identification and RFID

In terms of commercial applications, RFID systems may be considered an instance of a broader class of automatic identification (auto-ID) systems. Auto-ID systems essentially attach a name or identifier to a physical object by some means that may be automatically read. This identifier may be represented optically, electromagnetically, or even chemically.

Perhaps the most successful and well-known auto-ID system is the Universal Product Code (UPC). The UPC is a one-dimensional, optical barcode encoding product and brand information. UPC labels can be found on most consumer products in the United States. Similar systems are deployed worldwide.

Optical barcodes offer faster, more reliable, and more convenient inventory control and consumer checkout than checking out by hand. Several weaknesses of optical barcodes are that they require line-of-sight and may be smudged or obscured by packaging. In most circumstances, optical barcodes still require some human manipulation to align a barcode label with a reader. Supermarket shoppers have certainly experienced a checker struggling to scan an optical barcode.

Auto-ID systems that transmit data via RF signals, i.e. RFID, do not have the same performance limitations as optical systems. Data may be read without line-of-sight and without human or mechanical intervention. A key advantage in RF-based auto-ID systems is parallelism. Modern RFID systems may offer read rates of hundreds of items per second (RFID Tag).

2.2. Applications

As manufacturing costs dropped, RFID systems began to be used for lower-value items in industries besides transport. An example is in animal identification of both pets and livestock. Glass-encapsulated RFID devices have been implanted in millions of pets throughout the United States. These tags allow lost animals to be identified and returned to their rightful owners. These tags have a very short read range.

Livestock, particularly cattle, are often labeled with a RFID device that is clamped or pierced through their ear, attached to a collar, or swallowed. Unlike implanted pet tags, these RFID devices are rugged and able to be read from greater distances. Concerns over Bovine Spongiform Encephalopathy (mad cow) disease have motivated proposals for universal tracking of livestock with these types of RFID systems. Like transport applications, animal tracking is still essentially a low-volume, high-value market that may justify relatively expensive RFID systems.
RFID proximity cards or “prox cards” are commonly used for building access control at many companies and universities throughout the world. Similar systems have been used for ski-lift access control at ski resorts around the world. Many subway and bus systems around the world, for example in Singapore, use stored-value RFID proximity cards.

These applications also exposed some shortcomings of RFID. For instance, some RFID technologies do not operate well in proximity to liquids or metals. Each different technology has its own strengths and weaknesses, including variations in cost, size, power requirements, and environmental limits. There is no “one size fits all” RFID technology.

While RFID continues to lower the costs of tracking high-value items, an untapped and lucrative market lies in tracking cheap, everyday consumer goods. Tracking and managing the flow of goods through these supply chains is a complex and expensive enterprise.

Supply chain management and inventory control applications of this scale require an extremely low-cost tag to be economically viable. In settings like animal identification, proximity cards, electronic toll systems, or stored-value systems, RFID tags costing as much several US dollars could be justified. However, items in consumer supply-chain management and inventory control applications are much cheaper than in traditional settings. Ideally, RFID tags in these applications should be as simple and cheap as the traditional, UPC optical bar code.

EPCglobal, an RFID standards body, has developed specifications for low-cost electronic product code (EPC) tags as a replacement for the ubiquitous UPC. In the past, the lack of an open standard was a barrier to RFID adoption. The EPC standard, and to some extent, the ISO-18000 standard will make it easier for users to integrate their RFID systems.

The potential for EPC may be huge. Globally, over five billion barcode transactions are conducted daily. Even miniscule savings per transaction could translate into a huge aggregate cost savings. The market has already begun to adopt low-cost RFID on a large scale. A single RFID IC manufacturer, Philips Semiconductor, has already shipped several billion RFID chips.

### 2.3. Principles

Discussion of RFID technology tends to focus only on tag devices. It is more accurate to view RFID as a complete system that includes not only tags, but also other important components. RFID systems are composed of at least three core components:

- RFID tags, or transponders, carry object-identifying data.
- RFID readers, or transceivers, read and write tag data.
- Databases associate arbitrary records with tag identifying data.

#### 10.4. ábra - Figure 10.4: RFID Technical Information

<table>
<thead>
<tr>
<th>Type of tag</th>
<th>Typical usage</th>
<th>Typical environment</th>
</tr>
</thead>
<tbody>
<tr>
<td>PASSIVE</td>
<td>Object identification, process automation, Retail, maintenance and authentication</td>
<td>Retail sales, Warehouse and freight management, Logistics, Manufacturing and Supply chain management, Agriculture</td>
</tr>
<tr>
<td>ACTIVE (battery powered)</td>
<td>Position tracking, location-based computing, Aware systems that react to specified physical signals, constant monitoring with the use of active sensor</td>
<td>Hospitals and healthcare facilities, Warehouses</td>
</tr>
</tbody>
</table>
We illustrate the interaction of these components in Figure 10.5. In this figure, three tags are readable by one or both of two readers. The readers then may connect to databases with records associated with particular tag identifiers.

10.5. ábra - Figure 10.5: How RFID Works

2.4. Tags

Tags are attached to all objects to be identified in an RFID system. A tag is typically composed of an antenna or coupling element, and integrated circuitry. An important distinction that will be discussed later is a tag’s power source. Often tags carry no on-board power source and must passively harvest all energy from an RF signal.

There are many types of tags that offer different functionalities, have different power sources, or operate at different radio frequencies. Each of these variables helps determine which applications a particular tag may be appropriate for and what the costs of a tag may be.

Modern tags tend to implement identification functionality on an integrated circuit (IC) that provides computation and storage. In the manufacturing process, this IC is attached or “strapped” to an antenna before being packaged in a form factor, like a glass capsule or foil inlay, that is integrated into a final product.

In practice, different vendors often perform each of these manufacturing steps. Other RFID designs may be “chipless” or have identifying information hard-wired at fabrication time, i.e. “write-once, read-many” tags. Newer technologies that allow RFID circuitry to be printed directly onto a product will be discussed in Section 5.

2.5.

RFID readers communicate with tags through an RF channel to obtain identifying information. Depending on the type of tag, this communication may be a simple ping or may be a more complex multi-round protocol. In environments with many tags, a reader may have to perform an anti-collision protocol to ensure that communication conflicts to not occur. Anti-collision protocols permit readers to rapidly communicate with many tags in serial order.

Readers often power what are called passive tags through their RF communication channel. These types of tags carry no on-board power and rely solely on a reader to operate. Since these tags are so limited, may subsequently rely on a reader to perform computation as well.
Readers come in many forms, operate on many different frequencies, and may offer a wide range of functionality. Readers may have their own processing power and internal storage, and may offer network connectivity. Readers might be a simple conduit to an external system, or could store all relevant data locally.

Currently, many applications rely on fixed reading devices. Early trials of EPC at a major supermarket chain integrated fixed readers into docking-bay entrances. These readers scan tags at the pallet level as shipments of products arrive. In the long term, readers may be integrated at a shelf level as a “smart shelf”. Smart shelves would scan for tags at the item level and monitor when they are added and removed from a shelf.

RFID readers may also be integrated into hand-held mobile devices. These mobile readers would allow someone to, for example, take inventory of a warehouse by walking through its aisles. The cellular phone manufacturer Nokia is already offering RFID-reading functionality in some of their cell phones. If EPC-type tags become highly successful, interesting and useful consumer applications might arise. If this occurs, RFID reading functionality might become a common feature on cellular phones, PDAs, or other handheld computing devices.

2.6. Databases

RFID databases associate tag-identifying data with arbitrary records. These records may contain product information, tracking logs, sales data, or expiration dates. Independent databases may be built throughout a supply chain by unrelated users, or may be integrated in a centralized or federated database system. Databases are assumed to have a secure connection to readers. Although there are scenarios where readers may not be trusted, it is often useful to collapse the notions of reader and database into one entity. For example, if tags contain all relevant product information, there is no need to make a call to an off-site database. One may imagine a federated system of back-end databases, perhaps where each product manufacturer maintains its own product look-up service. In these settings, it may be useful to deploy an Object Naming Service (ONS) to locate databases associated with some tag identification value. An ONS allows a reader to find a set of databases associated with a particular tag identification value. This is analogous to the Internet Domain Naming Service (DNS) that returns addresses of name servers that can translate domain names to numerical IP addresses. ONS has not yet been adopted widely in practice.

2.7. Power Sources

As briefly mentioned before, tags may obtain their power in several different ways. The power source is an essential property of a tag, since it will determine a tag’s potential read range, lifetime, cost, and what kind of functionalities it may offer. The power source will also be important in determining how a tag may be oriented and what physical forms it may take.

There are three main classes of tag power sources: active, semi-passive, and passive. Active tags have their own source of power, such as a battery, and may initiate communication to a reader or other active tags. Because they contain their own power source, active tags typically have a much longer operating range than passive tags. Large asset and livestock tracking applications often use active tags, since the items they are attached to (e.g. railcars, shipping containers, or cattle) are high in value and have physical space for a bulkier, rugged tag.

A key feature of active tags is that they are able to initiate their own communication with readers. Advanced active tags, or “smart dust”, might even form ad hoc peer networks with each other. One useful application of active tags is in shipping containers, which can fall off ships over rough seas. These missing containers sometimes are not accounted for until well after the ship has docked. An active tag with an accelerometer sensor could detect when it was falling off a stack of containers and broadcast a log of its demise before it sank into the ocean. Active tags could also function as security alarms using the same functionality.

By contrast a semi-passive (or semi-active) tag has an internal battery, but is not able to initiate communications. This ensures that semi-passive tags are only active when queried by a reader. Because semi-passive tags do have an internal power source, they do offer a longer reader range than passive attacks, but at a higher cost.

An example application that often uses semi-passive tags is electronic tollbooths. Semi-passive tags are typically affixed to the inside of a car’s windshield. When the car passes through a tollbooth, it will initiate a query to the semi-passive tag and read an account identifier from the tag. The on-board battery lets the tag be read from a considerable distance. However, since the tag only needs to broadcast when queried, it can remain idle most of the time and save power. Semi-passive tags are also often used in pallet-level tracking or tracking components like automobile parts during manufacture.
Passive tags have neither their own power source, nor the ability to initiate communication. Passive tags obtain energy by harvesting it from an incoming RF communication signal. At lower frequencies, this energy is typically harvested inductively, while at higher frequencies it is harvested through capacitance. While passive tags have the shortest read range of all three powering types, they are the cheapest to manufacture and the easiest to integrate into products. Batteries are relatively expensive and cannot easily be incorporated into some items, like paper packaging. For this reason, passive tags are the most common tags. EPC tags are passive. Lacking an internal power source dictates many properties of passive tags. First, they cannot operate without the presence of a reader, although passive tag could temporarily cache some energy in a capacitor. Because of their necessarily weak response signal, passive tags are often

### 2.8. Operating Frequencies

Different RFID systems operate at a variety of radio frequencies. Each range of frequencies offers its own operating range, power requirements, and performance. Different ranges may be subject to different regulations or restrictions that limit what applications they can be used for. The operating frequency determines which physical materials propagate RF signals (Table 10.1). Metals and liquids typically present the biggest problem in practice. In particular, tags operating in the ultra-high frequency (UHF) range do not function properly in close proximity to liquids or metal.

#### 10.6. ábra - Table 10.1: Operating Frequencies

<table>
<thead>
<tr>
<th>Frequency Range</th>
<th>Frequencies</th>
<th>Passive Read Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Frequency (LF)</td>
<td>120-140 KHz</td>
<td>10-20 cm</td>
</tr>
<tr>
<td>High Frequency (HF)</td>
<td>13.56 MHz</td>
<td>10-20 cm</td>
</tr>
<tr>
<td>Ultra-High Frequency (UHF)</td>
<td>868-928 MHz</td>
<td>3 meters</td>
</tr>
<tr>
<td>Microwave</td>
<td>2.45 &amp; 5.8 GHz</td>
<td>3 meters</td>
</tr>
<tr>
<td>Ultra-Wide Band (UWB)</td>
<td>3.1-10.6 GHz</td>
<td>10 meters</td>
</tr>
</tbody>
</table>

Operating frequency is also important in determining the physical dimensions of an RFID tag. Different sizes and shapes of antennae will operate at different frequencies. The operating frequency also determines how tags physically interact with each other. For instance, stacking flat foil inlay tags on top of each other may interfere or prevent tags from reading properly.

### 2.9. Functionality

The basic RFID functionality is identification. When queried by a reader, tags return some identifier that may be used to retrieve other data records. However, tags may offer various other functionalities useful in different applications. The underlying principles and technologies of these various types of tags are so closely related to strict RFID tags, that they often collectively referred to as “RFID”. Although not strictly RFID, we discuss several major classes of RFID-related devices.

We split RFID-style tags into five broad classes: EAS, read-only EPC, EPC, sensor tags, and motes. These will be referred to as classes A through E. EPCglobal offers five similar classes of tag based on functionality dubbed Class 0 through Class 4. The EPCglobal classes closely align with ours, but differ somewhat. These five classes are summarized in Table 10.2.

Difficult technical and economic problems arise in class B and particularly class C devices. EAS tags are so limited in function that they are extremely simple and cheap to manufacture.

#### 10.7. ábra - Table 10.2: EPC Global Classes

<table>
<thead>
<tr>
<th>Class</th>
<th>Name</th>
<th>Memory</th>
<th>Power Source</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>EAS</td>
<td>None</td>
<td>Passive</td>
<td>Article Surveillance</td>
</tr>
<tr>
<td>B</td>
<td>Read-only EPC</td>
<td>Read-Only</td>
<td>Passive</td>
<td>Identification Only</td>
</tr>
<tr>
<td>C</td>
<td>EPC</td>
<td>Read/Write</td>
<td>Passive</td>
<td>Data Logging</td>
</tr>
<tr>
<td>D</td>
<td>Sensor Tags</td>
<td>Read/Write</td>
<td>Semi-Passive</td>
<td>Environmental Sensors</td>
</tr>
<tr>
<td>E</td>
<td>Motes</td>
<td>Read/Write</td>
<td>Active</td>
<td>Ad Hoc Networking</td>
</tr>
</tbody>
</table>
By contrast, class D and E devices offer enough functionality to justify higher manufacturing costs and can offer relatively ample resources. The challenge “sweet spot” lies in class B and class C devices, which are part of crucial systems, yet are still subject to tight resource and cost constraints. Multi-frequency passive tags operating at HF, UHF, and microwave do exist in 2006. Although they do not operate as UWB tags, supporting multi-frequency communications is possible in a passive setting.

2.9.1. Electronic Article Surveillance (EAS)

EAS tags are the most basic RFID-type tag and have been in commercial use for over 40 years. EAS tags do not contain unique identifying information, so technically are not RFID tags. They simply announce their presence to a reader. In other words, EAS tags broadcast a single bit of information – “Someone is here”. In practice, EAS tags are almost always passive and are often attached to compact discs, clothing items, or books in retail locations. EAS tags could be active or semi-passive, but the added cost of a power source would greatly outweigh adding unique identifying functionality. Because of their limited functionality, EAS tags are the simplest and cheapest to manufacture.

2.9.2. Read-only EPC

Unlike EAS tags, EPC tags contain some identifying information. EPCglobal refers to these tags as class B tags. This information may be a product code or a unique identifier. Read-only EPC tags have a single identifier that is written once when a tag is manufactured. Thus, class B tags offer strict RFID functionality. Class B tags will likely be passively powered. Although they could be semi-passive or active, again the cost of a battery would greatly outweigh the cost of re-writable memory. As the name suggests, EPC tags are used in basic item tracking applications. However, many other practical applications of tags, such as smart cards or proximity cards, are using tags with read-only memory that offer simple identification. Read-only EPC tags are fairly simple and may even be “chipless”, thus are relatively cheap.

2.10. EPC

Class C refers to simple identification tags offering write-once, read-many or re-writable memory. Rather than having an identifier set at manufacture time, identifiers may be set by an end-user. If an EPC tag offers re-writable memory, its identifier may be changed many times. Class C tags still offer strictly RFID functionality. Class C EPC tags may be used as a logging device, or can emulate Class B read-only EPC tags. In practice, class B EPC tags may be passive, semi-passive, or active. Strict RFID functionality includes class B tags. Supporting non-volatile, writable memory adds complexity to class B tags. Consequently, they may be significantly more expensive than read-only EPC or EAS tags.

2.11. Sensor Tags

Sensor tags may contain on-board environmental sensors, and may log and store data without the aid of a reader. These types of tags will be referred to here as class D. Sensor tags offer more than strict RFID functionality, and are typically not thought of as RFID. 3 Due to technical issues, “re-writeable” tags in practice can only be written some fixed number of times; perhaps several hundred re-writes. Many sensor tags may form a “sensor net” that monitors a physical area’s environmental properties. This may include temperature changes, rapid acceleration, changes in orientation, vibrations, the presence of biological or chemical agents, light, sound, etc. Because they operate without a reader present, sensor tags must necessarily be semi-passive or active. An on-board power source and sensor functionality comes at a much higher manufacturing cost.

2.12. Motes

Class E tags, or “smart dust” motes, are able to initiate communication with peers or other devices, and form ad hoc networks. Motes are essentially general pervasive computing devices and are much more complex than simple EPC-style RFID. Because they are able to initiate their own communication, mote devices are necessarily active. Commercial motes are available from Crossbow Technology. Ongoing research into smart dust and motes is being conducted at the University of California, Berkeley and Intel.

2.13. Standards

The two most relevant RFID standards are the International Organization for Standardization’s ISO/IEC 18000 standard and EPCglobal’s standards.
These standards are not competing, and it is conceivable that EPCglobal’s standard could eventually be adopted into an ISO standard. EPCglobal defines specifications for EPC-type tags operating in the UHF range. The ISO 18000 standard has 6 parts addressing different frequency ranges:

- Part 1 – General standards
- Part 2 - LF
- Part 3 - HF
- Part 4 – Microwave, 2.45 GHz
- Part 5 – Microwave, 5.8 GHz (withdrawn)
- Part 6 - UHF


2.14. Challenges

2.14.1. Technical

RFID systems still face many technical challenges and obstacles to practical adoption. A major hurdle is simply getting RFID systems to work in real-world environments. Systems that work perfectly in a lab setting may encounter problems when faced with environmental noise, interference, or human elements. Readers and tags often experienced interference caused by other wireless systems, or unknown sources. This type of interference was not systematic, and usually resulted from environmental idiosyncrasies. Addressing these issues required trial and error, and practical experience to recognize what was causing the problem. For example, simply repositioning or re-aligning readers would often address performance issues.

Software support for RFID is still in its early stages as well. Getting distributed back-end database look-ups to work in practice is a complex task that is often glossed over in RFID literature. In particular, key management and network connectivity issues are often underemphasized. Many vendors do currently offer RFID software solutions. However, in the coming years is likely that the industry will consolidate onto several standardized software interfaces.
The point of this digression is to emphasize that, like most information technology systems, RFID systems still require practical expertise to install, configure, and manage. End-users should expect to experience mundane technical complications that arise while implementing RFID. Despite marketing claims to the contrary, RFID is not a “magic bullet” that is simple to implement out of the box.

2.14.2. Economic

A key hurdle that still remains in RFID systems is simply cost. This is especially the case with EPC item-level tagging. A commonly cited price point where item-level tagging is supposed to be economically. As the market grows, RFID costs will drop and new applications will become economical, especially as more investment is made into back-end architectures. However, for the near future, the costs of many envisioned applications, particularly for EPC tags, are simply not justified.

2.14.3. Security and Privacy

Many concerns have been expressed over the security and privacy of RFID systems. Traditional applications, like large-asset tracking, were typically closed systems where tags did not contain sensitive information. Tags on railway cars contained the same information painted on the side of the cars themselves. However, as more consumer applications are developed, security, and especially privacy, will become important issues.

2.15. Future Technologies

Two promising technological developments especially relevant to RFID are printed circuits and organic components. These technologies have the potential to greatly lower manufacturing costs and to produce RFID tags built out of flexible plastic materials, instead of silicon. The long-term vision is that a large-scale packaging manufacturer could print RFID tags directly into paper or plastic as it is produced. Product makers would not use this RFID-enhanced packaging material as they normally would. One advantage in terms of privacy is that RFID tags would only be attached to product packaging, and not the product itself. This technology is still years away from being economic and there are many hurdles to overcome. Currently, circuits printed by an inkjet have a very low resolution; circuit gates take much more surface area than traditionally fabricated circuits. Other technologies like gravure printing also produce relatively large circuit surface areas. Regardless, much research is being focused on organic components for other purposes, like flexible displays. Developments in this area will benefit RFID, potentially opening the door to many inexpensive and interesting future applications.

3. Logistical Tracking & Tracing

Information is of primary importance in logistic processes. Improved transparency makes information on the location and availability of materials and equipment accessible to all authorized stakeholders regardless of their location.

10.9. ábra - Figure 10.7: EAN

This development is promoted by the implementation of automatic identification and tracking systems. Classical inventory systems keep track of lots or batches, whereas RFID systems allow the unique identification of items.
and a shift to a ‘per unit’ or serial-level inventory system, which allows the tracking of individual items throughout the supply chain. This poses new requirements for the existing IT systems. Integration of the RFID system in existing IT systems, such as ERP systems, is critical to the success of the implementation. Different commercial applications are available. In this context, RFID can be seen as an enabling technology for automatic tracking and tracing systems. To date, RFID has slowly started to replace and complement manual labels, trading stamps, barcodes as well as methods based on optical character recognition. However, barcode technology is by far the most widely adopted technology for identification in logistics applications. International retailers like Wal-Mart, Metro Group and Tesco are leading the development and have boosted the introduction of RFID in logistics by mandating their largest suppliers to implement RFID on containers and pallets.

10.10. ábra - Figure 10.8: EPC Global Network

Because of this, the retail sector has become a global development environment where the RFID solutions are being tested, piloted and implemented. Other business sectors are closely following these experiments and making preparatory work to utilise the new technology. Effective utilisation of RFID may require large changes in the existing processes, a fact that potential users might not have considered. Until now, RFID applications have mostly concerned separate, closed and in-house systems. The use of passive UHF RFID technology is increasing the most rapidly, due to the low price, good standardisation situation and sufficient performance (3-4 m reading range). Due to their increased use, UHF RFID tags which are designed for specific applications (e.g. mounting on metal) have also become commercially available. A growth in active RFID technology is foreseen in RTLS (Real Time Location Systems), due to improved standardisation, lower prices and the possibility for integration with existing IEEE 802.11 (WLAN) infrastructure. RFID development has mostly concentrated on components. As the maturity of RFID systems improves, the type of products which vendors offer changes from individual tags and readers to completely integrated systems (complete environment of identification portals including antennas and interfaces to existing IT systems). Even though the focus at present has been on tracing the higher level logistic units (containers, pallets) pilots involving item level tagging of products sold to consumers (open loop, B2C) have been introduced. This could dramatically extend the use and potential of RFID applications towards the end of the supply chain. This has aroused concern among some consumers who are worried about their privacy. Consumer protection organisations are advocating a policy whereby the consumer should always be told when an item contains an RFID tag and have a choice to have it removed or destroyed (or deactivated). Ratification of the EPC UHF Class 1 Gen 2 Standard at the end of 2004 and the ISO-18000-6C based on the Gen 2 specification in 2006 has removed a major obstacle to the implementation of low-cost RFID. In addition to the technology standardisation, efficient utilisation of RFID requires agreement between stakeholders on common modes of operation. Interoperability will be particularly important in global supply chains. More pressure and demand for new identification solutions are generated by the new modes of operation such as pull-based flow control, reducing inventories, legislation (tracking of food, pharmaceuticals etc.), cost reduction, automation, communication between partners and visibility of the supply chain, decreasing waste and improving security. The main factors that have slowed down RFID implementation are the investment costs which are regarded as high and the uncertain expectations for the return on investment (ROI). There is not enough or not specific enough information about the benefits, especially for manufacturers and suppliers of parts, who bear the cost of tagging. On the hardware side especially the UHF readers are rather expensive. Unsuccessful pilots with low readability rates have created an image of immature technology for passive UHF RFID technology and delayed implementation, although there are also reports of continuous improvements in the technology, especially when the latest firmware modifications are applied to readers. Automatic identification develops towards systems, which can read the IDs of all products automatically without the need to stop the process or manual intervention. The technology used has to be able to read each of the individual tags. The long term vision in logistics is a system providing the necessary real-time information on the supply chain, this information being extensively utilised. RFID is the key technology to make this happen. Real-time information on the location, contents and conditions of individually identified shipments,
products, transport units, and transport vehicles can be gathered in a controlled manner. The collected data can be combined with the planning information and processed into appropriate information to be used at different stages of the process. Product history will become available via B2B networks. The information can be distributed effectively and in real time to the stakeholders. The price of RFID tags will decrease below 5 cents and item-level tagging will be applied on a larger scale. Miniaturised reader modules and microchips promote the integration of RFID readers into mobile phones and PDA devices.

4. Production, Monitoring & Maintenance

Most of the RFID-based solutions used today in the field of manufacturing & maintenance are limited to tracking and tracing of parts or tools. Compared with the previous identification means, RFID provides a fast and automatic capability to identify parts without the need to see or contact the ID. However, the reading performance of RFID tags (reliability, range…) for several products/processes is not yet totally sufficient to guarantee the economic benefit of adopting RFID compared with the well established systems using barcode, 2D matrix or even name plates with written part or serial number. Most of the RFID-based solutions are either “closed loop” (RFID is not used outside a company) or “semi closed-loop” (RFID is not used outside a supply chain), although “open loop” solutions (RFID is used anywhere during the product lifetime) would be of great use, especially for maintenance purposes. Monitoring of product usage and status during their service life is important to allow the development of condition-based or predictive maintenance services. RFID tags are a key element to this since they can wirelessly communicate the identity of a unique product to an information system where this identity can be merged with other data and processed. However, the available RFID technology has not yet the performances to allow industrial use: in terms of robustness, temperature capability, sensing and communication functions, miniaturisation.

10.11. ábra - Figure 10.9: Electronic Product Code

The standards for identification/coding of parts, objects and services related to manufacturing or maintenance are generally already established (although different from one domain to another) and RFID shall comply with these standards. The competitiveness of manufacturing and maintenance enterprises is strongly determined by their ability to quickly integrate customer requirements for new products and services within complex business networks (i.e. from SME type organisations up to global actors with diverse organisational and technical maturity levels). In this regard, the integrated enterprise management still lacks solutions for both self-organisation of single network entities, as well as for continuous dynamic adjustment of the overall network of actors, with particular emphasis on the need for a ubiquitous-but-secure access to product related data and services.

Initially, manufacturing and maintenance will benefit from RFID-based solutions developed for track and trace of components allowing better automation of supply chain and resource management. Beyond this, further benefits from RFID-based solutions will be obtained by using the unique capability given by RFID tags to be read wirelessly and automatically without human triggering and processing in order to inform the surrounding information system in real time of its identity, through which a connection can be made to possibly significant data content directly “attached” to it.

10.12. ábra - Figure 10.10: Product Identification
Note that in some scenarios, additional data will be recorded to an RFID tag, while in other scenarios, a simpler tag is used, carrying a unique ID or ‘licence-plate’, which is then used for identifying sources of information on the network, as well as retrieving relevant records, since the unique ID read from the tag can also be used as a database key. The networking of information systems will then make it possible to know in real time, from anywhere, the configuration of an individual product throughout its lifetime, to access necessary information on this product and provide tailor-made added value services like:

- Event-driven auto-adaptation of process, workflow, tools and resource allocation to the flow of parts being processed (references and status)
- Automated manufacturing or maintenance documentation and distributed decision making tools tailored to the actual product configuration (references and status),
- New and more complete “all included” services for product operation including real time remote monitoring of products.
- Automated detection of counterfeit, misused or poorly maintained parts.

The RFID technology itself will develop from simple identity devices to include broader sensing, mobile data storage and networking capabilities. This will allow both more sophisticated and local processing, taking into account the product history and status and the changes of configuration. These services will be provided either locally (by the machine or product monitoring systems) or remotely (by service providers via the internet or dedicated connection). The real time, complete, accurate and shared information on the status of individual products would directly facilitate the dynamic detection and integration of changing market needs or disturbances. This perspective for RFID developments shall not be considered by itself, but only in a global perspective of ICT developments. In the same way, RFID may not be the best solution for every case of product identification and will not immediately replace all other means for part identification. RFID-based solutions shall therefore remain interoperable with them.

5. Product Safety, Quality and Information

At the current point in time, RFID technology is beginning to pervade the domain of trade with tags being introduced mostly to logistical units such as pallets, but not so much on the item level, yet. There are already several examples of pilot installations also targeted at consumers, most notably the Future Store initiative of the Metro group, which is pioneering the trials of RFID related applications for the benefits of the end consumers. Manufacturing and production industries are in general still quite hesitant regarding the adoption of RFID technologies for reasons of missing interoperability standards and also because of lack of knowledge about best practices. There are however already some success stories in the field of product safety and quality, even for fast moving goods, such as a project on RFID for the food supply chain. In general however, most producers lack both the knowledge of the respective processes (esp. with regard to safety regulations) and the potential
connections to related technologies that make sense for integration in the field of product safety and quality such as temperature or humidity sensors.

When we regard the issues of product safety and especially product quality, we have the clear objective to provide detailed information about the history of a product to the end consumer. This will help to create trust and transparency in sensitive product areas such as perishable or sensitive goods, but also pharmaceutical and luxury goods and high-value goods, some of which are highly composite products. Depending on the consumer’s context and demands, correct and complete information has to be provided. The classical “pull” mode, where consumers actively search and query for information, could easily be extended by a “push” mode in order to ensure the provisioning of up-to-date information at the right place, at the right time. We want to achieve fully RFID-enabled product lifecycle processes for the businesses in the trade domain. After packaging and labelling the respective cases, pallets, and potentially even items with RFID tags, information about shipments will be stored and tracked at different stages in the product lifecycle throughout the complete supply chain. By augmenting this track & trace information with sensor data regarding relevant properties such as temperature, humidity, velocities of movement etc. we can enable all intermediaries and the buyers of products to get an insight into a product’s history and to verify the quality of goods. Especially when sufficient transparency across company borders is given, the tracking and tracing of composite products can potentially lead to more efficient quality management when it comes to exactly locating a faulty batch, production line or manufacturer, and meeting optimal quality improvement measures or precisely directed recall campaigns. In the long run and in certain domains, we will see an increase in the utilization of sensor technologies in general, which allow for contributions to product safety far beyond the typical track & trace based approaches that are currently being implemented. Especially for certain product types such as perishable goods there are strong indications that measuring e.g. temperature or humidity on the way through the supply chains and altering delivery models accordingly can greatly enhance product quality. Likewise, in some areas, e.g. hazardous goods, it has already been demonstrated in research trials that embedded, communicating sensors can help prevent industrial accidents by giving feedback about storage and handling conditions. Naturally, the ultimate goal of data exchange with the customer requires the previous installation of respective B2B exchange infrastructures based on global standards. Not-for-profit standards bodies such as EPCglobal are overseeing the development of an architecture of open standard interfaces for the necessary large scale B2B infrastructures, based on requirements from end users and with the active participation of several technology solution providers. The benefit of such open standards is to foster a competitive marketplace for solutions, in which the end-user can choose among interoperable solutions from multiple providers. In addition to this, we also need legislative bodies and institutions in order to establish trust and security among the respective businesses in the trade domain. The implementation of RFID enabled data exchange infrastructures at a broad level will significantly and sustainably influence the importance and the future development of all related IT systems.

10.13. ábra - Figure 10.11: Tracebility Platform

6. Payment

Payment applications have many interesting new technologies and development directions. Closed-end payment schemes, interoperability of ticketing and payment instruments, mobile payment applications, more stable security solutions, electronic purse and e-cash are just a few ones to be mentioned. From this list we selected electronic cash as the research subject in focus, because it has the highest relevance due to its economic impact
and its strong reliance on contactless technology. Europe still relies on the independent, national payment systems, a situation which is not expected to change any time soon. The introduction of EURO in 12 European countries was a major step forward. However the depth of integration probably stops at this stage with further improvements only to be expected from the enlargements of the EURO zone with some of the new EU members. With this higher level of interoperability achieved on a pan-European level, substantial expenses could be saved for the general public as well as for industry and services. The avoidance of currency exchange and the associated risks improved operating efficiency and contributed to a higher level of transparency, resulted in increased competition and the establishment of a truly European market. Technology on the other side did not keep pace with this development. Cash is still king. The vast majority of payments both in terms of transaction numbers and value are still performed with cash. The decrease of cash usage is minimal and barely discernible especially in the case of low value payments. In a payment system or instrument there are four aspects that matter: speed, security, expense and general acceptance/usage. Cash is performing very poorly compared to other possible solutions in all of the first three aspects, but it is unbeatable in the fourth, but most important aspect. General acceptance and usage is the most important characteristic of any payment method and none of the newly introduced and emerging solutions could even come close to cash in this aspect. There were recently a number of initiatives by banks to introduce alternative payment instruments like mobile payment and electronic purse, but none of them succeeded so far; none became a generally accepted payment method. (Part of the reason of failure was that proprietary solutions were pursued where the necessary network effect could not be achieved and also often the technology used was inappropriate, like in case of stored value payments.) In the last 50 years the payment card was the only really successful, innovative solution that achieved widespread acceptance and adoption, and even its penetration only really accelerated during the past decade. New offerings of closed loop payment schemes and prepaid cards operated by new 3rd party service providers try to capitalize on the inability of the incumbent financial service providers to satisfy the new requirements but this road only leads to further segmentation of the market and deterioration of economics.

The future of European small value payments, payments that cover most of our everyday purchases, should be based on a combination of state of the art technologies. The solution is an open payment scheme based on a contactless proximity application that is combined with remote mobile communication and PKI based security. The new payment instrument is an offline purse that stores electronic value and acts as a kind of stored value account issued by banks. The purse can be loaded remotely from existing bank accounts resulting in full integration into the present financial infrastructure. The e-cash would be denominated in EUR ensuring overall acceptance across the whole EURO zone. The new e-cash or payment purse has to be a generally accepted, official payment instrument. The issuing organisations, regulators would ensure that the money can be used and is generally accepted and that existing acquiring and clearing infrastructures can be used. This capability would take care of the most important usage barrier. The contactless feature of the application will ensure speed and convenience. To date, many of the new approaches failed because users were unfamiliar with the way of operation of the new solutions and inconvenient usability features and frequent failures practically prevented customers from trying the new methods. The speed of payment plays directly into the bottom line of merchants and makes a huge difference in the customer service of the shoppers. The remote communication feature will provide direct access to one’s bank account allowing customers to always recharge their purses whenever necessary, wherever they are. The purse can be implemented within specially designed mobile handsets – based on NFC technology – but may also be simple plastic cards, or may have any other type of form. The mobile handset housing the purse or a PC with an internet connection would just provide the necessary communication channel with the back office system of the partner financial institution for the top-up function. In the non-phone based embodiments, the contactless interface would also be the access channel of the application. The third important component of the new payment instrument is the PKI based security architecture. The use of encryption and digital signature does not only provide a very high level of security much tighter than that of existing solutions but also enables offline P2P money transfer and definition of funds which are earmarked for special purposes. Such special, dedicated funds could be assigned to people under a certain age if limitations are in place regarding their consumption of specific products and services. The same feature can be used if money is provided to someone for a special purpose and not for general use. (buy a book not candy, social support not to be spent on booze) The security architecture will also be used to ensure protection of sensitive personal information and to prevent misuse of user or transaction data. To build an acceptance network with relatively low cost and high density mobile handsets with contactless interface will also be used as acceptance devices. With such an architecture plain everyday mobile phones, or more specialised devices for the merchants can be used as payment purses as well as POS terminals supporting all kinds of transactions, even P2P ones, between any two parties. Besides being cheaper and more economic than the traditional paper money, the new e-cash would have additional benefits, like increased security – would comply with anti money laundering regulations – like contribution to the whitening of the economy, and to health and environment protection.
7. Management and organization in agri-food chains and networks

The study of chains and networks has grown rapidly as an area of academic interest from the early 1990s onwards. Economists, sociologists, and management scholars acknowledged that studying firms individually could not sufficiently explain real life phenomena, and did not lead to useful recommendations for practitioners. Inter-organizational relationships became a whole new area of research, emphasising the collaboration and coordination between firms. Interestingly, from both a social science perspective, looking at human and organizational behaviour and performance, and a technical perspective, studying the serially ulinked processes and their interfaces, scholars have greatly contributed to this emerging field of chain and network studies. The study of (agri-food) chains and networks can be divided into three streams. One stream is the Supply Chain Management (SCM). SCM seeks to optimize the processes within and between the firms participating in the chain, such as logistic processes, (electronic) information exchange and quality control. The second stream is the economic organization approach, e.g. Transaction Cost Economics. This approach, with its roots in both economics and organization theory, seeks to analyse transactions and to design the most efficient governance structure for a particular transaction. The third stream is the network approach, emphasising multiple relationships among firms. This approach investigates the network instead of the chain as firms are often participating in multiple chains, and have relationships with companies that are not part of the supply chain (such as providers of equipment, labour, advisory services, accountancy, etc). Network studies go beyond the functional perspective (as in SCM) or the purely economic perspective (as in TCE), by incorporating the embeddedness of a firm in a (social) network.

7.1. Governance

New institutional economics has been one of the main bodies of literature applied in the study of the organization of agri-food chains and networks. New institutional economics uses a transaction costs perspective. The focus is on the transaction (understood as transfer of rights) because taking advantage of the division of labour requires complex devices at the micro level (modes of organising the transfer) as well as at the macro level (institutions facilitating and enforcing this transfer). The emphasis on transaction costs is explained by the costs of those devices: comparing these costs is crucial for understanding how (and what) institutions and organizations allow benefiting from specialization. This economic organization perspective on transactions enables to study the efficiency of the various organizational structures that can be found in the agri-food industry.

7.2. Collaboration

One of the organizational issues in chains and networks of particular importance in the agri-food industry is how (small) primary producers are organised in a horizontal way to strengthen their position in vertical relationships. Producers often have established collective organizations, to benefit from bargaining power, risk sharing and/or economies of scale. These so-called economic producer organizations (POs) play a major role in national and international agri-food supply chains, as they are an intermediary between an often large number of individual farming households and a small number of customers (or suppliers). POs are also an important instrument for rural development, because of their ability to support both economic development and social cohesion. POs have several economic functions, such as collecting, processing and marketing agricultural products, implementing quality assurance programs, and giving advise and training to their members. By exploiting economies of scale and scope as well as by reducing transaction costs, POs can improve the efficiency and efficacy of agri-food supply chains. In developed countries, POs are experiencing restructuring processes, due to their changing role and function in (international) supply chains and networks. Changes in competitive conditions, consumer demands and public policies have forced the producer organization to shift from a producer (or supply) orientation to a customer (or demand) orientation. POs as well as their members have increased their effort in marketing, innovation, and internationalization. As the PO - customer relationship has become relatively more important compared to the producer - PO relationship, we have witnessed shifts in investment patterns, in coordination mechanisms, and in control relationships.

8. Questions

1. What are the main problems of food safety?
2. What kind of technologies are developed for identifications?
3. What is the EAN and GS1?
4. How works the RFID technique?
5. What are the parts of a RFID System?
6. What is the EPC Global?
11. fejezet - Decision Support Systems

1. Introduction

Decision making is a major component of living and, therefore, a fascinating topic for discussion and investigation. Several fields in science also occupy themselves with the nature of different aspects of decision making: philosophy, psychology, sociology, economics, etc. In most institutional decision-making problems, there are three main aspects of concern:

- The information about the current situation and, possibly, about the past.
- The processes that are to be influenced by the decisions.
- The actual decision-making process.

For instance, decision making concerning acid rain illustrates the above three aspects clearly, as follows:

- The information consists of huge amounts of data regarding industrial, agri-cultural, and automotive activities. It also consists of data about wind and sun activities and about the types of power stations used.
- The processes to be influenced by the decisions consist of the production and emission of polluting material, together with the atmospheric transformation (chemical reactions and transportation) and the deposition process. Therefore the processes are basically of a physical and economic nature.
- The decision-making process consists of several interacting subprocesses at the local, national, and international levels.

In the history of decision support and decision analysis, one sees that many tools and methods have been developed to help make decisions. We also see that most of these methods and tools concentrate on one of the three main aspects of institutional decision-making problems. For instance, in most college textbooks dealing with operations research, we find several generic models describing the relation of the basic processes with the possible decisions. Each model is usually presented together with at least one technique for constructing a most favorable decision. Linear programming models and the simplex method provide the most classical example.

The first aspect, information, has been the source of inspiration for particular types of information systems that consist of one or more data bases and special methods for arranging the information. Typical examples are geographic information systems (GIS) and management information systems (MIS), each with their own way of structuring and storing the information and with their own way of dealing with the information.

Finally, the decision-making process has inspired several approaches to the structuring of the process of reaching decisions and ways of comparing different alternatives. The process of reaching decisions may be complex because the decisions themselves are complex and consist of many subdecisions, which should be taken in the right order (decision trees can be an appropriate tool to model this type of situation). The process of reaching decisions may also be complex because of the number of persons, departments and other groups involved, each with their own interests, constraints, and ambitions. In such cases, group decision-making methods or even negotiation support can be helpful.

One of the most striking features of the existing decision support methodology is that most tools and methods concentrate entirely on one of the three aspects and often only on a part of that aspect. This is understandable, because often it is difficult to integrate the other aspects. For instance, GISs are able to handle a wealth of information in a very user-friendly and enlightening way. However, it is difficult to translate this information for the basic processes to models in order to use this information for deriving consequences of possible decisions. A direct linkage is possible only if the process models are closely related to the information structure, as in guiding car-systems, where an optimal route is indicated on a map. This decision selection can easily be adapted if information about traffic jams or blocked bridges becomes available.
For more strategic decisions, however, it is unlikely that the process models relate so closely to the information structure. Even if something can be done, the approach is probably specific to a particular problem and therefore not generally applicable. One reason for this situation is that there are accepted standards for keeping information in data bases, but there are no accepted standards for model building.

Quite often, methods are implemented as if the designers are not even aware of the other aspects. As a result, implementations have been developed that are practically infeasible. For example, approaches to modeling and optimization of the basic processes have been commonly advertised that do not fit with the way information becomes available, or, more seriously, with the way the decisions are chosen.

Many institutional decision processes relate to well-defined processes of a technical, physical, or economic nature. This is particularly true for environmental problems, like the acid rain problem. In such cases there is usually a lot of knowledge about the basic processes and about the way they are influenced by possible decisions. However, not all institutional decision processes possess this property. Consider, for instance, the structuring of the funding organization for research in Poland. In such a case there is certainly a basic process to be influenced by the choices to be made, namely, the research process in Poland including quality, quantity, and distribution of research topics. However, there is not much accepted hard knowledge about the basic process and the core of the decision problem consists of a careful structuring of the influences of relevant institutions and of the decision process.

On the other hand, if there is a well-defined basic process and if the core problem essentially consists of deciding how to influence that basic process, then it is obvious that the starting point for any reasonable approach should be the modeling of the basic process, including the changes caused by possible decisions. As stated above, environmental policy problems usually belong to this class.

2. Historical Overview of Decision Support Systems (DSS)

During the late 1970s the term “decision support systems” was first coined by P. G. W. Keen, a British Academic then working in the United States of America. In 1978, Keen and Scott Morton published a book entitled, Decision Support Systems: An Organizational Perspective defined the subject title as computer systems having an impact on decisions where computer and analytical aids can be of value but where the manager’s judgment is essential. Information systems (IS) researchers and technologists have developed and investigated decision support systems (DSS) for more than 35 years.

Van Schaik (1988) refers to the early 1970s as the era of the DSS concept because during this period the concept of DSS was introduced. DSS was a new philosophy of how computers could be used to support managerial decision-making. This philosophy embodied unique and exciting ideas for the design and implementation of such systems. There has been confusion and controversy in respect of the interpretation of the decision support system notion and the origin of this notion originated in the following terms:

• Decision emphasises the primary focus on decision-making in a problem situation rather than the subordinate activities of simple information retrieval, processing or reporting.

• Support clarifies the computer’s role in aiding rather than replacing the decision maker.

• System highlights the integrated nature of the overall approach, suggesting the wider context of machine, user and decision environment.

DSS deal with semi-structured and some unstructured problems. With the ever-increasing advances in computer technology, new ways and means of computer-assisted decision-making was born. As a result hereof, over the passage of time, different DSS definitions arose:

• DSS a “model-based set of procedures for processing data and judgments to assist a manager in his decision making.”

• “Decision Support Systems couple the intellectual resources of individuals with the capabilities of the computer to improve the quality of decisions. It is a computer-based support system for management decision makers who deal with semi-structured problems.
• A decision support system is an interactive system that provides the user with easy access to decision models and data in order to support semi-structured and unstructured decision-making tasks.

• DSS “a computer-based information system consisting of hardware/software and the human element designed to assist any decision-maker at any level. However, the emphasis is on semi-structured and unstructured tasks.”

• DSS a computer-based systems that help decision makers confront ill-structured problems through direct interaction with data and analysis models.

• DSS are computer-based systems that bring together information from a variety of sources, assist in the organisation and analysis of information and facilitate the evaluation of assumptions underlying the use of specific models.

• DSS “a computer-based information system that combines models and data in an attempt to solve semi-structured and some unstructured problems with extensive user involvement.”

From these definitions it seems that the basis for defining DSS has been developed from the perceptions of what a DSS does (e.g., support decision-making in semi-structured or unstructured problems) and from ideas about how a DSS’s objectives can be accomplished (e.g., the components required and the necessary development processes).

There are several requirements for a DSS which must embrace a definition of a DSS. These are that a DSS:

• requires hardware;

• requires software;

• requires human elements (designers and end-users);

• is designed to support decision-making;

• should help decision makers at all levels; and

• emphasises semi-structured and unstructured tasks.

The technology for DSS must consist of three sets of capabilities in the areas of dialog, data and modelling. The researchers make the point that a good DSS should have balance among the three capabilities. It should be easy to use to allow non-technical decision makers to interact fully with the system. It should have access to a wide variety of data and it should provide analysis and modelling in a variety of ways. Many early systems adopted the name DSS when they were strong in only one area and weak in the other. Figure 1 shows the relationship between these components in more detail and it should be noted that the models in the model base are unlinked with the data in the database. Models can draw coefficients, parameters and variables from the database and enter results of the model’s computation in the database. These results can then be used by other models later in the decision-making process. Figure 1 also shows the three components of the dialog function wherein the database management system (DBMS) and the model base management system (MBMS) contain the necessary functions to manage the data base and model base respectively. The dialog generation and management system (DGMS) manages the interface between the user and the rest of the system (Figure 11.1).

11.1. ábra - Figure 11.1: The structure of a DSS
Three levels of technology are useful in developing DSS and this concept illustrates the usefulness of configuring DSS tools into a DSS generator which can be used to develop a variety of specific DSS quickly and easily to aid decision makers—see Figure 2. The system which actually accomplishes the work is known as the specific DSS, shown as the circles at the top of the diagram. It is the software/hardware that allow a specific decision maker to deal with a set of related problems. The second level of technology is known as the DSS generator. This is a package of related hardware and software which provides a set of capabilities to quickly and easily build a specific DSS. The third level of technology is DSS tools which facilitate the development of either a DSS generator or a specific DSS.

Some authors identified five specialised types of DSS:

- Text-oriented;
- Database-oriented;
- Spreadsheet-oriented;
- Solver-oriented; and
- Rule-oriented.

11.2. ábra - Figure 11.2: Types of DSS
Others classified DSS as ad hoc DSS or institutional DSS. An ad hoc DSS supports problems that are not anticipated and which are not expected to recur. An institutional DSS supports decisions that reoccur. Hackathorn and Keen (cited in Power, 2003a) identified DSS into three interrelated categories:

- Personal DSS;
- Group DSS; and
- Organisational DSS.

11.3. ábra - Figure 11.3: Types of DSS Support

The following DSS frameworks help categorise the most common DSS currently in use:
• Communications-Driven DSS: These systems are built using communication, collaboration and decision support technologies.

• Data-Driven DSS: These systems analyse large “pools of data” found in major organisational systems and they support decision-making by allowing users to extract useful information that was previously buried in large quantities of data. Often data from various transactional processing systems (TPS) are collected in data warehouses for this purpose. Online analytical processing (commonly known as OLAP) and data mining can then be used to analyse the data.

• Document-Driven DSS: These systems integrate a variety of storage and processing technologies to provide complete document retrieval and analysis.

• Knowledge-Driven DSS: These systems contain specialised problem-solving expertise wherein the “expertise” consists of knowledge about a particular domain (and understanding of problems within that domain) and “skill” at solving some of those problems.

• Model-Driven DSS: Early DSS developed in the late 1970s and 1980s were model driven as they were primarily standalone systems isolated from major organisational IS that used some type of model to perform “what if” and other kinds of analysis. Such systems were often developed by end-user groups or divisions not under central IS control. A DSS is not a black box—it should provide the end-user with control over the models and interface representations used. Model-driven DSS emphasise access to and manipulation of a model.

Despite all the rapid developments of the late 1980s, 1990s, and early 2000s, DSS as a field is now at a crossroads. Some functions that were once considered part of DSS now appear to be migrating to other areas. For example there is an increasing trend to integrate and embed decision support applications into operational systems (e.g., fraud detection system embedded in credit card processing).

3. Future trends

In future, it is envisaged that traditional DSS applications will be extended to a larger number of potential applications where the data required is only an interim stage or a subset of the information required for the decision. This will require the construction of DSS where the end-user can concentrate on the variables of interest in their decision while “other” processing is performed without the need of extensive end-user interaction. Some future trends for DSS are suggested:

• Organisations that consolidate there is into a single environment reduce administration and license costs. By consolidating organisational data into a Web visualisation application, will facilitate better decision support.

• All organisations use metrics and key performance indicators to undertaken business and remain competitive. With the advent of Web-based technologies (e.g., portal technologies), a decision support portal will be able to present key information to the right audience.

• In the future all data collection and analysis will be automated. This will “free up” domain experts from verifying the validity of data from TPS and data warehouses allowing them to act on the information from DSS instead.

• There will be an increase in visualised information in context with user-centric displays. By having the most recent data correlated and aggregated, will allow for better decisions and which are more relevant to a user’s current conditions.

• There will be a surge to use advanced display techniques to highlight key issues. Consequently the design of future DSS interfaces will receive greater prominence since the interface should bring attention to the most important areas almost immediately.

• Decision support technology will continue to broaden to include monitoring, tracking and communication tools to support the overall process of unstructured problem solving. The broadening of this technology will be as a result of an increased availability of mobile computing and communication.
4. Decision Support Systems in Agriculture: Some Examples

Agricultural knowledge, that is the combined experience of how to grow and produce food, fiber, and bioproducts while securing a livelihood from the land, is extremely complex, comprised of multiple disciplines, multiple persons, with multiple levels of abstraction. Producer decisions range from considering the details of how a plant needs protection from pests and diseases, to planning commodity trading and marketing—all sometimes in a matter of minutes. Other producers’ worries range from which variety of food crop to plant, to which field to plant first, to issues of food availability and alternative sources of income should food production fail. With such complexity, uncertainty, and variation over time, it is not surprising that agriculture as an enterprise is considered highly risky.

4.1. Crop Production

One important goal in agricultural crop production is to develop less intensive and integrated farming systems with lower inputs of fertilizers and pesticides, and with restricted use of the natural resources (water, soil, energy, etc.). The main objectives of these systems are to maintain crop production in both quantitative and qualitative terms, maintain or preferably improve farm income, and at the same time reduce negative environmental impacts as much as possible. Achieving all of these objectives is a prerequisite for sustainable agriculture.

The Thematic Strategy on the Sustainable Use of Pesticides adopted in 2006 by the European Commission aims to establish minimum rules for the use of pesticides in the Community so as to reduce risks to human health and the environment from the use of pesticides. A key component of this Strategy is implementation of Integrated Pest Management (IPM), which will become mandatory as of 2014. In the context of IPM, the EU will develop crop-specific standards, the implementation of which would be voluntary. IPM creates synergies by integrating complementary methods drawing from a diverse array of approaches that include biocontrol agents, plant genetics, cultural and mechanical methods, biotechnologies, and information technologies, together with some pesticides that are still needed to control the most problematic pests and to manage critical situations. Concepts of IPM, IP, and IF are based on dynamic processes and require careful and detailed organisation and management of farm activities at both strategic and tactical levels. This means that time must be invested in management, business planning, data collection and detailed record keeping, and identification of required skills and provision for appropriate training to ensure safe farm operation. In IPM, IP, and IF, farm managers must also know where to obtain expert advice, and they must be willing to accept scientific and technical advances that benefit the environment, food quality, and economic performance, and that therefore can be integrated into the crop management as soon as they are reliable.

Decision Support Systems (DSSs) collect, organize, and integrate all types of information required for producing a crop; DSSs then analyse and interpret the information and finally use the analysis to recommend the most appropriate action or action choices. Expert knowledge, management models, and timely data are key elements of DSS and are used to assist producers with both daily operational and long-range strategic decisions. Computer-based DSSs have gained increasing importance since the 1980s, and a large number of DSSs have been developed to assist extension agents, consultants, growers, and other agricultural actors in crop management. Despite their promise, DSSs have contributed little to practical IP in field because of a series of problems. For example, many simple DSS tools are not widely used because they address only specific problems, whereas agricultural producers must manage a wide range of problems generated by the entire production system.

4.1.1. Structure of a DSS

Many parts of this “super consultant” have already been developed, but these components need to be integrated to produce a holistic system. Pre-cultivation and cultivation decisions are important because they cannot be postponed, are often irreversible, represent a substantial allocation of resources, and have a wide range of consequences that impact the farm business for years to come; all of these possible consequences must be considered by using economic and environmental indicators. These decisions are also difficult because they are complex (they involve many interacting factors and have trade-offs between risk and reward) and/or involve uncertainty (mainly due to the erratic climate).
The super consultant must be delivered through the World Wide Web. A web site eliminates the need for software at the user level and provides a mechanism for a merging of push and pull approaches. Furthermore, it allows the DSS to be updated easily and continuously, so that new knowledge can be provided to farmers even before it is published in research journals. The super consultant should also have greater automation of interpretation than the current DSS. This requires that decision supports are based on static-site profiles and site-specific information; the static-site profile information includes factors about the site that do not change substantially during the growing season (such as previous crop, soil characteristics, cultivar, etc.), while site-specific information may change continuously and must be transmitted directly to the DSS as measurements (such as weather data) or scouting reports (such as the current crop status). Therefore, the DSS for durum wheat was designed to be used in an interactive manner via the Internet.

Lack of clarity about the role of DSSs in decision making, as well as organisational problems related to user support, are among the causes of failure of several DSSs. DSSs should not be designed or used to replace the decision maker but to help the user make choices by providing additional information; the user remains responsible for the choice and the implementation of actions. Both static-site profiles and site-specific information (data) are viewed as flowing from the environment via instrumented sensors or human activities (scouting, analyses, etc.) to a database. The information is manipulated, analyzed, and interpreted though comparison with available expert knowledge as part of the decision process. The information is processed for producing a decision support. As noted earlier, the decision itself is the responsibility of the user, and the DSS is not designed to replace the decision maker but to help in making choices by providing additional information. A decision results in an action to be executed within the crop environment. After the action is carried out, the environment is again monitored to begin a new cycle of information flow. Thus, information flows to and from the environment in an endless loop that begins with sensing and ends with action.

4.1.2. Actors and infrastructures of the DSS

The DSS provider also manages the network of weather stations and of control crops, which provide input data for the DSS. The users of the DSS are the client enterprises (i.e., a single farm, or an organisation that represents many farms, that stipulates an agreement with the provider for accessing the DSS) and the crop manager(s). The crop manager is a person (usually a technician or an advisor) who makes decisions about crop management or suggests the proper actions to the grower. The crop manager directly interacts with the DSS for creating one or more crop units (i.e., a field sown on a uniform piece of land, with the same wheat variety, and cropped in an uniform manner all season long), inputting the crop specific data, and viewing the DSS output. She/he can also interact with the provider for help in interpreting the DSS output.

4.1.3. Monitoring the crop environment

Agro-meteorological stations measure air temperature (°C), relative humidity (%), leaf wetness (yes/no), and rainfall (mm) at a ceratin high above the soil.

A network of “reference crops” is created near the agro-meteorological stations. These crops are periodically monitored during the wheat-growing season by the DSS provider in order to collect field data on the crop status. This information is used by the DSS provider for ongoing evaluation and for improved interpretation of the DSS output.

Both static-site and site-specific information are needed for running the DSS in commercial crops. Static-site information depicts the profile of each crop unit, the soil characteristics (texture, fertility, organic matter content, etc.) and the contribution of organic fertilization. Site-specific information is collected during the wheat-growing season by scouting or field observation. This information represents easily collected data describing plant growth, structure of the weed population, and health of the crop.

4.1.4. Management of data fluxes

Both weather and crop data are automatically stored in specific databases of the DSS. Each weather station is equipped with a TCP-IP gateway that sends the data via GPRS/EDGE every 3 to 15 minutes, depending on the weather variable. When weather data are supplied by external providers, an internet-based procedure makes it possible to download the data automatically at fixed time intervals. The crop data are inputted via the Internet into the specific database by the crop manager through an easy-to-use interface of the DSS.

4.1.5. Data analysis
The weather and crop data are analyzed to produce decision supports for the key aspects of durum wheat cultivation. A step-by-step problem-solving procedure based on important factors relating to the specific process is used for producing decision supports for sowing, nitrogen fertilization, and weed control; decision supports concerning crop growth, pests, and diseases are produced through mathematical models. The problem-solving process consists of a sequence of sections that fit together; these are: problem definition, problem analysis, generation of possible solutions, analysis of the solutions, and selection of the best solution(s). The process initially involves formally defining the problem to be solved. This first step not only involves formalizing the problem but also ensuring that the correct problem has been identified. The next step in the process is to determine the current situation and what components of the situation have created the problem; a set of criteria by which to evaluate any new solutions are also defined. The next step in problem solving is to generate a number of possible solutions. At this stage, the process generates many solutions but does not evaluate them. In the analysing section of the problem-solving process, the various factors associated with each of the potential solutions are investigated; good and bad points and other factors relevant to each solution are noted but solutions are still not evaluated. In the last step, the various influencing factors for each possible solution are examined and decisions are made about which solutions to keep and which to discard. This selection procedure is frequently iterative; a shortlist of potential solutions is prepared first and then further refined by increasing the depth in the analysis of each solution. Usually the process yields one or a few viable solutions.

A plant disease model is a simplification of the relationships between a pathogen, a host plant, and the environment that cause an epidemic to develop over time and/or space. The disease and plant models used in the DSS were developed following a fundamental approach, where ‘fundamental’ is the alternative to ‘empirical’. Empirical models describe behaviour of the system on the basis of observations alone and explain nothing of the underlying processes. Fundamental models (also referred to as explanatory, theoretical, or mechanistic models) explain the same behaviour on the basis of what is known about how the system works in relation to the influencing variables. Fundamental models are also dynamic in that they analyse components of the epidemic and their changes over time due to the external variables influencing them. Dynamic modelling is based on the assumption that the state of the pathosystem in every moment can be quantitatively characterised and that changes in the system can be described with mathematical equations. The models are also weather-driven, because the weather variables are the main inputs of the model.

4.1.6. Decision supports

The DSS produces several kinds of output, at different scales of complexity. The DSS provider can access the results with the highest level of detail because the provider must have a complete understanding of the biological process that underlie the production of the decision support. The provider constantly compares this output with the real situation observed in the reference crops. The crop manager accesses the output concerning the crop unit(s) she/he has created. Two kinds of output are available for each crop unit. The first output is a “dashboard” with images that summarize current weather conditions, crop growth, and disease risk for a selected station; in this dashboard, the other functionalities (fertilisation, weed control, etc.) are displayed by icons (not shown). The manager can also click on the image of any disease and observe the level of risk of the selected station in comparison with that of the other stations.

4.1.7. Technological infrastructure

The technological infrastructure of the DSS comprises the four interrelated components. The “Weather” component manages the collection and storage of the weather data as well as the procedures for the quality control of these data. The “Crop” component manages administration and storage of the data from the crop units. It has two subcomponents: “User Tools” and “Crop DB”. The “Analyze” component contains the procedures for calculating the decision supports (i.e., the main output of the DSS) and for storing them in a database that can be accessed by the users through the “Access” component. The “Analyze” component includes three subcomponents: i) “DSS Calc”, which contains the algorithms that use the inputs for producing the output; ii) “DSS DB”, which stores the results of the calculation procedures (i.e., the output); and iii) “DSS Viewer”, which makes it possible to view the output stored in the “DSS DB” (for those modules that are batch calculated) or to start a new, on-demand calculation of output. The batch-calculated modules are “Crop Growth” and “Diseases”. The “Access” component includes folders and procedures required for managing the users, connecting to the different modules, and accessing the DSS. This component is supplied by the infrastructure of the web-portal http://www.agrishare.com, which makes it possible to manage the different users, including: i) the provider of the DSS, who can access all the information and interact with the whole system; ii) the client enterprise; iii) the crop manager(s); and the crop unit(s) created by each crop manager.
4.2. Decision Support on Experiential knowledge

While the totality of agricultural knowledge is, as indicated above, exceedingly complex and diverse, we will consider a small subset of that knowledge in this chapter. We will focus on knowledge related to the growth and production of agricultural food crops and the role of nutrients, either in deficit or excess in that relationship. Agricultural knowledge is extremely descriptive with many adjectives and nouns, but few of the axioms, postulates, and theorems enjoyed by sciences such as physics and mathematics. Also as suggested above, agricultural knowledge tends to be encyclopedic with relatively few universal, nearly inviolable rules. In addition to exercising relatively few universal rules it is also clearly interdisciplinary, requiring close interaction among disciplines to adequately capture the experience.

Acknowledging the interdisciplinarity is important because the methods and norms of the various disciplines differ and should be respected in order to obtain the best knowledge from each of the disciplines. A personal experience illustrates differences among social and biological scientists, for example. Among biological scientists data almost always refers exclusively to numerical knowledge, weights of maize, metric tons of root crops, dollars per kilogram, kilograms of fertilizers or amendments, duration of crop cycles, while social science data can be notes taken during an intensive interview, during a focus group discussion, or as a result of a recollection. It is important in working with such diverse, interdisciplinary knowledge that disciplines are respected for their methods, techniques, approaches and culture.

Accurate collection and recording of agricultural knowledge, not surprisingly, must reflect the complexity of the knowledge itself. Such collection is difficult and success, not surprisingly, seems to require methods appropriate for the knowledge. Probably some of the best methods from the point of view of completeness are those used by anthropologists. Their holistic perspective requires unusually complete, thorough knowledge collection and recording using the most current methods available.

4.3. Propa (Papaya expert system)

That agricultural knowledge is highly interdisciplinary presents a challenge to the classical concept of an expert in a single discipline. When a grower or producer contacts the University with an issue they sometimes are referred to several experts before determining which expert is the right one for the specific problem. Confusion and failure to succeed in the diagnostic effort may occur. The goal of the Propa decision-aid was to explore this dynamic by attempting to construct a decision-aid that would identify and solve typical problems possibly requiring multiple disciplines. The Propa decision-aid illustrated that it was possible for a group of experts from various disciplines to assess a case of a papaya problem and sort out which expert would be the primary expert to solve the problem. This was achieved through the use of a monitor and blackboard system that evaluated the interaction between the experts and the person with the papaya problem information. Each expert was assigned a dynamic relevancy factor which represented the success of their interaction with the papaya problem information. The disciplines brought together for the problem-solving session included experts in 1) Insect pests, 2) Nutrient management, 3) Disease identification, and 4) General management and good practice. Propa was able to show the user images of the various insects to assist and confirm their identification, which greatly assisted the insect expert’s diagnosis and recommendation process.

4.4. PDSS (phosphorus decision support system)

The goal was to capture the knowledge, including both successful practice and the supporting scientific knowledge associated with the Diagnosis, Prediction, Economic Analysis, and Recommendations associated with managing nutrient phosphorus (P) in tropical food production systems. PDSS builds on the results of the structuring of the knowledge for the soil acidity decision-aid ACID4. As a result of the meta-analysis of the soil acidity decision-making process, we identified four components in the general process of nutrient management: 1) Diagnosis, 2) Prediction, 3) Economic Analysis and 4) Recommendation. These components served the basis for constructing PDSS and will now be discussed in succession.

4.5. NuMaSS (nutrient management support system)

The NuMaSS Project was designed to join the individually developed decision-aid tools ADSS and PDSS, with a new system to be adapted from a nitrogen decision-aid. The NuMaSS Project was developed to integrate and disseminate decision-aid tools that diagnose soil nutrient constraints and select appropriate management practices for location-specific conditions. The strategy was to develop globally applicable, largely
computer-assisted, integrated decision-aids that could both diagnose and prescribe appropriate solutions to soil nutrient constraints. The Project had three objectives 1) Improve the diagnosis and recommendations for soil acidity and nutrient problems. 2) Develop an integrated computerized knowledge-base, and 3) Develop auxiliary tools resulting from the integrated knowledge-base to assist producers to diagnose and solve soil acidity and nutrient constraints.

4.6. Visual ACN

The mission of the Central Institution for Decision Support Systems (DSS) in Crop Protection (German acronym ZEPP) is to develop, collect and examine existing forecasting and simulation models for important agricultural and horticultural pests and diseases and to adapt these models for practical use. More than 40 weather-based forecasting models for pests and diseases have been successfully developed within the last years. The occurrence of diseases/pests and periods of high-intensity attacks can be calculated with high accuracy. The forecast models are based on different concepts. These range from simple temperature sum models to complex population matrices with integrated rate based algorithms to calculate growth, reproduction and distribution of noxious organisms.

DSS are employed for the

- estimation of disease/pest risk
- estimation of the necessity for pesticide treatments
- forecast of the optimal timing for field assessments
- forecast of the optimal timing for pesticide treatments
- recommendation of appropriate pesticides.

Results of DSS are distributed to the farmers via warning services, using different transmission media (bulletins, letters, faxes and telephone answering machines) and via the internet platform www.isip.de (Information System for Integrated Plant Production). The predictions are suitable for integrated as well as organic farming.

The three basic parts that lead to the creation of a comprehensive and modern DSS for forecasting and warning in integrated crop protection will be analysed. Meteorological data are needed as well as assessed field data as input for decision support systems. With these input data the decision support systems calculate an output result, e.g. the date of the first appearance of a pest. In the first part of this chapter a software is presented which was developed to administer the data of weather stations and to make it available for prognosis models. In the second part the results of a study how to increase the accuracy of simulation models by using Geographic Information Systems (GIS) are presented. The influence of elevation and geographical location on temperature and relative humidity are interpolated using GIS methods, whereas precipitation data was obtained from radar measurements. These meteorological data were then used as input for the simulation models. The output of the models is presented as spatial risk maps in which areas of maximum risk of a disease are displayed. It is expected that by using GIS methods the acceptance of model outputs will be increased by the farmers. Finally model validation is one of the essential requirements of the model development process to guarantee that models are accepted and used to support decision making. Validation ensures that the model meets its intended requirements in terms of the employed methods and the obtained results. The ultimate aim of model validation is to make the model useful ensuring that the model addresses the right problem, provides accurate information about the system which is being modelled and makes it actually be used.

5. Questions

1. Define the term of Decision Support System!
2. Give examples for DSS application areas!
3. What are the main components of a DSS?
4. What is the Model Management subsystem for?
5. What types of DSS do you know (Classifications of DSS)?
12. fejezet - Expert Systems and Applications

1. Introduction

In the early 1970’s there was substantial interest in studying decisions by experts that did not use statistical or other mathematical tools, and determining if and how such decisions could be modeled in a computer. In particular, there was an interest with investigating conceptual and symbolic methods appropriate for modeling physician and other expert decision making. Out of this environment, the notion of an expert system evolved. The concept of expert systems is almost magical: simply capture human expertise and put it into a computer program. Rather than worry about a person, a computer program that includes all of the relevant and appropriate knowledge could be developed and shipped around the world.

The term expert system apparently began to be replaced by the term “knowledge-based system” in the mid 1980’s to mid 1990’s. The shift was one that would begin to remove the need for labeling a system with “expert”, and reduce the hype, but still would require that the system be “knowledge-based”. This name shift put less direct pressure on developers to build systems that were equivalent to experts, but also was sign of a commercial and research shift away from expert systems and an evolution to other forms of problem solving approaches.

2. Expert Systems and Human Reasoning

Initially, computer scientists were interested in capturing non-quantitative decision making models in a computer, and they used expert systems to generate those models. What were some basic assumptions about human reasoning that drove expert systems? Perhaps the initial primary assumptions were:

- Experts know more than non experts.
- People use information and knowledge based on their past experience.
- People use heuristics.
- People use specific, a priori rules to solve problems.
- People use focused knowledge to solve problems.

2.1. Experts Know More Than Non-Experts

Expert systems assume that experts know more or at least something different than non-experts in a field. Accordingly, expert systems assume that experts are differentiated from non-experts by what knowledge they have. As a result, capturing that expert knowledge, can potentially change the knowledge of non experts.

2.2. People use Past Experience

People use their past experience (actions, education, etc.) as the basis of how to solve problems. As a result, system developers interested in building systems to solve problems can consult with people to try to capture that past experience, and use it to solve problems where that experience could be used.

2.3. People use Heuristics

Heuristics are so-called “rules of thumb”. Past experience often is captured and summarized in heuristics. Rather than optimize every decision, people satiscfice using heuristics that they have found from past experience drive them toward good, feasible solutions. In order to solve complex problems, expert systems assume that it is possible to capture those heuristics in a computer program and assemble them for re-use.

2.4. People use Rules
Much of everyday and business problem solving seems based on rules. For example, when choosing a wine for
diner, simple rules such as “if the diner includes a red meat then the wine should be red”, help guide dinners to
the choice of a wine. People use rules to solve problems. Rules were seen as a simple and uniform approach to
capturing heuristic information. Heuristics and other knowledge are captured and kept in a rule-based form. If
people use rules then computer programs could use those same rules to solve problems.

2.5. Problem Solving Requires Focused Knowledge

Expert systems researchers note that one view of human intelligence is that it requires knowledge about
particular problems and how to solve those particular problems. Accordingly, one approach to mimicking
human intelligence is to generate systems that solve only particular problems.


Because it was assumed that human problem solvers used rules and knowledge could be captured as rules, rule
bases and their processing were the critical component of expert systems (Figure 12.1). Accordingly, the
structure of a classic expert system was designed to meet those needs. Ultimately, expert systems were
composed of five components: data/database, user interface, user, knowledge base/rule base, and an inference
engine to facilitate analysis of that knowledge base.

12.1. ábra - Figure 12.1: Components of Expert Systems

3.1. Data

The data used by the system could include computer-generated data, data gathered from a database, and data
gathered from the user. For example, computer-generated data might derive from an analysis of financial
statement data as part of a program to analyze the financial position of a company. Additional data might be
selectively gathered straight from an integrated database. Further, the user might be required to generate some
assessment or provide some required data. Typically, initial expert systems required that the user provide key
inputs to the system.

3.2. User Interface

Since the user typically interacted with the system and provided it with data, the user interface was critical.
However, an expert system user interface could take many forms. Typically, there would be a question from the
system, and the user would select one or more answers from a list, as in a multiple choice test. From there, the
system would go to another question, and ultimately provide a recommended solution. In some cases, the user
would need to analyze a picture or a movie clip in order to answer the questions.

3.3. User

In any case, in a classic expert system, the user is a key component to the system, since it is the user who
ultimately provides environmental assessments, generates inputs for the system and as a result, disambiguates
the questions provided by the system to gather data from the user.
Research has found that different user groups, e.g., novice or expert, given the same interrogation by the system, would provide different answers. As a result, generating that user interface and building the system for a particular type of user are critical (Figure 12.2).

12.2. ábra - Figure 12.2: IE-and Expert System Interfaces

3.4. Knowledge Base/Rule Base

The knowledge base typically consisted of a static set of “if…then…” rules that was used to solve the problem. Rules periodically could be added or removed. However, the knowledge needed for solving the particular problem could be summarized, and isolated. In addition, the very knowledge that was used to solve an inquiry also could be used to help explain why a particular decision was made. (Some researchers also include explanation facility as its own component.) Accordingly, gathering, explaining, and verifying and validating that knowledge is the focus of most of the rest of this discussion in this paper.

3.5. Inference Engines

Inference engines facilitate use of the rule-base. Given the necessary information as to the existing conditions provided by the user, inference engines allow processing of a set of rules to arrive at a conclusion by reasoning through the rule base. For example with the system “if a then b”, and “if b then c” would allow us to “reason” that a led to b and then to c. As rule-based systems became the norm, the inference engine saved each developer from doing the same thing, and allowed developers to focus on generation of the knowledge base. Developers became referred to as knowledge engineers. Ultimately, data was gathered, combined and processed with the appropriate knowledge to infer the matching solution.

3.6. Expert System Software

Because the expert system components were distinct, software could be designed to facilitate the ability of developers to focus on problem solution, rather than building the components themselves. As a result, a wide range of so-called “expert system shells” were generated, for example, EMYCIN, ART (Automated Reasoning Tool by Inference Corporation).

4. What is an Expert System?

The term expert system has been broadly applied to a number of systems, apparently for a number of different reasons. At various points in time, the term “expert system” has implied a type of knowledge representation, a system to perform a particular task, the level of performance of the system. The main components of an Expert System can be seen on Figure 12.3.
4.1. Rule-based Knowledge Representation

As noted above, people appeared to use rules to reason to conclusions, and experts were seen as supplying rules that would be used to guide others through task solution. As a result, most of the so-called “expert systems” probably were “rule – based systems”. Researchers had seen that this type of reasoning apparently was often used by people to solve problems. So-called experts seemed to reason this way, so the systems were “expert”.

4.2. Activity/Task of the system

Another rationale for labeling a system an “expert system”, was because the system performed a specific task that human experts did. Experts seem to structured reasoning approaches that could be modeled to help solve various problems, e.g., choosing a wine to go with diner.

4.3. Definitions

Accordingly, over the years there have been a number of definitions of expert systems, including the following:

- A program that uses available information, heuristics, and inference to suggest solutions to problems in a particular discipline (answers.com).

- “The term expert systems refer to computer programs that apply substantial knowledge of specific areas of expertise to the problem solving process.”

- “…the term expert system originally implied a computer-based consultation system using AI techniques to emulate the decision-making behavior of an expert in a specialized, knowledge-intensive field.”.

As a result, we will call a system an expert system when it has the following characteristics:

- A rule-based approach is used to model decision making knowledge, and that those rules may include some kind of factor, to capture uncertainty.

- Interacts with a user from whom it gathers environmental assessments, through an interactive consultation (not always present).

- Designed to help facilitate solution of a particular task, typically narrow in scope.

- Generally performs at the level of an informed analyst.

5. Characteristics of Expert System Applications

Since expert systems related to the ability of a computer program to mimic an expert, expert systems were necessarily about applications, and comparing those human experts and systems. The initial goal of expert systems at some level was often to show that the system could perform at the same level as a person. But as they put these systems in environments with people we began to realize a number of key factors. First, typically, in order for there to be a rule-base to solve the problem, the problem will need to be structurable. Second, systems may support or replace humans. Third, one of the key reasons that a system might replace a human is the amount of available time to solve the problem, not just knowledge.

5.1. Structured vs. Unstructured Tasks
Expert systems and their rule-based approaches rely on being able to structure a problem in a formal manner. Rules provided a unifying and simple formalism that could be used to structure a task. Thus, although the problem may not have had sufficient data to be analyzed statistically, or could not be optimized, there was still information that facilitated structuring the problem and knowledge about the problem in a formal manner.

5.2. Support vs. Replace

Expert systems were often seen as a vehicle to replace human experts. Many systems apparently initially were designed to replace people. However, in many decision making situations, the focus was on providing a decision maker with support. For example an accounting system ExperTAX, the expert system is not designed to replace accountants, but instead enhancing and supporting advice for people.

5.3. Available Time

Another important issue in the support vs. replace question was how much time was available to make the decision. If a problem needed to be solved in real time, then perhaps support was out the question, particularly if there were many decisions to be made. Further, even if the system was to support an expert, perhaps it could provide insights and knowledge so the expert did not need to search for information elsewhere.

6. Applications

Because of its focus on modeling and mimicking expertise, ultimately, the field of expert systems has been application oriented. There have been a large number of applications of expert systems, in a broad number of different areas, including Chemical Applications, Medical diagnosis, Mineral Exploration, Computer Configuration, Financial Applications, Taxation Applications. Applications have played an important role in expert system technology development. As expert system technologies and approaches were applied to help solve real world problems, new theoretical developments were generated.

7. Knowledge Acquisition

Early expert systems research was not so much concerned with knowledge acquisition or any other issues, per se. Instead the concern was mostly about the ability of the system to mimic human experts. However, over time as more systems demonstrated the feasibility of capturing expertise, there was greater attention paid to knowledge acquisition.

In general, expert system expertise was initially solicited in a team environment, where programmers and the expert worked hand-in-hand to generate the system. Faculty from multiple disciplines were often co-authors on research describing the resulting systems. However, as the base of applications broadened, it became apparent that interviews with experts, designed to try and elicit the appropriate knowledge was generally the most frequently used approach. The importance of getting step-by-step detail, and that using some form of “quasi-English if-then rules” to document the findings. However, there have been a number of other creative approaches for gathering knowledge from experts, including the following.

8. Knowledge Representation

There are a number of forms of knowledge representation in artificial intelligence. However, expert systems typically refer to so-called rule-based systems. However, there have been some extensions to deterministic rules to account for uncertainty and ambiguity.

8.1. “If…then…” Rules

“If…then…” rules are the primary type of knowledge used in classic expert systems. As noted above those rules are used to capture heuristic reasoning that experts apparently often employ. However, over time researchers began to develop and integrate alternative forms of knowledge representation, such as frame-based or case-based reasoning, into their systems. Systems that included multiple types of knowledge sometimes were referred to as hybrid systems, or labeled after a particular type of knowledge representation, e.g., case-based.

9. Explanation
Researchers were able to develop techniques, so that given complex rule bases or other structured forms of knowledge representation, systems could analyze the knowledge to find a solution. However, a human user of the system might look at the systems and not understand “why” that particular solution was chosen. As a result, it became important for systems to be able to provide an explanation as to why they chose the solution that they chose.

9.1. Importance of Explanation Facilities

They found that novice and expert users, employed the explanation capabilities differently. In addition, they also found that users were more likely to follow a recommendation if there was an explanation capability. As a result, explanation is an important strand of expert system research, that includes the following.

9.2. Trace Through the Rules

Perhaps the first approach toward generating a system that could provide an explanation for the choice was to generate a trace of the rules. The trace was simply a listing of which rules were executed in generating the solution. Much of the research on explanation leveraged knowledge and context from the specific application area. Although primitive, this approach still provided more insight into why a decision was made, as compared to probability or optimization approaches.

9.3. Model-based Reasoning

In general, explanation is facilitated by the existence of a model that can be used to illustrate why a question is being asked or why a conclusion was drawn. One model-based domain that has gathered a lot of attention is the financial model of a company that depends on a number of accounting relationships. This financial model has been investigated by a number of researchers as a basis of explaining decisions.

9.4. Dialog-Based Systems

An interesting approach to explanation, suggesting that in the long-run that expert systems must be able to participate in dialogs with their users. Quilici suggested that providing a trace was not likely to be enough, but instead the system needed to know when and how to convince a user. This would require that the system understand why its advice was not being accepted.

9.5. Verification and Validation

As noted above, one of factors that makes a system an expert system, is the level of performance of a system. As a result, perhaps more than any other type of system, verification and validation that some system functions at a particular level of expertise is important in establishing the basic nature of the system. Accordingly, an important set of issues is ensuring that the system developed works appropriately and that the knowledge contained in the system is correct. Assuring those conditions is done using verification and validation.

Verification is more concerned with the syntactical issues. Verification refers to building the system right. Verification refers to making sure that the technology has been correctly implemented. Accordingly, verification is concerned that the structural nature of the “if...then...” rules is appropriate. For example, verification is concerned that there are no loops in the rule base (“if a then b” and “if b then a”) or that there are no rules that conflict (e.g., “if a then b”, “if a then c”). Verification also is concerned that any weights on rules have been done correctly. A number of approaches help determine if expert system weights on the rules have been put together appropriately or if there are any anomalies that should be investigated.

Validation is more concerned with the semantic issues. Validation refers to building the right system. Some of the key functions of validation, all consistent with the nature of expert systems, are:

- ascertaining what the system knows, does not know or knows incorrectly;
- ascertaining the level of decision making expertise of the system;
- analyzes the reliability of the system.
10. Expert System Strengths and Limitations

Unfortunately, the mere term “expert” has put much pressure that the system performs at an appropriate level. This label is both a strength and a weakness. This section lists some other of the strengths and limitations of expert systems.

10.1. Strengths

Expert systems have provided the ability to solve real problems using the manipulation of syntactic and semantic information, rather than quantified information, providing a major change in the view as to what computer could do. In particular, if the problem being posed to the system is one for which rule based knowledge is effective, then the system is likely to be able to provide a recommended solution.

Further, expert systems can be integrated with other computer-based capabilities. As a result, they can do substantial “pre-analysis” of the data. For example, in the case of financial systems, financial ratios can be computed and analyzed, saving much time and effort.

10.2. Limitations

However, there are also some limitations associated with expert systems. One of the biggest “complaints” against expert system has been the extent to which they are limited in scope and that the systems do not know their limitations. Classic expert systems have rules that focus only on the problems that it is designed to solve, resulting in their limited scope. Generally expert systems do not know when a problem being posed by the user is outside of scope of the system.

From a practical perspective, expert systems “…require complex and subtle interactions between machines and humans, each teaching and learning from other.” Rather than being static, systems and people need to learn and change to accommodate each other.

In addition an early investigator who noted that people, other than the authors of the rules may have difficulty modifying the rule set. For example, Clancy noted “…the view that expert knowledge can be encoded as a uniform set of if/then associations is found to be wanting”.

Getting and keeping up-to-date knowledge is another potential limitation. For example, in the area of United States taxation, the tax rules change every year. Some rules are new and some rules are no longer valid. Such rule-base changes are not unusual in any setting where technology is involved that must change often more than once a year. For example, imagine developing a system to help someone choose the right mobile phone.

Finally, a primary limitation to expert systems is illustrated by comment from Mike Ditka, a hall of fame American Football player. On a radio interview on Los Angeles Area radio, while talking about evaluating football players, he noted “…the intangibles are more important than the tangibles”. Viewed from the perspective of expert systems, this suggests that although we can capture (tangible) knowledge, that other (intangible) knowledge is out there, but not captured, and in many cases that additional knowledge may be the most important.

11. Extensions to Expert Systems and Emerging Research Issues

The basic model of the expert system presented to this point is one where knowledge is gathered from a single expert and that knowledge is categorized as “if...then...” rules, as a basis for mapping expertise into a computer program. However, there have been some extensions to that basic model, including the following.

11.1. Knowledge from Data

Gathering knowledge from experts ultimately became known as a “bottle neck”. In some cases data was available, so rather than capturing what people said they did, an analysis of the data found what they actually did. Some researchers began to try to get knowledge from data, rather than going through classic interview
processes. Ultimately, the focus on generating knowledge from data ended up creating the notion and field of knowledge discovery.

Neural nets also provided a vehicle to capture knowledge about data. Ultimately, neural nets have been used to create rules that are used in expert systems and expert systems have been built to try to explain rules generated from neural networks.

11.2. Alternative Forms of Knowledge Representation

As researchers studied reasoning and built systems they found that rules apparently were not the only way that people thought, or the ways that the researchers could represent knowledge. For example, one line of reasoning suggested that people used cases or examples on which to base their reasoning. As another example, researchers built frame-based reasoning systems. Frames allow researchers to capture patterns, that allow heuristic matching, e.g., as was done with GRUNDY. As a result, case-based reasoning and other forms of knowledge representation helped push researchers to forms of knowledge representation beyond rules in an attempt to match the way that people use knowledge.

11.3. Alternative Problem Solving Approaches

Not only knowledge representation changed, even other types of problem solving approaches were used. “The term (expert systems) has subsequently been broadened as the field has been popularized, so that an expert system’s roots in artificial intelligence research can no longer be presumed … any decision support system (is) an expert system if it is designed to give expert level problem specific advice…”.

11.4. Expertise

Since expert systems were intent on capturing human expertise in a computer program, this led to a need to better understand expertise and what it meant to be an expert. As a result, since the introduction of expert systems there has been substantial additional research in the concept of expertise, not just how expertise can be mapped into a computer program.

11.5. Uncertainty Representation

Generating expert systems for different domains ended up facilitating the development of a number of approaches for representing uncertainty. However, additional research has focused on moving toward Bayes’ Nets and influence diagrams and moving away from the MYCIN certainty factors and the Prospector likelihood ratios.

11.6. The Internet and Connecting Systems

Generally, the expert system wave came before the Internet. As a result, the focus was on systems for a specific computer, and not networked computers. As a result, there was limited research about networks of expert systems. However, since the advent of the Internet, expert system concepts were extended to knowledge servers and multiple intelligent agents. In addition, technologies such as extensible mark-up language (xml) are now used to capture information containing rules and data, and communicate it around the world (e.g., xptrule.com).

11.7. Ontologies

Further, developers found that as expert systems grew or were connected and integrated with other systems that more formal variable definition was necessary. Large variable sets needed to be controlled and carefully managed, particularly in multilingual environments. As a result, extending those expert system capabilities led to some of the work on ontologies.

11.8. Embedded Intelligence vs. Stand Alone Systems

Increasingly, rather than highly visible stand alone applications, rule-based intelligence was built into other production applications. Because the systems were not stand alone expert systems, users did not even “see” the embedded expertise: People don’t go check on what the expert system has to say – programs now are just more
intelligent. For example, fixing spelling errors and grammar errors in Word, requires a certain amount of intelligence.

11.9. Business Rules

As another form of evolution, there is now interest by businesses in so-called “business rules”. As might be anticipated, business rules assume that businesses use rules in their interaction with other businesses. Rather than wait for people to make decisions, business rules capture those decision making capabilities. Business rules have virtually all of the same concerns as we saw in expert system rules, in terms of knowledge acquisition, knowledge representation, verification and validation, etc.

12. Conclusion

Expert systems have provided an important starting point for understanding and mimicking human expertise. However, they were only a start. Expert systems focused on heuristic decision making and rules, generally as manifested in “if-then” rules, possibly employing weights on the rules to capture uncertainty or ambiguity. Expert system provided the foundations on which many other developments have been made.

13. Questions

1. What is the Artificial Intelligence?
2. What are the potentials of Artificial Intelligence applications?
3. What are the knowledge-based Systems?
4. What are the expert systems, what are the application areas?
5. What are main components of the Experts Systems?
6. Describe the Neural Networks, and explain the differences between Neural Networks and Experts Systems!
7. Give Example for Agricultural Applications!
13. fejezet - eGovernment Services in Agriculture

1. Introduction

Although a number of new concepts and expressions have originated in the field of information society, eGovernment is considered to be one of the first and, because the whole field is rapidly developing, this concept is also constantly changing.

eGovernment has become an indispensable tool in reforming state administration and the work of local government, it is increasing the satisfaction of citizens regarding services, and creating a more flexible, transparent, public administration. The European Union’s relevant guidelines for its Member States stress the importance of eGovernment.

In an ageing Europe, developing competitive industry is extremely important, when we consider stagnating economic development and the consequent need to reduce social expenditure. As well as the policy of creating a citizen centered public administration, promoting our industrial competitiveness in international markets is a central theme of the newest eGovernment thinking. The main aims are to increase effectiveness, implement the necessary structural changes and reduce administrative costs.

The role of government and public administration is becoming more and more important, as the state’s potential as an active service provider contributes to any improvement in economic capacity in international markets. Nowadays, there is a lot of pressure to modernize public administration: on the one hand, public administration itself must become efficient; on the other hand, a more efficient public administration serves to make the economy and society more competitive as well. Creating good governance is possible through a “service provider” public administration that reacts promptly to changes, is flexible and able to meet the demands of users.

However, the successful realization of this may be hindered by the passive resistance of the back-office side of governance service providers, as well as the unsatisfactory quality of the front-office, customer-side service providers, and a lack of interest shown by citizens.

The first web home page of the United States government opened just 10 years ago, and at first it provided only static information. Since then, a process of professionalisation is underway, and with it the age of “amateur” eGovernment services has ended:

• systematically built and accessible fields of knowledge have been developed;

• eGovernment activity has become institutionalised as a profession (international organisations, comprehensive programmes, specialized periodicals, specialist institutions, awards/prizes);

• the specialization of relevant professionals has accelerated;

• a whole battery of training programmes have started across the world;

• the research infrastructure of this area is becoming stronger and stronger;

• governments are starting to act like large companies: on the one hand, prioritising innovation, and on the other hand with regard to the planning, implementation and management of programmes.

2. Definition

According to the definition given by the European Commission, “eGovernment” means using the combination of information technology, organisational changes and new skills in public administration.

13.1. ábra - Figure 13.1: Definitions of eGovernment
The aim is to improve the quality of public services, reinforce the democratic process and support community objectives. According to the Commission’s initiative, eGovernment is:

- open and transparent: public administration capable of comprehending citizens’ expectations, and it is accountable and open towards democratic participation;
- cannot exclude anyone: user-centred public administration must reach everyone with personalised services;
- effective public administration: operates to use taxpayers’ money in the most efficient way saving time and cost.

According to the resolution of the European Union, eGovernment consists of the following three activities:

- the use of infocommunication tools in public administration,
- the reorganisation of work processes and operational units to ensure the modernisation of public administration,
- training of civil servants and government officials as well as customers (citizens) in the use of new tools and technologies.

Public administration may be divided into two main areas, namely: the service-side (back-office) and the customer-side (front-office). The balance between the two sides – the distribution of public goods (the content) and administration (control) – can be achieved with the help of eGovernment by the system. Today, the expression “eGovernment” is used as the collective name for a, complex social and technical system, often including the following components:

- the reform of public administration;
- the technical modernisation of public administration;
- transforming services and the channels through which they are delivered, and making them multi-functional;
- developing an institutional partnership relationship between (local) government, citizens, and their local communities.
In practice, eGovernment means a new culture, a comprehensive and radical transformation in the course of which public administrative organisations make use of all the possibilities of electronics in order to improve the availability, quality and transparency of public services, and try to reduce the costs of public administration.

**13.2. ábra - Figure 13.2: Key Definitions of eGovernment**

This is in direct opposition to those frequent misconceptions concerning eGovernment, namely that the letter “e” only implies “electroisation”, i.e. computers and software. “Electronic government means the comprehensive, smooth reorganisation of processes and endowing them with opportunities made possible by new technologies, whereby administrative and governmental tasks can be performed on the interfaces of agencies, citizens and politics, as well as within and between government agencies”.

**13.3. ábra - Figure 13.3: eGovernment Transformation**
3. Main characteristics of electronic government

Electronic government is made up of two components:

- Renewal of the internal operation of public administration institutions – back-office, i.e. service-provider side.
- Communication between these institutions, the population and the business sector – front-office, i.e. customer-side.

Front-office service means the direct relationship between public administration and its “customers”. This is the actual interface where the “exchange” of information takes place. The typical infrastructure of the customer-side is the Internet (homepage, portal) and the telephone network.

The task of back-office services is, on the one hand, to “serve” the front-office, to receive and process the documents from the customers, to ensure all the necessary conditions for integrated administration and processing (workflow, integrated databases, electronic signature, data protection, data safety etc.), then to return the result or results to the front-office modules. It is also the task of the back-office to support the efficient operation, management and control of public administration institutions and local government institutions.

13.4. ábra - Figure 13.4: Front-end Vs Backend
In the past, the importance of the two components has been unequal. At first, owing to the enthusiasm following the appearance of the Internet, attention was directed towards online presence, and then came the increasingly efficient diffusion of information about public administration. At the present stage of the process, the popularity of web homepages is now widespread.

There are four possible forms of the relationship between back-office and front-office, that is why these indicate four different strategic models:

- The model of “moving towards online services”: this is characterized by a low degree of integration between different processes, as well as by a service based on a single homepage (channel). The main objective is to have the existing services appear online as well. This has the already mentioned advantages (time saving, flexibility).

- The model of “channel-integration”: this is also characterized by a low degree of integration on the service-provider side but customers can access the services through several channels. The emphasis is on systems for channels (on-line, off-line), and on various systems used for increasing effects and attracting attention (ulinks, pop-up windows, etc.).

- The model of “process-integration”: here the level of integration on the service-provider side is high, the number of channels used is limited to one web homepage. Emphasis is on simplifying processes, abolishing parallel systems and creating automation. As a consequence, the government can function more efficiently (faster, more transparent administration), and this has a positive effect on services, as well.

- The model of “service-integration”: this has all the advantages of the former models, but as well as the high-level integration of the back-office, customers are served by as many information service channels as possible. The number of times citizens have to come into contact with the public sphere (often for the same data) is minimised.

The real challenge is the change from one phase to the other, since there is an “evolutionary break” between the different levels.

In the course of the transition from simple presence to the interaction phase, the task is to bridge the so-called “fear-gap”. This means that the transition from giving simple static information to online communication is causing the public administration bodies serious problems, since this poses data security issues and implies serious learning tasks.

There is an “organisational gap” during the transition between interaction and transaction. In this phase, the public administration bodies, hitherto working in relative isolation, have to contact one another more and more often, and solve certain problems together, since the more complex transactions usually involve several areas of public administration.
Finally, the change from transaction to transformation is possible through the “value-transformational gap”, which means thinking according to the citizen-centred model, cooperation between the units of public administration, perfect information-division, the precise definition of the scope of responsibility, and the efficient management of the consequent legal and ethical questions.

4. The levels of maturity of eGovernment

At the moment, we can distinguish between five levels of maturity in the field of eGovernment services:

- Stage 1: information: the customer receives only general information about the process of the case in question and the necessary documents.

- Stage 2: one-way interaction: as well as the above, it is possible to download and fill in electronically the documents (forms) required for dealing with the case, with or without guidance control, but handing in the documents takes place in the traditional way.

- Stage 3: two-way interaction: electronic data can be entered and it is also possible to check/advise on the entered data. Application in person is not necessary to initiate processing of the case, but delivery of public administration documents (e.g. certificates), receiving decisions, decrees etc., and the payment of dues and fees connected to the case all take place in the traditional way.

- Stage 4: transaction: service ensuring the whole transaction (administrative process) online. The citizen receives the form (document) appropriate to the case electronically, and payment of the relevant fees or dues can also be arranged electronically.

- Level 5: personalisation: pro-active, customer-centred service. This fifth level of maturity will be introduced in the European Union from 2007. In the case of regularly used services such as tax and contribution declarations, it is completely unnecessary to submit personal data again and again when public administration already has them: in these cases, public administration is able to supply forms that have already been filled in.

5. Multi-channel access

Electronic public administration cannot reduce the freedom of choice of citizens, since the state cannot exclude anyone from accessing and using public services.

By the word “channel”, we mean the way in which citizens can gain access to services offered by the public sector – personally, through the post, by telephone, fax, on terminals in public places (kiosks), with browsers on personal computers.

13.5. ábra - Figure 13.5: e-Government Service Delivery Infrastructure
The development of public administration portals is primarily adjusted to the expectations of the population and those of enterprises. This, however, is just the tip of the iceberg. Government reorganisation can ensure transparency and the participation of citizens in political decision making (especially at a local level). From this view, the renewal of processes is more important than the electronisation of existing processes. This means people can effectively access services, and exercise their rights and duties. In order to be able to do this, an extremely high degree of integration is necessary between the different services and the various levels of administration.

eGovernment does not mean that in the future, the population will have to spend more time sitting in front of computer monitors, nor does it mean that all transactions performed by the state must be done via the Internet. EGovernment must make the functioning of the public sphere simpler, and reduce, wherever possible, the necessary number of operations performed in the course of a task. One of the beneficiaries of the resources saved in this way would be the citizens themselves.

6. Customer-centred thinking

The government, as opposed to enterprises, cannot choose its customers, and people are in fact more than just customers. They are connected to the state as taxpayers, users of information and in many other capacities: as citizens, they want to be well informed, they want to take part in political processes, and they wish to express their opinions concerning certain issues.

Developing a customer-centred portal means that the implementation must be adapted to the needs of future users regarding content and the nature of application. Good portals are those that apply a mixture of two different approaches: they are “life-situation” and “target-group” oriented. Governments show a preference for the life-situation approach: they adapt the services of public administration to citizens’ needs (by making sure, for example, that the desired information and services can be found with the help of key words). Being target-group-oriented means that services are differentiated according to other aspects besides that of a particular life-situation; for example, the various user groups (citizens, enterprises) or users of differing expertise (experienced or beginner) or with different rights.

13.6. ábra - Figure 13.6: Different citizen, business, and government users rely on commodity data to inform business and policy decisions
7. The advantages of eGovernment

In 2004, the European Public Administration Network (EPAN) assessed the effects of the European eGovernment projects, and on the basis of the survey, it defined seven kinds of advantages that are typically the results of eGovernment development programmes. We should also take into consideration that deeper social and economic purposes may also be served through eGovernment initiatives. Among these purposes might be strengthening of civil society, promoting publicity and transparency, as well as strengthening democracy.

7.1. The supply and quality of information improves.

With the use of ICT, and even more with the digitisation of information, in most cases a higher quality and more widespread information fortune was created. As digital systems become more widespread, it is no longer necessary to input paper-based data in a tiring manner, since the data was already created in this digital form. The management of digital information is a lot easier too, what is more, the different data sequences can be compared and combined. With the interoperability of databases and their division between different offices, information provision improves significantly, and there is no need to repeatedly supply every authority and office with certain data.

7.2. Procedure time decreases.

Digitalizing information brings more advantages than solely quality. Electronic information disclosure is faster; hence data can be made available to citizens and offices faster (and usually in a more up-to-date state). Services can ensure that the users of public administration services receive forms that have already been partly filled in. There are numerous indications that in the future, there might be no need for the work of data controllers in offices. The storage of information in electronic form may also speed up certain decision-making processes.

7.3. Administrative burdens decrease.

As a consequence of these changes, there is an opportunity to reduce unnecessary administration. Instead of applying mechanical records of data, the work of civil servants can produce added value. Furthermore, in the case of the partly filled-in electronic forms (above), the burdens of citizens can also be reduced. Users need only to skim through the form, generated with the existing data, which they can finalize with a click/push of the button. Analyses of European economic policy point out that in proportion to the GDP, the measured administrative burden is - at present - between 6.8% in the Baltic states, Greece and Hungary, and 1.5% in the United Kingdom and Sweden. Naturally, it is no coincidence that this burden is usually lower in countries with higher GDP. The importance of this area is proven by the fact that the 25% reduction of administrative burdens on enterprises, which the European Commission would like to realize by 2012, could actually lead to a 1.4 - 1.8% increase of the GDP level.
7.4. Costs can be reduced.

When listing the advantages, we must not forget one of the most important aspects, i.e. improving cost effectiveness. This cannot be felt so much by the users, since what they experience is the shortening of administrative time. If, however, the user appears as a legal person, then the monetary savings in cash can be significant. The main components of cost saving could be fewer hours of work and a smaller labour force (e.g. as a consequence of not utilizing the above mentioned data controller tasks), and electronic communication could also be cheaper than the traditional method.

7.5. Higher level services

The components are as follows: greater flexibility, a higher level of transparency and the management of unique, individualised cases. In the first case, we are talking about the much vaunted continuous, "7/24/365" availability: background information can be downloaded at any time, forms can be obtained and filled in without having to actually travel to the office. Even financial transactions might be possible. The most important component may be flexibility, ensuring multi-channel (paper, Internet, CD-ROM, call-centre, SMS, WAP, digital TV) access; that is the citizen can maintain contact with the government in a way that suits him/her the best. To demonstrate the improvement in transparency, we should mention the monitoring of cases, opportunities for better interpretation of the relevant legislation, the "monitoring" of information which gives the opportunity to make exact queries concerning the phase of a certain case or the history of an official correspondence. Even in individual cases, the management of tasks that cannot be carried out by standard procedure and perhaps serving customers who require individual treatment may prove to be profitable for the service provider in both time and money. The official may be able to spend the time saved using standardized procedures on dealing with cases that require special attention.

The last two obligations, increasing efficiency and improving the satisfaction of customers, can be synthesised since these occur almost automatically as a consequence of the advantages above. Service providers have a hard time measuring these (especially the last two) although there are some possible solutions here (for example, measuring the degree of utilisation or placing it in some kind of scale/level system).

7.6. E-democracy opens up new opportunities for the public.

Modern democracies have been struggling with unsatisfactory processes for years: citizens have little trust for democratic institutions, participation in elections is low, and membership of political parties is decreasing. Many people say that that one of the greatest advantages of using new, interactive technologies is that it will help them find new ways to stop unsatisfactory processes. Technology is not a panacea, nor is it omnipotent; but with the help of the Internet, new areas of democracy will open up for citizens.

Thanks to the Internet, many users can communicate simultaneously with a lot of other people, on a mutual basis. Interactivity, the possibility of communication without frontiers plays an ever-greater role in European eGovernment activities. Primarily, widespread participation is made possible by the Internet, while at the same time, it enables the (elected) representatives to familiarize themselves with the interests of those they represent, and also, allows them to take their views into consideration in the course of the decision making processes.

On this basis the concept of electronic democracy can be best defined as the use of interactive technologies for strengthening democratic processes, as a result of which people may feel there is greater scope for their views and opinions, and they can be more active participants in democracy. Professor Stephen Coleman of Oxford University, who is researching electronic democracy, interprets the expression as the relationship between citizens and the government, and recommends the following definition: Electronic democracy means making use of the opportunities provided by digital technology with a view to improving the democratic process interacting between the governing power and those governed, and between representatives and those represented.

Thanks to systematic, intensive social dialogue, there is hope that democratic public culture and the willingness to participate will grow stronger, and at the same time, the understanding of democratic, political participation will become wider. Votes in elections will continue to be the main, precisely measurable scale of participation, but the communication and the discussions preceding the elections which belong to the process of everyday politics, will also be seen as an important aspect of participation.
For this reason, “real” electronic government is characterised by the balanced combination which ensures electronic services and the various possibilities of electronic participation. If we are to define “electronic government” as above, then “electronic democracy” can be further divided into “electronic participation” and “electronic voting”.

“Electronic participation” means not only the digitisation of existing processes of planning and decision making, its aim – with the help of information and communication technology, of course – is rather “to develop new possibilities for participation, and to establish itself as part of the new administrative and decision making culture.”

The main elements of a democracy based on the activity and direct participation of citizens are as follows: the “one-to-many” type of communicative regime, direct representation based on interactivity, creating involvement in the process of political decision making, restoring trust, and reducing the democratic deficit manifesting itself in parliamentary institutions.

Information and communication technologies can facilitate the fulfilment of a deliberative-based democracy in several dimensions:

• making the operation of the state apparatus more transparent on an institutional level by ensuring freedom of information,

• changing the system of connections between the citizen and the state into a system of relations based on two-way communication,

• the self-organisation of citizens, by creating networks organised from the grass roots.

Making the state more transparent at an institutional level can be seen as the embodiment of electronic freedom of information. The transparency of the state administration’s operation at institutional level is considered to be of key importance by the supranational directing bodies of the European Union, therefore the memorandum concerning the creation of the electronic government service system was prepared for the eGovernment conference held in Manchester in November 2005. This memorandum acts against the concealment of information and against bureaucracy excluding the public.

In Hungary, for example, the adoption of legislation concerning the freedom of information is a sign indicating the transparency of the state activities, while the presence of a growing number of online communities organised from the bottom up shows that there might be a consensus concerning public affairs; the citizen may change from being a passive receiver into a person whose creative activities are based on individual decisions. However, the majority of programmes aiming at developing e-democracy still originate from the government sector, apart from a few initiatives supporting special situations, the number of civil initiatives is in- significant.

The change from the system of connections between the subordinate citizen and the state to the more equal two-way relationship based on communication is under way.

The realisation of e-democracy also means, on the one hand, that electronic communication with the public authority becomes a civic right, while on the other hand, data of public interest becomes accessible in different formats, and from various sources.

• The aim of the programme called eParticipate4, which was started in March 2005, is to create a network between institutions of the public sphere, which increases the opportunity for citizen participation in democratic processes. The initiative is based on the “webcasting” technology that might allow a transmission, perhaps live from the council-chamber of a local authority, with access for audio and video content as well.

• In the past few years, podcasting has become more and more possible, and the public sphere has realized how important this is. Anyone can upload their oral (or audiovisual) message to the Internet with the help of a computer and a microphone. Governments recognized the importance of this in 2005: the Democratic Party of Singapore was the first to issue a political podcast, then Senator Larry Craig, of the United States of America, made it possible to download podcasts from his homepage, the following day, the White House provided access to President Bush’s speeches in the form of podcasts.

• The introduction of the institution of e-petition can be fitted into the conception of services supporting legislation (e-Parliament). The approach, which aims at the institutional modernisation of parliaments, can be achieved, on the one hand, through better organisation, managerially-based efficiency, openness,
participation and transparency, and on the other hand, by aiming to eliminate an increasingly obvious democratic deficit, by which means we may regain the citizen’s trust. In this spirit, the German Bundestag launched its electronic petition service from September 1 2005, offering every single citizen the possibility to initiate e-petitions, sign petitions written by others or express their opinions concerning any ongoing online discussion.

7.7. Networks organised from the grass roots up are created through the self-organisation of citizens.

The Web 2.0 revolution of today has resulted in the extremely significant advance of content generated by users. Ever growing numbers of citizens record events taking place in public spaces, with their mobile phones or handycams, then submit them on the World Wide Web. With the help of blogs and RSS (Really Simple Syndication) formats (“very simple information-division”), citizens are creating new contents, sharing their opinions, and posting their political views. There is a growing need in society for people to contribute to the development of the public good not as passive consumers, but rather as citizens taking part in the political discourse.

8. Measuring the effects of eGovernment

It is inevitable that in parallel with the development of eGovernment models, the development of research should follow a similar route. The emphasis must soon be centred on the study of the new developments with regard to both research questions and applications. Today, the approach is that the eGovernment services, ministries or states are ranked according to statistical measures, and the world is divided into those who lag behind and those who forge ahead. This may be insufficient, even flawed. We should proceed to examine the causes of new development, showing the possible direction ahead, and examining the social, economic and administrative effects of policy aims. These will require new questions and new research methods; that is, a new approach from researchers, from politicians and decision makers.

We might see a parallel between the development of eGovernment initiatives and the trend of research concerning the phenomena of the information society (including those of eGovernment). On this basis, the research into indicators of development in information society and any related investigations can be categorised into three groups:

- Indicators and research concerning readiness,
- indicators and research monitoring intensive development,
- indicators and research recording the effects of the development of information society.

The rate of development of eGovernment initiatives can also be characterized by a similar S-curve, but we are aware from research results that at the moment, we are finding a radical, paradigmatic change. The motor of change being that customer-friendly, user centred services are developing and in the interest of increasing effective cooperation between the various public administration bodies, the different regions and the different states is getting stronger.

Of course, we can draw a parallel between the two development curves, and a statistical system can be drawn up into which the research concerning eGovernment initiatives can be input. In this system, we can differentiate between the three consecutive subject areas:

The direction of development shows us that research is moving towards the better investigation of the real effects of eGovernment development programmes. The important questions are no longer about how many computers there are in households or offices, but the following, for example:

What effect does all this technical and social change have on the quality of life? How do the people’s satisfaction and their trust in public administration change? To what extent is the efficiency of public administration increasing?

Thus, relevant change to the most important type of research question and the corresponding gathering of statistical data can be summed up in the following figure.
All this presents a three-part system of requirements in those various countries when involved in eGovernment research and into information society in general:

- information supply regarding the country’s state of preparedness through investigations of readiness,
- monitoring the intensity of development in the interest of comparative analyses,
- monitoring the changes that have taken place in society generally and in the economy and public administration, in particular through the widespread use of information and communication technology.

The investigations must examine society and the economy in a sensitive, complex way, revealing then analysing the phenomena and then, by determining the directions of development and implementation, they can assist social and economic policy decision making.

The newest developmental phases of eGovernment demand a move from research that expresses simple quantity to a more scientific, benchmarking type of research that examines the effects of spontaneous development and the economic, social and administrative effects of development programmes. There is a growing need for more differentiated research and analysis, underlining the need that for those countries in the “push” phase, impact studies regarding the population’s reception of eGovernment services are especially important.

9. Key terms

Back-office: A back-office system serves the effective cooperation within or between public administration bodies.

Benchmarking: A research procedure suitable for the qualitative and quantitative comparison of the performance level of an intervention with one that qualifies as the best in a similar field. The method makes the analysis and correction of key processes, as well as the elimination of errors, possible. It also improves performance and the definition of targets. Benchmarking is an important tool for finding “best practice”, which can also lead to higher performance befitting the targets aimed at.

CustomerRelationshipManagement(CRM): This means all the methods, information technology applications and Internet access with the help of which a company is able to maintain relationships with their customer-base, within an organised framework.

Deliberative democracy: In the traditional conception of democracy, citizens play only a passive role as consumers, since they practice the right of democratic control mainly by voting. At the same time, they have almost no influence on determining the public good. In deliberative democracy, however, public debate plays a central role, which is usually centred on different ideas concerning the public good. In public debates, citizens express not only their existing ideas, but also formulate their point of view as a result of constant reflection.

Electronicdemocracy: The use of interactive technologies in the interest of strengthening democratic processes, as a result of which people may feel they have more space in which to express their views and opinions, and they can be more active participants of democracy. In other words, exploiting the possibilities offered by the digital technologies to strengthen the relationship between citizens and the government, in order to improve the democratic processes between the governing power and those governed, between the representatives and those represented. Electronic democracy can be further divided into “electronic participation” and “electronic elections”.

Electronic government: Using the combination of information technology, structural changes and new skills in public administration in order to improve the standard of public services, and make the operation of public administration more simple, more efficient and more economical, and to further strengthen the democratic processes.

EU 20 services: The 20 basic public services defined in the eEurope 2002 ActionPlan (Common List of Basic Public Services - CLBPS). The online sophistication of these services is measured annually in every Member State of the European Union.

Front-office: A customer service and information technology system through which the IT back-office systems of public administration bodies are made accessible to authorised users.
Interactive services: Services that go beyond simply offering information - downloadable forms, search systems, thematic guides that require the active participation of the customer.

Interoperability: The ability of systems to cooperate with one another. Interoperability can be:

- technical, concerning the necessary standards for the cooperation of systems
- semantic, concerning the standardisation of the description of concepts and objects
- political, human, concerning the disposition of resources
- between communities, concerning the distribution and common use of resources
- legal
- international.

10. Questions

1. What is e-Government? Try to give definitions!
2. What types of digital government do you know?
3. What are the key drivers of e-government?
4. Give examples for e-government services!
5. What are the differences between e-Government and e-Governance?
6. What are the benefits of e-government?
14. fejezet - Mobile and Wireless Technologies and Applications

1. Introduction

Mobile communications has enabled us to use many equipment unlinked to the internet without a ‘wired’ LAN (Figure 14.1). Simply put, if the internet gave us the ability to access any web address on a desktop, mobiles have given us the access at any time and from anywhere. This capability, derived from modern telecommunication technology, is crucial in conducting international business operations.

14.1. ábra - Figure 14.1: Mobile Communication Tools

![Diagram of Mobile Communication Tools]

Nevertheless, due to various mobile protocols and networks available in different parts of the world nowadays, for example, analogue, GSM, TDMA or CDMA, it becomes challenging for the airtime providers to expand their services across technological incompatibility (Figure 14.2). The developing Third-Generation (3G) standard is attempting to unify all new-generation mobile devices in a single platform. With the new standard, the mobile gadgets may replace desktop PCs, laptop PCs, credit cards or even wallets in the near future.

14.2. ábra - Figure 14.2: Mobile Technology Landscape

![Diagram of Mobile Technology Landscape]
Mobile communication allows people to stay in touch with each other at anytime, almost anywhere and through handheld devices. Unconstrained by wires, mobile-system users can communicate while travelling as fast as about 60 miles (100 km) per hour. The mobile phone converts the speaker’s voice into radio waves that travel through the air until they reach a receiver at a nearby base station. The base station then sends the call through the telephone network to the intended recipient.

The initial impetus for developing and marketing mobile telecommunications systems was to offer consumers mobility. At first, many consumers were not enticed by this capability due to its higher cost compared with fixed lines. However, that difference is declining as companies create national or regional networks and alliances that offer pricing plans without roaming fees (charges for calls outside the carrier’s service area). Unlike most countries in the world, in the USA, mobile phone users incur charges, whether the call is incoming or outgoing, thus bearing higher total cost.

The full-feature capabilities of digital phones along with declining service charges have reduced the importance of pagers in the wireless industry. The introduction of two-way paging (which enables users to receive, store and play digitised voice messages) met with a disappointing response. The cellular phone is far more versatile in comparison.

Short message service (SMS), based on GSM technology, is one of the fastest-growing services in mobile communications today on a global basis (Figure 14.3).

**14.3. ábra - Figure 14.3: Cellular Network**
The terms, mobile and wireless, used in the Telecommunications industry, have their own specific meanings, but they share some common characteristics. For example, mobile vs. stationary indicates the ability to access while the device is moving. Similarly, wireless vs. wired means ability to access while the device is not physically connected by a wired line. Even though mobile/wireless technologies can transmit voice and data by means of radio waves, infrared rays, microwaves and electromagnetic waves, this paper mainly discusses merely applications utilising radio waves as the medium due to its popularity.

2. Mobile applications

A number of widely used mobile applications are briefly described in this section. These include: mobile phones, mobile satellites, handheld devices, wireless computing and mobile commerce (m-commerce).

2.1. Mobile phones

This would be counted as the most obvious example of mobile applications based on number of users, as many as 1.3 billion worldwide in early 2003. Table 2 shows a statistics snapshot of the mobile industry as of February 2003 according to Cellular Online.

To complement the cellular phone and wireless computing networks, mobile satellites offer a combination of all-digital transparent voice, data, fax and paging services to and from handheld telephone devices. The systems share an air interface standard named Geostationary Mobile Satellite Standard (GMMS) that is similar to GSM. This means that the Satphone customers will be able to use mobile phones that are compatible with satellite systems in any country where GMSS is offered; in effect, creating roaming capabilities that normal land-based mobile phone users need to pay extra for when the handsets are used in areas outside the network coverage.

Nowadays, there are over 1000 satellites orbiting the globe. The number will climb to over 1500 satellites in 2008. They can be positioned in orbits with different heights and shapes (circular or elliptical). Based on the orbital radius, all satellites fall into one of the following three categories:

Another method to classify existing satellite systems is by their functions.

- Voice communications. These universal satellites provide telecommunication services consisting of not only voice, but also data, fax, and paging. The providers include Iridium, GlobalStar, and ICO.
- Satellite radio. There are two major satellite radios providing music on demand aiming to niche markets, i.e. coast-to-coast drivers who like to stick to their favourite channels wherever they go. Thus, XM Radio allied with
General Motors (GM) to install receivers in its cars, while Sirius Radio partnered with Ford Motor and BMW to put Sirius radios in their cars. In the portable electronics segment, Sony and Sharp are also interested in embracing XM Satellite Radio in their Walkmans.

• Broadband networking. In 2005, Teledesic will offer broadband data services through the current 288 (rather than 840 as originally designed) satellites orbiting about 500 miles away from earth. The in-operational rival, SkyBridge – a satellite-based broadband access system – has started offering internet access and videoconference services since 2000.

• Data messaging. Orbcornm provides narrow band two-way digital messaging, data communications and global positioning services. http://www.orbcornm.com/about.htm

• Geodesy & Navigation. Global Positioning System (GPS) is a space-based triangulation system utilising satellites and terrestrial computers to measure positions anywhere on earth. It was first and foremost developed by the US Department of Defense for navigation purposes. Now, the system is more utilised in the public sector for positioning persons or objects that carry transmitters. Glonass, a GPS in Soviet version, is mainly utilised in military and aviation units in the former USSR.

• Remote sensing. The remote sensing market has traditionally been the domain of single specialist dedicated satellites for sensing global climates and natural events. Comprising of 12 satellites, FUEGO is devoted to wild fire detection and named ‘Forest Fire Earth Watch’. RapidEye, a satellite-based geo-information service, provides global climate information, i.e. hail, storm, frost, drought, etc., as well as agricultural produce prediction services. Cosmo-Skymed provides remote sensing services, such as disaster monitoring, urban monitoring, law infringement and environmental and agricultural monitoring, only to customers living around the Mediterranean Sea.

In general, it seems somewhat gloomy for this industry since all terrestrial networks like phone lines and fibre optics as well as land-based wireless infrastructure have spread into almost all urbanised parts of the world leaving only few abandoned places where there is low density of population to utilise the bandwidth. How could the satellite communications survive in business when the total cost of ownership for end-users is relatively high? No surprise, they cannot!

There have been a number of satellite operators like the examples given above who eventually went bankrupt due to lower incomes than expectations. The reasons for this hardship can be described by the following:

• the satellite industry has a long wait between design and profitability, i.e. it can make money only after 10 years in orbit and needs to build the entire network before signing up its first customer,

• manufacturers must lock down technology more than 3 years before launches the industry bets on a market up to 15 years.

• manufacturers must lock down technology more than 3 years before launches the industry bets on a market up to 15 years.

These are the major negative factors most satellite communication providers are encountering nowadays. Some analysts still see good opportunities for this industry, for example, Futron, a technology management consulting firm based in Maryland, predicts that the satellite transponder business will grow more than 75% within 10 years from now. This would be a good opportunity for satellite operators to turn to focus on video/audio and broadband markets rather than sticking to only the voice market, which is forecasted to stay flat for the coming decade.

2.2. Handheld devices

These devices have come a long way in the past few years, growing from little more than electronic organisers into useful business tools like pocket PCs or PC tablets. The devices can be categorised based on their applications and complexity as follows:

• Personal Digital Assistant (PDA). A term for any small mobile handheld device that provides computing and information storage and retrieval capabilities for personal or business use, often for keeping schedule calendars and address book information handy. The hardware controls are supplied by either Palm OS or Microsoft Pocket PC.
• Smart Phone. A wireless phone with text and internet capabilities. It can handle wireless phone calls, hold addresses and take voice mail and can also access information on the internet and send and receive e-mail and fax transmissions. It can be viewed as a combination of a mobile phone and a PDA in a single gear.

• PC Tablet. A new generation of slate-style portable computers from Microsoft and its partners promise to combine the flexibility of paper notepads with the best attributes of powerful notebook PCs. Microsoft has added a program called Microsoft Journal, which is intended to be a note-taking replacement for the pad of paper you would typically take to a meeting. Everything you write on the pad is stored as graphics – called digital ink – unless you highlight an area and ask the machine to recognise what you wrote. PC makers working on Windows XP Tablet PCs include Acer, Compaq, Fujitsu, Tatung and Toshiba.

• Pager. A small telecommunications device that receives (and, in some cases, transmits) alert signals and/or short messages. This type of device is convenient for people expecting telephone calls, but who are not near a telephone set to make or return calls immediately.

2.3. Wireless computing

Wireless computing or Wireless Local Area Network (WLAN) is the technology that enables a user to receive information such as e-mails and files directly from the internet or any networks to one’s laptop, without the sender’s knowledge of the serving network IP address, which may be a wireless LAN (Figure 14.4).

14.4. ábra - Figure 14.4: WiFi

The technology allows for the rerouting of information to the served network for wireless computing just as it does for mobile data services based on 2.5G and 3G technologies. For this application, there are two competing technologies in this field; Bluetooth and 802.11b or Wi-Fi:

• Bluetooth. An open specification for seamless wireless short-range (less than 30 feet) communications of data and voice between both mobile and stationary devices. For instance, it specifies how mobile phones, computers and PDAs interconnect with each other, with computers, and with office or home phones. The first
generation of Bluetooth permits exchange of data up to a rate of 1 Mbps, even in areas with a large amount of electromagnetic disturbance.

- Wi-Fi (802.11b). Provides for wireless Ethernet transmission primarily between laptops or PDAs and local access nodes that attach to a standard corporate LAN. Today’s 802.11b products transmit in the unlicensed spectrum at 2.5 GHz, and are capable of speeds of up to 11 Mbps. This standard, despite higher speed, is not suitable for mobile applications or moving devices.

2.4. M-commerce

M-Commerce is the use of radio-based wireless devices such as cell phones and PDAs to conduct business-to-business and business-to-consumer transactions over wired, web-based e-commerce systems (The wireless mobil environment is show non Figure 14.5). This application should be seen as a complement of existing e-commerce, which focuses on different groups of customers. Nowadays, we might see examples of m-commerce in forms of m-banking, m-payment and latest, e-vending:

- M-banking. Piloted by NTT DoCoMo, now Japanese mobile phone users can make payments and withdraw cash via mobile phone handset. The technology allows people to withdraw and deposit money at cashpoints in convenience stores and supermarkets using mobile phones instead of cash cards.

- M-payment. The m-Payment service allows users to initiate a payment by sending a text message with their password, the amount they want to send and the recipient’s mobile phone number. The users can also include a short message. As long as the person is on the GSM network, the service will transfer the funds and the recipient is notified via SMS. The system providers anticipated launching the technology with ticket reservation and restaurant businesses.

- E-vending. Cellenium, a mobile commerce technology provider, and Coca-Cola Beverages (CCB) AG in Switzerland are launching a text of wireless intelligent vending that will communicate real-time data to allow remote monitoring of machine functions and inventory. The vending machines will also be equipped to accept m-payments from consumers using their cell phones to purchase drinks. Cellenium will provide both the telemetry and the m-payment technologies combined in one solution.

In the competition landscape of mobile phone market, we can group the battlefields into three areas – protocols or standards; airtime carriers; and handset manufacturers – as described in this section.

14.5. ábra - Figure 14.5: The Wireless Mobile Environment.
generation voice and data devices based on the GSM technology. In early 2000, third-generation high-speed multimedia devices were introduced based on IMT 2000 Project standards. The progression is shown in Figure 14.6.

14.6. ábra - Figure 14.6: Network Overview

- First generation: Analogue. Looking at the above timeline, the protocol competition can be routed back to the first-time hand phone, which was introduced in the market with analog cellular networks in 1978. Analogue cellular operates in the 800 MHz frequency range and is available across 95% of the USA. Analog cellular service sends a voice through the air using continuous radio waves. As the voice signals travel through the air, they weaken with distance. Equipment in the cellular network returns the signal to its original strength, or amplifies it. This technology is the predominant system in use today and is also known as Advanced Mobile Phone System (AMPS).

- Second Generation: Digital. Digital cellular shares the 800 MHz frequency band with analog and is usually available wherever analog service is offered. In digital transmissions, a conversation is converted into the ones and zeros of computer code. Unlike analog transmissions that are sent out as a continuously varying electrical signal in the shape of a wave, digital transmissions are a combination of on-and-off pulses of electricity. Several incompatible air interfaces are used to implement digital cellular networks, including Code Division Multiple Access (CDMA) and Time Division Multiple Access (TDMA).

- Code Division Multiple Access (CDMA). This is a spread spectrum approach to digital transmission. With CDMA, each conversation is digitised and then tagged with a code. The mobile phone is then instructed to decipher only a particular code to pluck the right conversation off the air. The process can be compared in some ways to an English-speaking person picking out in a crowded room of French speakers the only other person who is speaking English.

- Time Division Multiple Access (TDMA). This is a digital–air interface technology designed to increase the channel capacity by chopping the signal into pieces and assigning each one to a different time slot, each lasting a fraction of a second. Using TDMA, a single channel can be used to handle simultaneous phone calls.

- Global System for Mobile Communications (GSM). GSM is a type of TDMA digital wireless network that has encryption features. GSM is being rapidly deployed worldwide and is the standard in Europe and Asia at 900 MHz. In the USA, carriers are deploying GSM at 1900 MHz, making GSM phones sold in the US incompatible with European GSM phones, and vice versa. It has the capability to transmit data at a rate of 9.6 kps.

- Third Generation: 3G. Third Generation (3G) is considered a new standard that promises to offer increased capacity and high-speed data applications of up to 2 MB. The most important feature of 3G is the fact that it is designed to allow global roaming. Third Generation should be considered the next generation of wireless
technology beyond personal communications services. The World Administrative Radio Conference (WRC) assigned 230 MHz of spectrum at 2 GHz for multimedia.

2.5. 3G networks.

These networks must be able to transmit wireless data at 144 Kbps at mobile user speeds, 384 Kbps at pedestrian user speeds and 2 Mbps in fixed locations. The International Telecommunication Union (ITU) seeks to coordinate 3G standards through its International Mobile Telecommunications-2000 (IMT-2000) project.

3. Trends and forecasts

The following is a collection of excerpt predictions by experts about the future of mobile/wireless technology and markets based on a publication in the ComputerWorld magazine. The heterogeneity can be seen on Figure 14.7.

14.7. ábra - Figure 14.7: Heterogeneity in Network

With the notable exception of the i-mode service in Japan, the most significant mobile/wireless applications have been in the enterprise market rather than the consumer market, and this situation is likely to persist through 2004. The existing enterprise applications have largely been used by ‘field-force’ employees (sales people, service technicians, delivery people, etc.). During the next two years, these will be supplemented by applications used by a wider variety of professionals and by more horizontal applications. Tablet PCs will make mobile access to conventional desktop applications more attractive and could lead to mobile computing being used as much in the office as it is out in the field – Eric M. Berg, technology forecaster, PricewaterhouseCoopers, Menlo Park, California.

By 2020, the use of mobile computing in healthcare will extend average life spans by 20–25 years. Implanted wireless devices will continuously monitor our health, enabling the medical profession to treat most diseases in their absolute infancy. Mobile computing will also be used to monitor our diet and its effects on our health, control unhealthy habits such as smoking and alcohol consumption, and enable us to maximise the effects of exercise. Diseases such as diabetes will be virtually controlled through wireless monitoring and corrective-action devices, which will automatically adjust insulin levels without the patient even knowing – Phil Asmundson, deputy managing director of the Technology, Media & Telecommunications Group, Deloitte & Touche LLP, Stamford, Connecticut.
By 2004, more than 1 million remote and mobile devices will be integrated with enterprise applications. Early adopters will include industrial, oil and gas, manufacturing, and utilities. Typical applications will include homeland defence sensors, monitoring flow and pressure of petroleum production, meter readings, and field communications – Bob Ross, WebSphere integration program director, IBM Software Group, Somers, New York.

Mobile devices such as cell phones and PDAs will merge and become indistinguishable. The device itself may even take the place of all credit cards and physical money, and become an automatic transmitter for recorded personal preferences such as room temperature, favourite TV programs and food preferences. Wherever we go, this information will be with us; for example, when you check into a hotel room, your device will automatically set the temperature, TV and dinner menu choices – Brian Terr, director of advanced products, Edmunds.com Inc., Santa Monica, California.

By the end of next year, there will be more than 50,000 publicly accessible ‘hot spots’ around the world for Wi-Fi communications. The vast majority will be created by major wireline carriers around the world, but some will be created by packet wireless carriers and cheap ‘Wi-Fi-in-a-box’ products. Virtual network aggregators will be the glue that binds together all these Wi-Fi ‘islands’. By 2004, different network variants will begin to merge into a seamless, ‘wireless broadband’ global network for roaming purposes. End users would not care what acronym or standard is used; they will just want ‘wireless broadband’ – John Rasmus, vice president, GRIC Communications Inc., Milpitas, California.

By 2007, PDAs and cell phones will have merged into single devices. They will have 802.11 (whatever flavour), Bluetooth, 3G and, possibly, direct satellite capability. They will be voice-controlled and use a heads-up holographic display. Laptops will become unnecessary for most folks – Doug Jackson, director of technology customer services, University of Texas at Dallas.

The really interesting platform for mobile applications is the automobile. It has a big battery and the ability to generate electricity. It has space for all kinds of devices. People spend a lot of time in them. Look for in-car telematics to include GPS, data storage, docking for multiple types of handheld devices, hard-copy output and so on. All of this already exists in law enforcement – and the new bus-based, 48-volt auto system standards will accelerate the vehicle telematics explosion – John Parkinson, chief technologist, Cap Gemini Ernst & Young US LLC, Rosemont, Illinois.

Within the next five years, all front-end user interfaces for computing will be wireless – Sumit Deshpande and Don LeClair, technology strategists, Computer Associates International Inc., Islandia, New York.

In five years, we will see a dramatic build-out of wireless LAN hot spots – there will be as many hot spots as there are ATMs, a lot of them actually colocated with the ATM so that a banking customer can retrieve cash as well as their e-mail – Pontus Bergdahl, president and CEO, Columbitech AB, Stockholm.

Instant messaging in the wired world has been the fastest-growing communication channel in history. Instant messaging in the wireless world is going to even outpace the wired adoption. Evidence is strong: Short Message System is already popular on wireless devices in Europe and Asia. Online ‘presence detection’ – the ability to know when someone you want to communicate with is immediately available – is one of the unique compelling features of instant messaging. This will be the ‘killer app’ for wireless devices – Glen Vondrick, president and CEO, FaceTime Communications Inc., Foster City, California.

Evidently, the mobile communications industry has grown very rapidly in the past decade. There are a number of major players involved in mobile communications. Consumers around the globe have a mixed response to this industry. One of the challenges faced by this industry is that not all consumers value the wireless internet enough to use it or care about it. On the other hand, consumers value mobile phones as sales are reported to be going up every year. The goal to attain globalisation of the mobile communications industry would be achievable if strong long-term partnerships are developed that give equal benefits to all partners. Each partner will have to contribute a needed service or technology, or a large number of consumers. There is no doubt that mobile communications will be a key influence in the world economy in this century (Figure 14.8 sows the evaluation).

14.8. ábra - Figure 14.8: Evaluation to IMS
14.9. ábra - Figure 14.9: Application in Vineyard

4. Questions

1. What is the mobile service?

2. What are the main types of wireless telecommunications networks?
3. What is the mobile computing?

4. What is pervasive computing?

5. What is the sensor network?

6. Give examples for agricultural applications of mobile and wireless technologies!
15. fejezet - ICT Systems in Farm Production Management

1. Introduction

Information technology (IT) enables farmers to access real-time market information and buy and sell through e-commerce sites; manage their cropland at ever smaller scales (to meet both economic and environmental objectives) through precision agriculture; and use modern accounting, recordkeeping, and tax management through computer and Internet resources.

15.1. ábra - Figure 15.1: Agriculture and Food Sectors

Telecommunication infrastructure in rural areas is crucial if farmers and rural residents are to adopt and utilize IT. Many government agencies, including those servicing farmers, are offering clients the ability to receive information and program benefits via the Internet.

1.1. Information Technologies for Crop Production

Recent advances in the computer, aerospace, and communications industries allow farmers to monitor and manage soils and crops on small areas of individual fields. Precision agriculture or site-specific crop management are the terms often applied to the suite of information technologies used for sensing subfield spatial and temporal variability and customizing applications across the field. A number of spatially oriented information technologies are commercially available for most crops to help with fertilizer, pesticide, seed, irrigation, and tillage decisions. Rather than treat fields uniformly, producers can use these technologies to manage soil, pest, landscape, or microclimate variability by adjusting input use within a field to enhance returns and to reduce potential environmental risks. Such technologies include yield monitors; the Global Positioning System (GPS); Geographic Information Systems (GIS); guidance systems; satellite, aerial, and on-the-go sensors; and variable-rate applicators producers have been cautious about using this technology for changing production practices.

Remote sensing, variable-rate applicators, and guidance systems are among the most recent, as well as most rapidly evolving, precision agriculture technologies. Geo-referenced soil data, such as pH or nitrate levels and soil type, can also help producers intensely manage their crops. Recent ARMS data indicate that the adoption of these technologies, like yield monitors and mapping, differs by crop. While remote sensing can detect variation in vegetative reflection, the cause of that variation may still require confirmation on the ground. Also, cost, timeliness, and image resolution issues may be inhibiting the spread of this technology.

1.2. Factors Influencing Adoption
A number of factors—such as profitability, farm and farm operator characteristics, university research and extension activities, and government agency use of IT—will likely affect adoption trends in precision agriculture (PA). Most studies of PA technologies have shown positive economic benefits from the adoption.

2. Precision farming

Precision farming as a concept was defined in the previous section, with ICT and information highlighted as key components. These facets are further defined and explored in this section with respect to their role in enhanced farmer decision-making. The collection of data is highlighted as only a step on the path to knowledge creation, and it is suggested that hardware and software devices have limitations in the application of knowledge and wisdom to decision-making problems. Precision farming is presented as being greater than the devices which collect and manage data with farmers, as end users of the data, being integral to achieving farm management system change.

15.2. ábra - Figure 15.2: Levels of Precision Farming

2.1. Perfecting the Farm through Precision Agriculture

Site-specific information that allows producers to make management decisions about discrete areas of the field is called precision farming or precision agriculture. Determining soil and crop conditions to improve whole-farm efficiency—while minimizing impacts on wildlife and the environment—is the crux of precision farming. It has been used successfully in many developed countries and has the potential to change agriculture dramatically in this century. A variety of tools can be used in precision agriculture. GPS, satellites, sensors, and aerial images can help to assess variation in a given field. Farmers can match input applications and agronomic practices with information received from these ICTs. Precision agriculture has been applied to many types of agricultural produce (hay, pasture, fruit, and cereals, for example) and to fisheries under many different climatic conditions. Many of these efforts have been limited to largescale farming because of the significant investment required, but applications under smallholders’ conditions are gaining visibility. Remote sensors, sonar-based technology, and other ICTs can also improve aquaculture and livestock production. Essentially precision farming provides a framework of information for farmers to make management and production decisions. It can answer questions pertaining to land preparation (including tillage depth and type, residue management and organic matter, and reductions in soil compaction); seed (planting date and rotation, density and planting depth, cultivar selection); fertilizer (nitrogen, phosphorous, potassium, and other nutrients, as well as pH additives, application methods, and seasonal conditions); harvest (dates, moisture content, and crop quality); and animals and fisheries (pasture management, animal tracking, and school identification). Precision Farming through Wireless Sensor Networks Consistent advances in microsensing, smaller devices, and wireless communication have resulted in new comprehensive technologies that offer even more consistent and reliable systems for smallholders and policy makers alike. Wireless sensor networks (WSNs), which combine many kinds of sensory data in one location, are some of the most innovative technologies available for farming and agricultural planning. With the right...
components, these networks can form knowledge management systems, research databases, and response systems that can guide local communities and governments in agricultural development. A WSN is a group of small sensing devices, or nodes, that capture data in a given location. These nodes then send the raw data to a base station in the network, which transmits the data to a central computer that performs analysis and extracts meaningful information. The base station acts as a door to the Internet (typically a local area network), providing operators with remote access to the WSN’s data. Because the networks can have multiple sensory devices, the data can contain information on soil, climate, chemicals, and other relevant subjects. The wide application of WSNs allows them to be used not only in managing agriculture but in testing water quality, managing disasters, detecting volcanic activity, and conducting environmental evaluations. These networks have several key features. First, WSNs have both active and passive sensors. Active sensors release a signal to detect a physical phenomenon like seismic activity and radar. Passive sensors, which transform a physical phenomenon into electrical energy, can detect a vast array of phenomena, including temperature, humidity, light, oxygen, and chemicals. Once sensors (for example, temperature and soil moisture) are selected, node locations are needed. Node density in developing countries should be scarce to better guarantee network connectivity for each node, reduce maintenance, and improve the network’s reliability (though it will limit field-mapping techniques). In addition, because low-income countries often experience poor network and telecommunications connectivity, nodes will often require a “buffer,” where data can be rerouted or stored in another node if connection to the base station fails. If an active node fails to transmit data to the base, the network will “wake up” the closest neighboring buffer node, providing a “multihop transmission. The design and implementation of WSNs requires a number of important features. The nodes should monitor the field(s) continuously and for a significant period—it is best if maintenance is not required for at least one cropping season (or 4–6 months). The nodes should cover a wide area, be small to prevent animal and human interference (like stealing), and tolerate harsh environmental conditions like monsoons and extreme heat. Self-organization is also important: The network should automatically detect removed or newly arrived nodes and adapt the messaging route. WSNs offer extensive benefits to farmers producing plants and animals. Agriculturalists can detect problems at an early stage and use more precise applications of fertilizer, water, and pesticide. Figure 15.3 shows Wireless Sensor Network (WSN), Distributed Collection Architecture

15.3. ábra - Figure 15.3: Wireless Sensor Network

![Wireless Sensor Network](image-url)
Precision Farming through Satellite Technologies

Precision farming through satellite technology utilizes three technologies: GPS (which can position a tractor within a few feet in the field), GIS (which can capture, manage, and analyze spatial data relating to crop productivity and field inputs), and variable rate technology (which provides site-specific, “on-the-fly” estimates of field inputs for site-specific application). The three ICTs combined provide information that allows producers to apply inputs, such as fertilizer and insecticide, precisely where they are needed.

15.4. ábra - Figure 15.4: Precision Farming through Satellite

Agricultural information is typically captured spatially, making it more convenient to handle on a regional scale. GIS technology is promising because it allows for a more specific focus. Variable rate technology has helped to identify weed infestations and water stress in areas where crop pest levels are high, which improves targeting of chemical applications and reduces waste associated with conventional blanket spraying. In addition to the potential productivity gains and cost savings, precision farming through satellite technology enables governments to study how agricultural practices affect the ecosystem and develop better regulations. Once data are collected through GIS, scientists can interpret the images and analyze the soil and crop conditions to achieve better results. Although satellite imagery cannot detect soil quality directly like sensors can, it can record soil properties like light reflections and color. As crops start growing, precise pictures of the crops are captured more efficiently. The condition of the fully grown plants can then provide a clearer picture of the quality of the crops and what they require for successful harvest. Based on soil and crop conditions, farmers can estimate the precise amounts of seed, pesticide, and fertilizer they need, organize the distribution of inputs, plan which crops to plant in which areas, and make new investments. Knowing the size and shape of fields can also help rural communities plan for future developments and investments like mechanization. Small, fragmented, or awkwardly shaped fields are difficult to work with a tractor or even animals. Above a certain minimum field size, it becomes cost-effective to use a tractor. Precision farming provided through satellite imagery can determine this threshold before a community invests in new equipment. If an area is suitable for mechanization, the benefits can be extensive.

15.5. ábra - Figure 15.5: Monitoring System for Farm Operating
2.2. Precision dairying

Precision agriculture, or precision farming, has been recognised as a specific farm management approach for almost two decades. Current and future uses of precision technologies in dairy farming are outlined and related to the information requirements of dairy farmer decision making. The pivotal role of information and communications technology (ICT) is acknowledged with a theoretical discussion of what information and technology involve as individual concepts.

The farm management concept of ‘precision agriculture’ (PA) entails managing land according to its specific characteristics with the core goals of increased production, reduced input costs, and reduced environmental impacts. These goals are achieved through utilisation of information technology to measure and manage the inherent variability within land-based production systems. Precision agriculture is not a new concept. Underlying precision agriculture systems is the assumption that ‘more and better information can reduce uncertainty in decision making and unmeasured variability in agronomic conditions.

Originally focussed on cropping practice, the precision agriculture philosophy has recently extended to include livestock farming, of which precision dairy farming (PDF) is a subset. Whereas in arable systems this variability is very much centred on the soil and plant resources, in dairy farming an important source of production variability is also the individual cow. Technologies are considered to be relevant to PDF including electronic animal identification, in-line milk sensors, decision support software, hand held computers, robotic milking, spatial land resource mapping, geographic information systems, variable rate technology for fertiliser and irrigation application, and spatially-based pasture sensors.

Definitions of the precision agriculture concept vary in published literature however several common themes are present. The underlying mechanism is information technology, both for data collection and interpretation; the phenomenon being measured is variability; the issue of scale is important; and the outcome is improved management control.

The smallest production unit in the dairy is the individual cow. Precision agriculture aims to manage the basic production unit in order to exploit its maximal production capacity. Nevertheless, under conventional conditions, feeding and milking management is usually executed for the whole herd or for a group of cows. The definition can be use of information and communication technologies for improved control of fine-scale animal and physical resource variability to optimize economic, social, and environmental farm performance.

Becoming a precision dairy farmer does not represent a conscious mindset shift. Rather it occurs as a consequence of a drive to improve the management approach, as discussed later in this chapter. Precision dairy farming is currently a classification used by researchers rather than by the farmers themselves and its subjective delineation makes it difficult to determine exactly when someone becomes a ‘precision dairy farmer’. Therefore the above definition is useful for providing an initial point of reference for delineating precision farmers and to provide a boundary for this research study. Under this definition an example of precision dairying could involve
using daily measurement of individual cow milk yields to set supplementary feeding rates. Whereas setting feeding rates based on monthly milk yields (from herd test data) would fall outside the definition of precision dairying due the coarse temporal scale.

Precision dairy farming is applicable to both animal and feed production subsystems. In Australia, individual cow management is typically the first step dairy farmers take towards precision farming due to the presence of established commercial products, for example milk meters and activity sensors. Also, Australian farmers are experienced with the individual animal management mindset from their use of herd testing data for breeding and culling decision-making, and therefore the shift to precision farming represents more evolution than revolution. An area where precision technologies are currently less advanced is in the feed production subsystem involving pasture production and utilization, and nutrient management.

2.3. Precision Livestock Farming

Efficient information management is very much part of profitable livestock production. The main purpose of Precision Livestock Farming (PLF) is to improve the efficiency of production, while increasing animal and human welfare, via applying advanced information and communication technologies (ICT), targeted resource use and precise control of the production process. Through the adoption of electronic data collection, processing and application, precision farming has the potential to improve production efficiency and reduce costs, as well as increase animal and human welfare. There is currently an abundance of information available to livestock managers, but it is not generally structured in a way that can be applied readily.

Many producers perceive that adopting high productive management systems involves increased risk. The perceived risks include financial failure because of unforeseen environment or market circumstances, damage to the farm infrastructure such as soils and pasture, compromises to animal health and welfare, and increased stress on farmers from managing an intensified system. These risks are real. Thus, it is important to develop a management system that ensures only the most essential procedures are carried out, they are all carried out correctly and consistently, and in a way that controls risk.

2.4. Integration of traceability with PLF

Traceability within livestock management has largely been limited to movement and disease control applications such as the European passport system for cattle. There are a number of objective reasons why the integration of traceability and PLF has not progressed further, which include (1) availability of easy to implement and affordable automated identification systems, (2) overemphasised privacy concerns related to data captured on-farm, (3) inconsistent offering of traceability products to farmers, and (4) too much focus on particular numbering technologies (simple numbering, barcode, RFID).

The most interesting example of the integration of traceability with PLF is the exchange of information along with the feed-animal-food chain. This information exchange (Figure 1) has a number of benefits: (1) Feed and feed input providers can greatly improve the composition of their products if they have access to slaughterhouse statistics resulting from the feeding profiles applied on the farm; (2) Farms can use such a system for the selection of the right feed (or right feed provider). They can also optimize their feed use/intake from the statistics of other farms on the network; (3) Abattoirs can use the system as a basis for cooperation with farms to produce and source more animals on weight and conformation specification; (4) Industry statistics are a very important tool for both governments and the industry itself to steer the sector. Reliable statistics can be used for political decision making, benchmarking, lobbying and business decision making.

15.6. ábra - Figure 15.6: Tracebility Systems and ulinkage with PLF
Recent developments in communication technology through mobile phone technology, telecoms and the internet offer a huge potential benefit to the design, application and value of PLF. Whilst independent applications on individual farms may be desirable to some customers, the advantages of centralised data collection, processing, management and reporting are significant. For example, data collected by sensors on the farm can be sent to a central site for processing, storage and reporting. This could result in considerable time saving for farm managers to be allocated for more productive tasks, such as farm and animal husbandry related tasks. The centralised processing should supply him with only the data pertinent to his daily needs, with more detailed reports available as required, including through the centralised database the comparative performance of his unit, for example. In short, the benefits offered by a good PLF system should be obvious to the user and ideally should reduce his management.

2.5. Commercial issues

The commercialisation principles of PLF technologies need to include (1) a verification of the benefits of the PLF technique being proposed, (2) a clear communication of those verified benefits to customers, (3) identification of principle beneficiaries (i.e. operator vs. owner of the business), (4) provision of appropriate training and technical support, (5) correct specification, installation, commissioning and monitoring of the installed system. Unfortunately, PLF developments have been largely spear-headed exclusively by academic organisations so far. In general, there is an inadequate engagement of commercial companies in the PLF technology development process. In order to increase the interest of suitable companies in providing services to farmers, collaboration between smaller specialist firms and larger generalist firms is desirable. Transferring PLF technologies to companies supplying and managing the systems is a significant step towards developing commercial PLF tools/products that are wanted by customers and sold with confidence.

The greatest problem of commercialisation is the lack of a consistent service offering for farmers. Farmers are biologist by nature and only technologists occasionally. There is a need for a service sector that will be able to (1) take care of technology components, (2) interpret data captured by sensors, (3) formulate and send simple, relevant advice to farmers on a regular basis, and (4) involve users in technology developments. This service sector would need to use suitable business models that avoid high initial investment costs for farmers. Affordable monthly or annual fees might well be compatible with farmers’ cash flow; especially they are unlinked to performance improvements or animal sales. Although farmers usually invest part of their gains in technology, it is a typically machinery that they would look forward to buying (as opposed to software or sensors).

The food industry in general is a very conservative industry and with good reason. Although it is one of the largest industries world-wide, its margins are very small and its products are usually very delicate. Agriculture is in addition a fragile industry, because it depends directly or indirectly on climatic factors and seasonal demand/supply circles. In addition, even for the more adventurous farmer it is very difficult to judge the applicability of a particular technology and “guesstimate” its benefits. In other words, an important missing
element is the absence of clear cost benefit data on PLF that takes into consideration the complexity of farmers’ purchase decisions. Demonstrating and verifying the economical, welfare and environmental benefits of these technologies are essential in the commercialisation process.

The other key limiting factor of adoption rate of PLF technologies on farms is the lack of co-ordination between researchers, developers and technology suppliers. Achieving better co-ordination between the developers and suppliers of PLF tools is very difficult, but would result in the development of better integrated systems. That in turn would result in greater commercialisation of PLF systems as integrated systems to serve the farmers better. In addition, many of the PLF “products” actually never have been “productised” (developed into a proper “product”); but they went directly from the lab to the farm. Only some larger firms with enough development funds have taken up PLF as their guiding principle.

3. Enhancing Productivity on the Farm

How can farmers and governments use ICTs to increase agricultural productivity? At the local level, farmers can use ICTs to match cropping practices to climatic trends, use inputs and resources environmentally and sustainably, and cope with productivity threats. At the national level, public officials can adjust policies to reflect the data collected with ICTs, predict food supplies, and target social programs or promote yield technologies. Integrating ICT into national programs, creating a policy environment conducive for ICT investment, and designing digital systems that are compatible and common can help improve access for users, but social and financial challenges remain.

New ICTs help to characterize field conditions, sometimes at a very fine level of detail, and help farmers improve soil and land productivity. Correcting past damages and ensuring future yields will require farmers, governments, and development partners to mitigate the effects of climate change and environmental degradation. Significant, national progress with some of these technologies will require appropriate legal and regulatory frameworks, monitoring systems, and liability, access, and property rights laws and regulations, such as regulations on carbon limits.

15.7. ábra - Figure 15.7: Some information flows

ICTs have considerable potential to help even small-scale producers prevent losses after investments have been made by identifying and controlling pests and diseases, receiving timely weather information, and improving resource use. At the same time, ICTs allow governments and development partners to better monitor farm productivity, make more accurate projections, and plan better for the future. ICTs should be used to form two-way communication networks that gather and use local knowledge. Advances in ICT are best suited to helping farmers improve their management of one or two farm components at a time. Development partners and governments need to prioritize which yield technologies or agricultural strategies to introduce. Incentives for
partnering with the private sector in large-scale ICT projects may enable the investment to reach smallholders. Taking stock of the technical capacity in rural areas will clarify infrastructure needs.

Given that the future of food depends to such a great extent on small-scale agriculture, governments and development.

Partners are focusing on how to increase productivity in sustainable ways through new technologies that smallholders can use. Irrigation management, biotechnologies, pest management and eradication, soil assessment, improved nutrient and land management, improved market access, and innovative storage facilities are all strategies for increasing smallholders’ agricultural productivity and improving their access to markets, but the challenge lies in ensuring that smallholders can obtain and use them. ICT provides an incredible opportunity to reach farmers with the technical information they require to increase yields. ulinking Technology for Agricultural Productivity with ICTs This module discusses two sets of technologies and the ulinks between them:

• Yield technologies, like improved seed, crops developed through biotechnology, tractors, pesticide, fertilizer, and irrigation systems.

• Information and communication technologies, like geographical information systems (GIS), wireless sensor networks, data mediation software, and short message service (SMS).

When farmers have access to biophysical and other yield-enhancing technologies, frequently they do not know how to use them effectively to address their productivity challenges (for example, they may have fertilizer but not know the optimal amount to apply). ICT can fill this gap in knowledge.

Global positioning systems (GPSs), radios, mobile phones, digital soil maps, and other ICTs give farmers information to use biophysical technologies appropriately (for example, nitrogen sensors can help to determine the correct fertilizer dose). Similarly, governments or development partners may know that farmers are using new yield-enhancing technologies but may not have the capacity to understand their impacts. Data-mining technologies, decision-support systems, and modeling software that can clarify the impacts and outputs of yield-enhancing technologies are among the most promising means of linking productivity and ICTs. This module describes how farmers and governments can use ICTs in their strategies to increase agricultural productivity. The applications are quite broad: ICT can be used to monitor pest thresholds in integrated pest management, provide relevant and timely information and agricultural services, map agrobiodiversity in multiple-cropping systems, forecast disasters, and predict yields. Crop losses diminish as farmers receive relevant and timely information on pests and climate warnings through SMS technology. Just as important, information can (and should) go both ways: Farmers can alert local governments or other relevant actors about serious crop developments like disease symptoms. This information makes it possible to avoid disasters more effectively and improves economic management, both of which are crucial for adapting to climate change. ICT can also lead to more optimal use of inputs. Increasing producers’ knowledge of how to use and manage water, equipment, improved seed, fertilizer, and pesticide has improved the intensification of farm practices around the world. In the long run, and after collecting and analyzing multisite, multiyear data, ICT can be used to match cultivars to appropriate environments, increase the understanding of genotype-by-environment interactions, and adapt cropping strategies to the changing climate.

Each of these applications increases the profitability of agriculture, reduces transaction costs, facilitates climate change adaptation, and improves livelihoods for the rural poor. Strategies to increase yields (including strategies to avoid yield losses) include initiatives like soil nutrient assessments, weather forecasting, and crop or animal protection. “Achieving Good Farm Practices through Improved Soil, Nutrient, and Land Management,” focuses on soil testing technologies and tools that characterize field conditions, sometimes at a very fine level of detail. These technologies help farmers apply inputs appropriately and encourage the use of sustainable, profitable farming practices. “Preventing Yield Losses through Proper Planning and Early Warning Systems,” focuses on how ICTs can be used to identify and control pests and diseases, improve access to timely weather information, and improve the design and management of irrigation systems.

3.1. Remote Sensing Technologies: Raw Data Collection

The first type of ICT that improves productivity includes tools that collect agricultural data:

• Geographical information systems (GIS) collect geographic data through computer hardware and software to capture, store, update, and display all forms of geographically referenced information by matching
coordinates and time to other variables. Data sets formed by GIS constitute “layers” of information (for example, on topography, population size, or agricultural household income) that can be merged and analyzed to establish relationships and produce maps or charts that visualize geographical traits.

- Global positioning system (GPS) is a satellite-based positioning and navigation system with three basic components: satellites that orbit the earth, control and monitoring stations on the earth, and the GPS receivers owned by users. GPS receivers pick up signals from the satellites, including precise orbital information (latitude, longitude, and ellipsoidal GPS altitude) of a given object or location, as well as the time. GPSs can function in any weather and are free for public use.

- Satellite imagery is an image of Earth taken from satellites in orbit. There are four types of satellite imagery: spatial (size of surface area); spectral (wavelength interval); temporal (amount of time); and radiometric (levels of brightness)—which capture a variety of variables about a given area of varying size. The resolution (in meters) of these images depends on the satellite system used and its distance from Earth; weather can interfere mainly with satellite systems utilizing visible wavelengths of light. The cost of the technology depends on the satellite system used, on whether new or archive imagery is purchased, and on possible georeferencing to a coordinate system.

- Aerial photography and orthophoto mosaic. An aerial photo is an image (once a photograph, now a digital image) of the ground taken from an airplane, helicopter, or radio-controlled aircraft at a given altitude. Aerial images are presented as an orthophoto mosaic that is an alternative to a map. These images are higher in resolution (decimeter) than satellite images, proving useful for those who want more details of the terrain such as crop conditions or land use. In addition, modern digital aerial photography is georeferenced—that is, each point has geographical coordinates, whereas satellite imagery requires georeferencing to be geographically accurate and compatible with other geographical data.

- Laser scanning, or light detection and ranging (LiDAR), is an active airborne sensor using a set of laser beams to measure distance from an aircraft to features on the ground. Airplanes and helicopters can be used for laser scanning. The data from laser scanning are three-dimensional at very high accuracy, and they also allow ground elevation under the tree canopy to be measured. The elevation accuracy of laser scanning data is much better than aerial photography, which makes laser scanning useful for accurate topographic mapping where elevation is critical. The data can also be used to measure forest attributes such as the height and density of trees and thus the volume (aboveground biomass) of the forest.

Information Management Technologies: Making Sense of the Data The raw data collected above are fairly useless without analytical tools, both human and inanimate:

- Spatial modeling (among other models). Closely related to spatial analysis or statistics, models are an attempt to simulate real-world conditions and explore systems using their geographic, geometric, or topological properties. GIS (which can also perform analysis), among other ICTs, has increased opportunities to create models that predict occurrences like yield growth and ecosystem degradation.

- Data mining is the extraction of stories or patterns from large amounts of data. Data mining can find four major patterns: clustering (discovering groups), classification (forming a structure), regression (finding a function), and associations (finding relationships). These analyses help to make sense of agricultural data collected by remote sensors.

- Data mediation is the process of taking many different data sets to produce a single, coherent set of information. Data mediation software organizes different types of data (such as hourly versus daily) and synthesizes different approaches to classification (for example, the use of different classification vocabulary), helping to mediate differences between data sources—particularly those on the Internet.

Dissemination Tools: Getting the Results to the Stakeholders After analysis, the results must reach those who need to react to the findings, using tools like:

- SMS. Text options that allow interaction between fixed-line and mobile phones.

- Radio. Transmission of information through electromagnetic waves with low frequencies.

- WiFi. Wireless local area network that allows various devices to connect to the Internet remotely.

- Knowledge management system. Electronic system that provides relevant information as it is requested.
Increasing smallholder productivity is one of the greatest tasks in this century. Although the dimensions of the challenge are huge (growing populations, growing demand for food, rising poverty, economic stagnation, worsening environmental degradation, and climate change), the growing number and sophistication of ICTs offers some hope of raising agricultural productivity, even in smallholders’ fields. Variable rate technology, GIS, GPS, satellite imagery, and other data collection technologies have increased the information available about soil health, weather conditions, and disease outbreaks, making very site-specific farming possible. The key to using these technologies to boost productivity is to remember that complementary technologies are needed: Data analysis technologies (such as data mining or mediation software) and information dissemination technologies (such as mobile phones and radio) are essential to reaching smallholders effectively. Dissemination also includes the crucial human component: Extension agents and farmers themselves must transmit and share knowledge. As noted, productivity can be increased by expanding the land available for agriculture or by making the land already in use more productive. Given current global circumstances, it seems that the second option is more likely to close the productivity gap and meet demand. In conjunction with technologies developed to raise yields, the use of ICTs such as those discussed in this module may do just that. Mainstreaming the use of ICTs in agriculture will also enable them to be used more effectively. Integrating ICT into national programs, creating a policy environment conducive for ICT investment, and designing digital systems that are compatible and common can help improve access for users. Conducting impact studies and sharing pilot project information is also critical to success with ICTs, as more specific lessons and impacts are learned. In closing, it is important to emphasize that the benefits of ICT can be realized on multiple levels. As ICT capacities expand, local farmers and communities as well as nations and regions need to understand their potential uses to increase agricultural productivity. These stakeholders must learn how to tailor ICT solutions to macroeconomic needs as well as local agricultural bottlenecks, while exploring how current infrastructure can harness relevant and appropriate technologies.

### 3.2. Assessing Soil Properties for Climateresilient Agriculture

Accurate soil analyses and improved farming practices are needed urgently because productivity gains are highest in healthy soils and where pesticide, fertilizer, tools, and machinery are used properly. Instruments for mapping and analyzing soil properties have proliferated in the last decade, increasing farmers’ knowledge about the soils on their farms and the need for climate-resilient agricultural practices. The following section discusses these technologies and their associated challenges in broad terms. Subsections discuss innovative technologies specifically related to nitrogen and carbon, two essential chemical components for successful soil conservation and climate change mitigation. Digital soil maps are the most promising applications for visualizing soil properties and the gravity of soil nutrient depletion in a particular area. The International Working Group on Digital Soil Mapping (WG-DSM) defines digital soil mapping as “the creation and the population of a geographically referenced soil database generated at a given resolution by using field and laboratory observation methods coupled with environmental data through quantitative relationships”. A variety of technologies, including satellite remote sensors and cameras, can be used to survey soil and collect data to create digital soil maps. These technologies collect soil information faster than methods that require scientists to take soil samples from the field. In the latter methods, 80 percent of the work on soil mapping is dedicated to soil identification and boundary mapping, and only 20 percent of the time spent in the field is left to gather data on more complex and equally important topographical features, such as water-holding capacity. Innovative data collection technologies allow researchers to focus on a variety of soil features.

A number of ICTs can be used to measure soil properties for creating digital soil maps. Through near infrared and short-wave infrared sensors, satellites measure spectral reflectance in soils on the ground. Different materials reflect and absorb solar radiation at a variety of wavelengths. As a result, remote sensors can measure soil color, texture (sand, silt, and clay), organic matter, moisture, salinity, and absorption processes by detecting and observing the solar radiation reflected (orbit sensing). Reflectance changes depending on the soil’s contents; for example, reflectance is low in areas with low silt content. This technology gives researchers an accurate assessment of soil properties to use in GIS and computer modeling for digital soil maps.

Digital soil maps give practitioners a good picture of soil fertility, vulnerability, and potential. Statistically testing soil maps against other data on human or policy variables (like demographics, land administration, farming practices, and climatic changes) allows researchers and others to explore causes of soil damage and forms of restoration. At a national or regional level, models created from digital soil maps can be used to improve the selection of crops and varieties (based on which crops and varieties can withstand stressful soil conditions). They can also be used in early warning systems (predicting crop failure, for example), giving policymakers more time to react to shortfalls in domestic and export markets. In addition, fineresolution soil maps collected from a number of regions could enable climatologists, hydrologists, and crop modelers to more...
ICT Systems in Farm Production Management

accurately predict the effects of climate change or new technologies on food production and environmental health. After soil data are collected, analyzed, and reflected in digital soil maps, results need to be shared with policy makers, scientists, and especially farmers, who would otherwise not have such detailed information on soil fertility in their respective farming communities. Recent developments in ICT increase the cost-effectiveness of soil maps: The spread of mobile phones and Internet access can transfer relevant soil information even to remote locations. Collaborating with extension staff, farmers, agrodealers, and others, development institutions can generate integrated soil fertility management schemes that improve a wide range of farming practices. Challenges in Soil Mapping Although technological developments have improved access to digital soil maps, major technological and economic challenges remain to be addressed in soil science and development institutions. Broadly speaking, the impacts and outcomes of using digital soil maps in smallholders’ fields have not been captured. Soil assessment techniques certainly contribute to the knowledge of production potential, but the transformative effects of this knowledge (such as the adoption of new practices) have not been tested empirically. Another technical challenge is that some digital soil maps cannot be used in quantitative studies or in models of food production or carbon management. Such studies generally require information on the functional properties of soils, such as available water capacity, permeability, and nutrient supply, which many mapping procedures do not capture. Finally, individual soil map units are shown as discrete polygons with definite boundaries. The data used in polygon maps are difficult to integrate with other forms of data, which are grid-based (like satellite images and digital elevation models). Social and financial challenges remain as well. Detailed yet inexpensive soil analysis tools are not widely available for small-scale farming in most developing countries, although they are being developed and piloted. Even where technologies are free to the public (like online satellite images), the resolution is too low to capture soil characteristics on individual plots. Without accurate, affordable soil analysis technologies, resource-poor farmers are unlikely to adopt sustainable and resource-optimizing farming practices. These practices are often more expensive in the short term and are typically more labor intensive. Finally, disseminating knowledge about soil management and farming practices is challenging. Soil science is complex. Soil restoration activities vary based on a diverse set of properties and the agroecological system. Even digital soil maps that create opportunities for soil assessment at the local level will require major dissemination and training efforts by extension staff and other stakeholders. These challenges are being overcome as technologies advance.

3.3. Preventing Yield Losses through Proper Planning and Early Warning Systems

ICTs can help to prevent and reduce losses in crops through well-planned investments and disaster warnings or timesensitive alerts. Water management and disease or pest prevention are crucial to increased productivity. Advances in ICTs such as GPS, GIS, mediation software, mobile phones, and satellite imagery have improved smallholders’ ability to adjust farm strategies and reduce risk. At the same time, these advances allow governments and development partners to better monitor farm productivity, make more accurate projections, and plan better for the future. Water is a primary topic in this thematic note. Although water is scarce and is becoming more so due to climate change, many water resources in developing countries are simply not exploited.

Weather data, along with improved irrigation management and system engineering, are more important than ever. This note also discusses disease and pest control. Pests and pathogens continually evolve, making it particularly difficult for small-scale farmers to increase productivity. Without inputs like pesticides and the knowledge to use them correctly, pests and diseases reduce global harvests by upwards of 30 percent for maize, rice, and potatoes. With ICTs, governments find it easier to reduce crop losses from flies or rodents and livestock losses from disease like bovine spongiform encephalopathy (less formally known as “mad-cow disease”), more efficiently use the total amount of pesticides employed in crop protection. Farmers often are unaware of or cannot accurately assess plant diseases, which may reduce agricultural productivity and raise costs if pesticides are overused. Concerns for animal health are similar. Herdsmen and fishermen spend resources and time treating sick animals or identifying disease outbreaks. Using a variety of ICTs, producers can better identify, track, and protect their crops, animals, and livelihoods. Protecting farm animals from disease and other ailments also improves through ICT. Sensors and other remote technologies can be implanted in an animal, providing hersman with the exact location, health, and situation for livestock like cows, pigs, or sheep. In addition to enabling easier identification and tracking, in the future, some instruments may offer animal response systems. ICT is now being used in integrated pest management systems to improve farm management in a variety of ways. The Low Frequency Array Project (http://www.lofar.org) piloted in the Netherlands uses sensors to monitor and treat potato crops at risk for the fungus Phytophthora infestans, which causes late blight. Because the development of late blight depends heavily on climatic conditions, capturing climatic conditions like humidity and leaf temperature can help farmers prevent onset of the disease by optimizing fungicide
applications when climatic conditions warrant it. The project used three instruments: sensor nodes, a server, and a decision support system. One hundred and fifty sensor nodes, called TNodes, send soil information every 10 minutes through a TinyOS operating system to the server where data are stored. Users can access this information directly, or receive texts or emails from the ulinking decision support system. The decision support system gathers information from the server along with other meteorological data from weather stations to produce maps of the temperature distribution within fields. The system sends alerts to the farmer that identify the patches of land most susceptible to the fungus. Information technologies are vital for disseminating crop protection advice, but “crowdsourcing,” (using ICTs to leverage widespread collaboration) can prevent diseases from spreading in the first place. If sufficient numbers of farmers can text information on potential crop disease symptoms to researchers and receive appropriate disease control advice, researchers can also track and potentially forestall epidemics. If farmers or cooperatives have access to the Internet, online bulletin boards or mailing lists can spread information on disease incidence quickly. Online decision support systems5 that ulink data to possible action, such as the one used in the Low Frequency Array Agro Project, are becoming more popular because clients require minimal software, which reduces management and distribution costs. Additionally, it is useful to ulink weather information to pest or disease development over time. The Pacific Northwest Integrated Pest Management website through Oregon State University (http://oregonstate.edu/dept/nurspest/) collects temperature and precipitation data from 380 weather stations and ulinks it.

3.4. Weather Forecasting

Since 2000, new ICTs have given farmers and partners better opportunities to manage climate risk. WSNs and satellite images capture raw data that can be transformed into information useful for agriculturalists, helping them optimize decisions related to choosing crops (based on water requirements), planting (timing and planting density), buying inputs, and applying fertilizer. Climate information can also improve insurance markets. In eWarning and other systems, farmers request information through SMS in two forms. Push-type messages are regular, automatic updates obtained through a user subscription. Pull-type messages are sent only when a user requests them. When the user sends a letter (like “P”) in a message, the eWarning system will respond with information on precipitation for the user’s geographical location. Surveys show that the push-type message is most popular, providing farmers with an hourly forecast up to four times per day.

3.5. Irrigation Management

Major water resource constraints and climate change make it increasingly important for developing countries to develop sound water-use policies and well-functioning, well-managed irrigation systems. Innovative water management systems and ICTs are helping to improve water use and expand intensive irrigation facilities. Though the number of technologies for irrigation is vast, this section focuses on remote sensors, satellite imagery, and GPS cameras. Each of these technologies helps to connect the farmers to irrigation infrastructure and guide governments in designing and implementing irrigation strategies. ICTs help address some of the challenges inherent in creating and sustaining irrigation systems in rural areas. The function of water-user associations and their productivity improve through ICTs like mobile phones and personal device applications (PDAs), which increase the quality and frequency of producers’ communication and interaction. Sharing information about emergency maintenance problems, entitlement rights, and management schedules is facilitated through ICT, which allows real-time responses even between users from distant communities. Digital orthophoto quads (DOQs), a feature of GIS, are digital maps that combine the geometric information of a LiDAR (laser scanning) is a new technology for obtaining a highly detailed digital terrain model or, if equipped with an aerial camera, for topographic mapping. A digital terrain model is basically a digital representation of an area’s terrain on a GIS that provides accurate position and elevation coordinates. It is compatible with other digital spatial data, is more accurate, and has a higher resolution than satellite images. Elevations can be accurate within 5 centimeters, but accuracy typically is closer to 10 or 20 centimeters. In comparison, digital aerial cameras only provide only about a 20-centimeter horizontal resolution. Because of its detailed imagery, a digital terrain model can be used for meticulous engineering designs such as those for roads, drainage, gravity-fed irrigation works, and detention reservoirs. These models can also be used more broadly to manage land and water (for example, in flood control). When combined through GIS with other data such as soil types, these models can help to identify areas with potential slope instability and erosion, which are important for reducing soil degradation and its negative impact on soil fertility. At the field level, digital terrain models can monitor and improve areas affected by waterlogging or flooding. Overall laser scanning has considerable potential for planning irrigation schemes, designing infrastructure, managing irrigation operations, and modeling. Laser scanning is most useful for large areas because the aerial operation is expensive. The cost of laser scanning also
depends on the accuracy of the data required, location of the area of interest, and level of the data products (such as GIS layers).

This note has described the many ways that ICTs enable realtime adjustments in agricultural practices to prevent losses after investments have been made. These technologies also have considerable potential to help small-scale producers use scarce resources—water, nutrients, and others. Greater certainty about the weather, access to water, and disease outbreaks can lead to better decisions and higher productivity. These ICTs also face important challenges, however, and a number of considerations are important in improving their effectiveness, especially for smallholders. Strategies to improve agricultural practices change dramatically over time, just as strategies to manage irrigation have evolved from a nationally operated to user-operated model.

ICT in agriculture receive current information, communicated in the most cost-effective way. Local knowledge is critical to improving smallholders’ productivity. ICT not only creates opportunities to disseminate information but offers ways of capturing local expertise. Vast differences in ecological and agronomic conditions make farmers’ knowledge indispensable. ICTs should be used to form two-way communication networks, ensuring that local knowledge is acquired and utilized. The collective action problem is quite apparent in relation to the technologies described here. Water management and disease control require hundreds or even thousands of farmers to perform the same tasks in unison. By strengthening information sharing, ICTs like mobile phones will increase the potential for collective action. Self-policing may also be crucial to the technology’s success. ICTs to disseminate information like weather forecasts must match capacity in the focus area. Some phones handle complex messaging; others do not. Local ICTs may need to improve before some preventive technologies can work in developing countries. Taking stock of the technical capacity in rural areas will clarify infrastructure needs.


Four kinds of financial services help farmers to achieve their economic goals: credit, savings, transfer and payment facilities, and insurance. The major prerequisites for using information communication technologies (ICTs) to deliver these services in rural areas are robust national financial systems (for example, with national payment systems, credit bureaus, ATM switches, central platforms for microfinance) and the infrastructure that allows electronic financial transactions between institutions and individuals. Factors that are critical for ICTs to expand financial services in rural areas are a supportive economic policy and regulatory framework; appropriate financial and nonfinancial products; and mechanisms, processes, and technology applications that can deliver products and services, improve transparency and accountability, reduce costs, and become self-sustaining. New channels for delivering financial services (facilitated by ICTs), new players, and greater competition enable service providers to offer a larger suite of financial products and services and acquire better financial information, some of which is useful to government regulation and policy development. A number of nonbank institutions have developed innovative approaches to financing agriculture, enabled by or integrated with ICTs, including mobile financial services, branchless banking, ATMs, and smartcards. Often governments lag in introducing the policies and regulations needed to extend cost-effective financial services throughout the economy, including underserved rural areas. To design supportive policies, provide the necessary infrastructure, and provide appropriate, affordable financial products meeting local needs, governments must explore partnerships with the private sector and rural communities. In turn, governments can devise and implement policies that give rural communities and private enterprises incentives to participate in the rural financial sector.

3.6.1. Services

- Credit, in the form of loans, personal loans, salary loans, overdraft facilities, or credit lines, is often used as working capital at the beginning of the growing season to purchase inputs and prepare land. They also need capital to invest in equipment such as tractors or drip irrigation and to harvest, process, market, and transport their produce. It is important to distinguish between short-term loans, which microfinance institutions usually provide, and the long-term financial services required for agricultural and livestock enterprises.

- Savings may be in the form of current accounts, savings accounts, or fixed or time deposits. Farmers have a significant need for savings, because their income is seasonally tied to the harvest, and for much of the year they rely on savings to smooth consumption.

- Transfer and payment facilities allow for local and international money transfers, remittances, government transfers, and check clearing.
• Insurance may cover crops and livestock as well as human life and health.

ICTs have now created the potential to deliver a greater diversity of financial products to greater numbers of rural clients than conventional financial service providers have been able to reach. ICTs can also enhance the government’s capacity to monitor and evaluate financial services provided to rural clients and design effective financial policies and regulations for the rural sector. A number of agents in rural areas—such as government departments, commercial banks, microfinance institutions, traders, telecommunications companies, community-based organizations, families, and friends—provide financial services, which can include credit, savings, insurance, transfers, and payments. Even so, tailoring and providing financial services for small-scale farmers remains challenging. Rural clients differ from the typical clients of financial service providers. They are located in remote and often sparsely populated areas, and they rarely possess the sorts of physical or financial assets that financial institutions customarily accept as collateral. Typical rural assets, such as livestock, pose challenges of inventory assessment and management, and collateral substitutes based on warehouse receipts or returns from future crops are unavailable in many countries. Farmers also have a special need for financial products with a time horizon extending over multiple crop cycles. Major prerequisites for using ICTs in financial services for agriculture are robust national financial systems and the infrastructure that allows electronic financial transactions between institutions and individuals. Two types of infrastructure and related services facilitate electronic transactions and are vital for extending financial services to rural areas.

The first is ICT infrastructure, such as high-speed Internet and mobile phones, available at affordable prices. This infrastructure is the backbone of electronic financial transactions. The second is financial infrastructure, which includes national payment systems, credit bureaus, ATM switches, or central platforms for microfinance institutions. Financial infrastructure enables financial service and technology service providers, as well as other providers vital for the integrity and stability of the financial system, to connect and perform transactions in real time. For example, financial infrastructure makes it possible for customers of one bank to use the ATM of a different bank or conduct a transaction (such as writing checks or wiring money) with customers of a different bank. It also channels financial information (such as the creditworthiness of a new customer) to financial institutions. These services and infrastructure do not benefit merely one operator or financial service provider; they cater to the entire rural and financial sector. For this reason, their provision is often initially regarded as a task for government, although in reality they can be (and often should be) provided by the private sector alone or in partnership with government.

3.7. The Use of ICT-Enabled Financial Services in the Rural Sector

ICT introduces new channels for delivering financial products and services to the rural sector, and it has the potential to reach farmers, intermediaries, entrepreneurs, and rural dwellers more directly than traditional brick-and-mortar bank branches or microfinance offices. These new channels enable financial service providers to offer a larger suite of financial products and services and acquire better financial information, some of which is useful to governments as they oversee, regulate, and develop policy for the agricultural and rural sectors. It illustrates how ICT expands the traditional relationships and service capacities in the rural finance ecosystem.

Interventions using ICT can introduce new players and lead to greater competition in the rural financial sector. Institutions or agencies that are not banks (nonbanks) may start providing rural financial services. Since the early 2000s, a number of nonbank institutions have developed innovative approaches to financing agriculture. They have sometimes adapted microfinance concepts to provide agricultural finance, used good banking practices, and above all, drawn on knowledge of agriculture and ICT to enter and succeed in this market. Many of these new approaches show great promise, but no single approach will work for all situations. Rather, organizations have the most success when they are not dogmatic, apply innovative and comprehensive risk management strategies and tools, and retain the ability to perform credit analyses of their intended rural clients without political interference. Nonbanks and banks can provide these ICT-enabled financial services for the rural sector:

15.8. ábra - Figure 15.8: ICT and the Rural Finance Ecosystem
Mobile financial services. Given the pervasiveness of mobile phones in developing countries, financial services providers can use them to reach clients in rural areas and provide a broad array of financial products and services, including credit, insurance, payments, and deposits. Financial service providers can tailor financial products offered through mobile phones to rural needs.

Branchless banking. Field agents, equipped with mobile phones or point-of-sale devices, can serve as mobile branches. Agents can provide financial services to smallholders, take deposits, provide financial information, and keep records of clients’ creditworthiness. In this way, branchless banking deepens financial inclusion throughout rural areas.

ATMs. Though ATMs are often associated with debit cards or smartcards, ATMs can serve as cash-dispensing machines in tandem with branchless banking, mobile financial services, and other ICT-enabled financial products. The availability of ATMs in rural areas can place cash-exchange points within reach.

Smartcards. Though not entirely in the category of ICT, smartcards (or stored-value cards) are an alternate means of providing services when mobile financial services are not readily available. Pre-paid cards, debit cards, or credit cards provide payment and credit facilities to rural clients. Stored-value cards have historically assumed some level of literacy (in particular, the ability to sign for a transaction), but the advent of smartcards that use biometric devices eliminates the challenges associated with literacy barriers.

As discussed, financial services rely on the availability of underlying financial and ICT infrastructure, such as payment systems, credit bureaus, central ATM switches, central financial platforms, mobile telephony, mobile data services, and Internet in rural areas. Governments have to work with the private sector to ensure that the underlying infrastructure is in place and extended to rural areas.

4. Policy Strategies and Regulatory

Farmer organizations can function more efficiently by using information communication technologies (ICTs) to attract and retain a wider membership, generate more funds, and provide better services to their members. Documented benefits of ICTs include improved connections to members, better accounting and administration, and stronger collective voice. Given the lack of basic infrastructure in much of the developing world, the most successful ICTs are robust and relatively simple. Governments, donors, and nongovernmental organizations (NGOs) generally initiate the development and testing of ICT solutions for farmer organizations, but in many instances partnerships with the private sector are essential. Two important challenges are to sustain the use of ICTs over the long term and ensure inclusiveness. Mobile phone systems appear to be the most flexible
technology for improving connections within farmer organizations and providing a wider range services. Technologies that do not depend on literacy (digital photography and video clips) are extremely effective for sharing information within and between farmer organizations. Computerized record-keeping has transformed efficiency in farmer cooperatives; approaches include commercial systems and systems using open-source software. Supportive government policy and willingness on the part of government organizations to join partnerships are important enablers. Farmers’ collective voice is stronger and reaches wider audiences with the help of radio and television. Interactivity is possible and even more promising through phone-in programs and text messaging. Radio and television are also effective tools for agricultural extension. Interactivity through websites is becoming more important for farmer organizations, but less so for individual smallholders.

Farmer organizations play an important role in tackling the systemic causes of poverty, because they give farmers—men and women—a legitimate voice in shaping pro-poor rural policies. By articulating farmers’ interests to public and private institutions, farmer organizations encourage those institutions to tailor their strategies, products, and services to farmers’ needs. Given a supportive policy framework, farmer organizations are well able to drive balanced social and economic development.

As well as forging institutional links and giving farmers a collective voice, farmer organizations provide services to their members. Smallholders can generate more income in a number of ways—such as by using better cultivation techniques and improved seed, reducing postharvest losses, and having better access to markets—yet as individual entrepreneurs, they may lack the knowledge or capital to change the way they operate. The collective strength of an organization can help its individual members become more efficient, if the organization’s services match its members’ needs. ICTs are integral to fulfilling both the lobbying and service functions of farmer organizations, speaking both for and to the farmer. Farmer organizations also have a third, commercial function, as seen in agricultural cooperatives and producer groups. Commercial activities become more efficient and transparent when supported by ICTs. “ICT” is a catch-all term for an increasing number of technologies, each offering corresponding opportunities for innovation. This module looks at a range of technologies, from the well-established and familiar technologies like radio and mobile phones to the more specialized technologies, such as computerized record-keeping systems and global positioning system (GPS). The discussion emphasizes technologies that can (or that have the potential to) reach large numbers of beneficiaries and perform reliably in the challenging context of the developing world. Different technologies offer different benefits, achieve different objectives, and have different limitations, so each is considered on its own terms.

When considering the value of ICTs to farmer organizations and cooperatives, it is worth bearing in mind that in remote rural areas of many developing countries, particularly in Africa, these organizations often are the only ones operating. Local government offices may be found in district headquarters, but often there is little else apart from frontline extension officers and schools. As a hub for business information, transportation, and storage, as well as a place where people share new systems and processes, farmer organizations have enormous potential—which should not be underestimated—for networking and bringing people together with the help of ICTs. The benefits offered by ICTs to producer organizations and agricultural cooperatives fall into three broad categories. Practical examples of ICTs in use sometimes cut across these arbitrary categories, and particular technologies may bring unexpected benefits, but the examples in this module are presented in three sections to reflect this categorization:

• Enhanced connections to members. Through the organization, farmers share market information and technical know-how, and they remain informed about the organization’s activities. For instance, topics discussed and decisions taken at board or executive committee meetings can be shared with members who, for reasons of distance or cost (direct and opportunity) cannot attend. Decision-making processes become more transparent, increasing trust between members, the board, and executive managers, and the overall functioning of the organization is improved.

• Improved accounting and administration. Farmer organizations are often responsible for handling very large amounts of money that may represent the cash income of thousands of farm families. Efficient record keeping allows an organization to serve its members better, and the transparency offered by computerization and other technologies enhances trust. Cooperatives that have invested in modern management and member information systems can improve their image to attract high-quality staff and gain members’ confidence.

• Stronger collective voice, including improved political voice. “Interactivity” as understood in developed countries with good infrastructure is still rare in many parts of the world. But individual farmers nevertheless “have their say” by phoning and texting their participation in agricultural radio broadcasts. They give feedback (and complain when necessary) about the services offered by their farmer organization and local
government. Comments are likely to have more influence expressed over the airwaves than expressed in a less public forum. Despite the potential benefits of ICTs, farmer organizations are rarely the first to adopt them, given that they usually work in difficult environments with low margins to generate income for their members. Neither managers nor members are preoccupied with the latest iPad. Where particular ICT solutions are available and necessary to guarantee better performance and benefits to members, farmer organizations can be expected to be late adopters of such technology without external support. In general, it is governments, donors, and NGOs that have the funds to develop and test ICT solutions that may benefit farmer organizations. Most if not all of the cases illustrated in this module are public-private initiatives to “include the excluded” by promoting ICTs in remote rural areas.

5. Futurefarm

We look at the different technologies and evaluate what is available today and what will be available in 2020 based on existing technologies. We further express our vision of how the new technologies and new concepts regarding organization will evolve. The factors that will have an enormous impact on ICT uptake during the next decade are the integration of technologies as ICT system integrators as well as the integration of organizations working on the overall agro-forest-environment-risk sector as a whole. Indeed a single component is not able to resolve integrated problems. We must all understand the urgent need for integration. Beyond the technologies and their evolutions we will also evaluate how these changes will impact on farming itself.

15.9. ábra - Figure 15.9: Innovation Funnel

The following technologies have been investigated:

- Satellite- and ortho- images
- GIS and Mapping technologies
- Telecommunication
- Sensors and mechanics
- Fertilizing
- Spraying
- Bedding out of plants
- Fleet management
Satellite- and ortho- images are already today a basic element in planning, driven by the EC orthoimage based control model in 27 member states. The images will have higher and higher resolution (satellite 1m and less, aerial images down to 30/10 cm and less), as well the processing of the images becomes an automated process with few human interactions, leading at the end to lower pricing.

The key question for the future will be to have these maps based on new sensors that are also able to, quickly and easily, manage the access to these maps, fast processing and transfer of the maps to the utilisation case – communicate with the experts or advisors and, from them, as well to machines that will be able to read and interpret these images. The information is used immediately to make better decisions. For precision farming needs, the image alone will is not sufficient, for decision making it will need to generate a bundle of base data like soil maps, the past and the planned future of the field, legal environment etc. making it possible to develop detailed planning for decisions.

5.1. GIS and mapping technologies

They have been expert driven systems in the past, heavy to learn and to run. Systems of the future – for the farmers, for advisors, for experts – all differently designed but able to cooperate - are already today on the market. They have to be easy to learn and to run, must be able to use more or less all different types of data, to integrate different technologies within the framework. As well the know- how of a typical ICT system integrator must run different applications within one region. This core capacity reflects that there is not only agriculture, not only forest, there is also environment with risk- management embedding this integration, and ICT has to reflect this. ICT must also be managed easily along vertical and horizontal chains. The farmer is not working alone but he is part of something bigger like cooperatives, sales channels etc. The software must reflect this.

Without of integrating the horizontal and/or vertical applications, systems will fail. It means at the end that the GI has to become part of an integrated system. The vision for 2020 is to have every farmer ulinked to farm management systems, either local or via internet and either himself or via an advisory service. This will push many benefits to the required increased levels that will make it possible to manage a sustainable agro- forest future supporting around 8 Bio people worldwide at that time.

5.2. Telecommunication

It has become a standard in the past few years. With technologies like GPRS or UMTS communication there is enough space for using them for IT communication usage. The cellular phones in the future will have - some already have today – GPS integration and with this it becomes possible to transfer data through their network. It has also become possible not only to send a contract to a mobile device on a tractor “where and what to do“ but also send a map with the contract that shows in detail what to do for precision farming. IT systems are also producing protocols of the done work and sending them back home. Further, telecom companies could become drivers of technology use in rural areas because such applications also means that the data traffic that is required can be profitable.

5.3. Sensors and mechanics

Sensors today are able to measure meteorological data, soil moisture data in different depths, height of harvested crops or the humidity of a biomass. The sensors of the future will be able to measure more content more precisely. In cooperation with telecommunication, an end- user will have access at any time to the results of sensors spread over the field or the country and will take this data as base data for his information needs that guide him to better decisions – e.g. spraying decided on the basis of the meteorology input within a specified
The farmers within this region – GIS based - can be informed automatically via an SMS message. Here again it will be necessary to integrate these data.

**5.4. Fertilizing, spraying, plant bedding**

There is a necessity in the future to optimize the use of fertilizers based on better information. The required information covers nutrient balances, sensor feedback, crop needs, as well as historical data of the field, soil- and even geological maps, and current status of the field at a given moment etc. This is justified through cost questions but also based on environment targets particularly in respect to water quality. We need better information to decide where to fertilize with what amounts. We, the experts and/or the farmer himself if he has that ability, need to produce detailed maps for that reason, transfer these maps to the machines and trigger the machines – or the driver in the case of an old machine – to work more precisely.

What is described about fertilizing must be fully carried over for spraying. We must have the capacity be increase the precision of spraying on a specific location with specific amounts of pesticide. Based on detailed maps the input for the sprayer is defined and this will in the future trigger single jet nozzles. Pattern recognition in close cooperation with images will support better input data for decisions. Also plant bedding will be done more and more precisely in combination with other technologies and tools.

**5.5. Fleet management**

Through the availability of maps, communication, mobile mapping systems, of GPS and dGPS etc. fleet management has become possible. Beyond the technological challenge of “screwing” together all the bits and pieces, fleet management raises the question of who must run the system: The farmers, the farmer organizations, the machine operators (not always the same), the industry etc.? At the beginning everybody wants to have access to the information and to have 100% of the benefits valued to the benefit of his organization. In a cooperation model this is not possible! The different organizations must learn to cooperate, to setup systems together by which every organization benefits and share the benefits among themselves. Then the overall benefit becomes significant: in time delivery of the industry or of the collectors becomes possible, traceability can be embedded. Beyond these benefits, the return on investment of such systems can be over only one growth period or less. Fleet management might become the horse that pulls ICT uptake in farming. Whether you have a developed or a non-developed region, the distribution of seeds, fertilizer etc. or the collection of harvests has to be organized better for the benefit of all stakeholders.

**5.6. Robotics**

Robots will be used in farming for special tasks such as pattern recognition ulinked to small robotics. As well, based on the pattern recognition results, specific diseases could be recognized leading to spraying on specific plant elements. This will require less energy, less CO2, less costs, and less environmental harm! The equipment is available as prototypes and it will be a question for local experts to build up the expertise on how a specific disease pattern looks like in a specific region.

**5.7. Traceability**

Traceability will give a unique identity to a specific crop, identifying which field it comes from, which farm and even which tree the fruit comes from. Today the technology is available to document, to transfer the core data and to organize IT-based handover of data from and to different chain elements. Related data such as temperatures during transport or the protocol of a detailed transport route can be traced. The question today is much more the question of who is owner of the data, who is willing to share which information, and data with whom, rather than the technology alone. More or less all elements can be handled based on specific legal environments or on bilateral agreements. It is again very much a question of cooperation and of organizations working together as it is not required that all food/value chain has its own traceability systems and database. They could share ONE technology and on the basis of this common system compete with better quality of products while traceability is in any case a legal requirement. Thus for CEO’s of food companies a traceability becomes a risk-management tool. The ICT solutions serving the overall system – let us call it Trust Centre – could be the management system for natural resources and beyond traceability of food/feed it can be extended to also cover broader natural resource management elements.

**5.8. Environmental achievements**
With the production of crops and logs farmers and foresters also take part in the optimization of “environmental caretaking or even risk reduction”. Through deeper analysis of the “systems” we can see that the target of crop or log production in a specific manner does not always coincide 100% with the production of better risk mitigation or better environment (Climate Smart Agriculture). In the future experts will be able to precisely define what has to be done to increase the capability of a specific piece of land to be better managed with risk determined on the basis of detailed measurements and reliable data.

5.9. Genetics

Genetics will help us with dealing with issues such as drought resistant crop sorts or species with higher growth rates; nevertheless as we have learned from the past that humans do not always the capacity to foresee the future consequences of their decisions. We must think of totally new models for evaluating in all details the consequences and how models to achieve this will look like.

5.10. Pressure in ecology

Today we have a high need to do things differently as we are faced with many environmental problems. Climate change, from a worldwide perspective, is understood as the most important challenge. We also have to deal with an enormous growth in population when 9.5 billion people will need to live and be fed in the world in 2050 while the biomass will be used more as a sustainable energy element. This means that the pressure on the ecology will increase enormously and that we must think about how to manage the ecology in a sustainable manner despite this increased pressures on the worldwide ecology.

5.11. Organizational structures

The organizational structures to solve problems around the agro-forest-environment complex are more than a hundred years old and do not in many cases fit anymore the needs of today. We see more and more how nature works in an integrated manner but have nevertheless still the old structures in place. As an example for a small integrated problem we need 5, 10 or even more departments of the governmental structure to be able to solve it. On the one hand these models are too expensive on the other hand decision processes are too slow and produce inefficient results. Innovation and technology will help to do things better and must be used in the future but there needs to be in parallel the creation of better integrated structures and decision processes including the need to involve farmers and foresters much more for defining the needs for environmental caretaking and nature-based risk management handling as well as the solutions.

5.12. Benefits of Technology Application

Future farms working with the technologies will benefit through the following:

- GIS and images and GPS/dGPS (Galileo, GNSS) give detailed sizes of the fields or forest elements as base for exact calculations.

- GIS and images and GPS give the exact location of the field or forest element area for later logistics use that will optimize the routes, reduce petrol use and in parallel reduce CO₂.

- Farm management systems with integrated GIS capabilities and access to expert information in structured or unstructured way will allow much better planning and also documentation as well will allow easy calculation of variations, helping to choose optimized models.

- Farm management systems alongside the business calculation will imbed all activities with their costs. They will also allow detailed calculations of nutrient- and even CO₂- balances - based on expert data that are also standardizes by legal frameworks.

- The subsidy tools – if subsidies are available in the future – this would be in a totally different direction – as they must be unlinked to a farm management tool. Outputs are directly sent to the subsidy server of the Ministry as it makes no sense to enter data twice. Both systems, old (subsidy server on its own) and new (subsidy server filled by management systems or run parallel for other clients on its own), will run in parallel.

- Forest management will be based on easy to use and GIS based forest management systems that are part of a forest inventory system with country specific growth tables, a possibility to model the growth of the trees, to
work out local growth tables with the purpose of managing the forest for optimized wood AND optimized environmental targets. Environment and risk management will become forestry services.

- Farm- and/or forest- management systems will allow automatic embedding of the data into traceability (legally required) systems like EC or US traceability laws or also in private sector systems like Global GAP. Beyond traceability these systems will also serve in meeting legal requirements regarding sustainability (e.g. EC biomass regulations).

- All these technologies will allow to setup advisory services in a manner to enable smallholders to benefit from the technologies.

- Logistic tools for fleet management serving complete regions or countries will significantly reduce costs, petrol use and CO₂ output. They will include as beneficiaries the complete chain and not only farmer’s alone. The benefit will be also shared by the chain partners.

- The integration of detailed meteorological data will allow better decisions and will optimize fertilization and also spraying and will also reduce costs as well as environmental impacts. As well this will be a key contributor to risk management and insurance purposes

- Precision farming as overall technology will be used more broadly when the integration of technologies is ready and due to that the benefits will increase significantly. For large farms it will be used, nevertheless for smallholders a network of service providers will also enable them to use these technologies. Bank and insurance companies will be integrated as also they need information out of the farm management systems like business plans or insurance data.

- Technologies like ISOBUS or CAN-Bus will push precision farming but we also have to understand that for doing precision farming we have several options: To use old machines and have a better map and a PC that shows during driving where it is – the reaction can be done manually; you can have an old tractor with a new fertilizer equipment and a ruggedized field-book computer on the tractor that gets the map via a service provider and via GPRS/UMTS communication directly on the field book that can trigger the fertilizer equipment or the sprayer directly without having the tractor interfaced; a third variant only will be to uplink the new tractor with ISO- or CAN-Bus interface that transfers its petrol use to the field book. The PC field-book will be the main intelligence on equipment that triggers all the processes from getting information via communication, process this information and distributes it to the accessory equipment and gets back information about what had been done and send this information back via a server for billing also to the management system.

- Advisors or even regional platforms will allow service providers to understand better the regional capabilities and needs of output and inputs.

- Environmental caretaking needs, supported by clever technologies will allow the farmers to be integrated into these targets and do environmental caretaking as a planned service for beneficiaries from the public or private side.

- Trust Centres will allow to integrate farmers or foresters data, only under legal aspects or within bilateral agreements with chain partners, into these centres and will allow regulated access to these data. Such trust centres will also be used to build up farmers management portals to enable them to come closer to his customers.

- GIS and applications will allow doing faster and better land consolidation for complete countries with the benefit for smallholders to optimize the structure of their fields.

6. Questions

1. What are farmers using, considering, wanting?
2. Who are the actors in knowledge dissemination?
3. What are the main components (areas) of the agriculture and agri-food systems?
4. What are the functionality of Farm Management Information System?
5. What purposes crop system models are used for?

6. What are the main important technologies of precision farming?

7. Describe the main important features of innovation in ICT in agriculture!
16. fejezet - Bibliography


Food Quality and Safety Standards. 2007. Margaret Will, Doris Guenther. Technische Zusammenarbeit (GTZ) GmbH. Eschborn, Germany

Gelb E., Offer A. 2009 ICT in Agriculture: Perspectives of Technological Innovation


Tomas Milford and Karel Charvat jr. (eds.). 2012. ICT for Agriculture, Rural Development and Environment. Where we are? Where we will go? Czech Centre for Science and Society Wirelessinfo

