Introduction to Database Systems

Responsible persons: Stephan Nebiker, Susanne Bleisch

Table Of Content

1. Introduction to Database Systems

data management and especially the management of geodata is not bound to a specific technology. It would be possible to use analogous map archives or file based record systems. However, the term Geoinformation System implies some demands that exceed the storage and retrieval of data. These additional needs can be satisfied sensibly with database systems only.

This lesson is focused on database concepts and architectures. After an introduction and the definition of some of the most important terms in the unit **Definition of Terms** we will devote ourseleves to the specific characteristics of the database approach (unit **Characteristics of the Database Approach**). The closer examination of various applications of databases in the unit **Example Applications** will allow you to become acquainted with the use of databases in different fields and contexts and to extend the knowledge about the characteristics of databases. A comparison of the database approach with file based solutions is made in the unit **Advantages and Disadvantages**.

Learning Objectives

- Be used to the terminology of data, information and data management and be able to explain the most important terms.
- Know and understand the characteristics of database systems and be able to transfer these to applications in your daily life.

1.1. Definition of Terms

From Data to usable Information...

Before we start with the use and architecture of data management solutions we discuss the terminology of this field to get to know the basic terms like information, data, database systems. A lot of these basic terms are used daily but most often not in the right context.

1.1.1. Data versus Information

Data (especially computer data):
> The presentation of facts, information or concepts which are created in a computer readable form or are translated into such a form.

Information:
> Information is a useable answer to a concrete question. (ZEHNDER 1998)

Something is information if a specific question is answered and that answer increases the understanding of the questioner and enables them to come closer to a specific objective. (Translation from (ZEHNDER 1998))

Information has the following aspects:

- *structured* [3] and *syntactic* [4]
- *semantic* (as regards content)
- *pragmatic* (relevant to applications)

Relationship between Data and Information

The terms data and information are often used interchangeably and in the wrong context. Therefore a list of distinguishing features is presented below

- Semantic aspects of data are often coded. These codes need to be defined and interpreted after conventions previously agreed upon (e.g. Grades from 1 to 6 with the convention 6 = very good).
- Generally, information needs to be reconstructed or derived from data (e.g. the average rainfall of the month July over the last 10 years).
- Normally, data do not contain aspects relevant to applications (e.g. it is not possible to derive information for applications like tax, developement, flood risk, etc. from the coordinates of a parcel of land).

---

[3] IIn the information theory three dimensions of information are distinguished: the syntactic, the semantic and the pragmatic. Let us take the example of a traffic light. In the syntactic dimension we differentiate the three colours red, yellow and green. But the traffic light makes more sense in the semantic dimension. In this dimension the colours are linked to meanings. Red means stop, green means go. However, only in the pragmatic dimension does the traffic light become useable for the traffic. Pragmatically, red means that the driver of a car must stop.

[4] Syntax can in linguistics be described as the study of the rules, or "patterned relations" that govern the way the words in a sentence come together.

我　方　去
讲真相劝番世
写下的标

2002年6月13日　星期

觉醒后的肺腑之

亲爱的先生，

四年多以前，当您来中国访问的时候，中华人民共和国广播
事业局和马里国家

悼艹嫌ó间签订了第一个广播合作协定，为中马两国在广播
方面进行友好合作创造了良好的条件，这对于增进两国
朋穊间的相互了解和发展两国人民之间的友谊是有益

谢谢你的注意．问候！

Trying to read a letter in a foreign language, we are able to recognise the structure and the syntactic aspects of it like the paragraphs, sentences, words, etc. but we cannot make out the meaning of the writing.

However, writing in our own language, cannot be called information in every case. At least we might be able to understand the content (*semantic*) but if this content is irrelevant or not interesting to us then the important aspect of the usefulness is missing.

A letter in chinese language

1.1.2. The Components of an Information System

Conceptually [5] an information systems has a layered structure.

Move your mouse over the terms of the following interaction and get to know what parts make up an information system.

Only pictures can be viewed in this version! For Flash, animations, movies etc. see online version. Only screenshots of animations will be displayed. [link]

The data management components of an information system are:

Data and Database:

An amount of data which are viewed by the operator as somehow going together plus additional data which is used by the DBMS to work correctly are called a database.

Database Management System:

---

[5] A concept is an abstract, universal idea, notion, or entity that serves to designate a category or class of entities, events, or relations. Concepts are abstract in that they omit the differences of the things in their extension, treating them as if they were identical. They are universal in that they apply equally to everything in their extension. Concepts are also the basic elements of propositions, much the same way a word is the basic semantic element of a sentence.

A database management system (DBMS) is a software product for the persistent, consistent and application independent storage and management of data. But also for the flexible and easy use of big and concurrently used databases.

Database System:

A database system (DBS) consists of a DBMS and one or more databases.

☞ Database management systems and database systems are in the focus of this module.

Information System:

An information system extends the database with a couple of software tools for querying, presenting, transforming and analysing the data.

☞ According to the first part of this unit where the **difference between data and information** According to the first part of this unit where the difference between data and information were discussed, the tools of an information system enrich the data with *semantic* and *pragmatic* aspects.

For sure, you have already heard the term geoinformation system and perhaps read the one or more definitions of it. The following paragraph defines the term geoinformation system and compares this definition with the one of an information system as discussed earlier in this unit. Geoinformation System:

Geoinformation System:

"*A geoinformation system allows capturing, storing, analysing and presenting of all data that describe a part of the earth's surface and all on this part located technical and administrative equipment but also geoscientific, economic and ecologic features. (Translation)*" (BARTELME 1989)

This definition contains the most important aspects of the definition of an information system but focuses on data with spatial referencing.

1.2. Characteristics of the Database Approach

The database approach has some very characteristic features which are discussed in detail in this unit.

In the unit **Example Applications** the use of databases in different fields is presented and according to these examples the most important features of the database approach revised.

A comparison between the file based approach and the database approach can be found in the unit **Advantages and Disadvantages**.
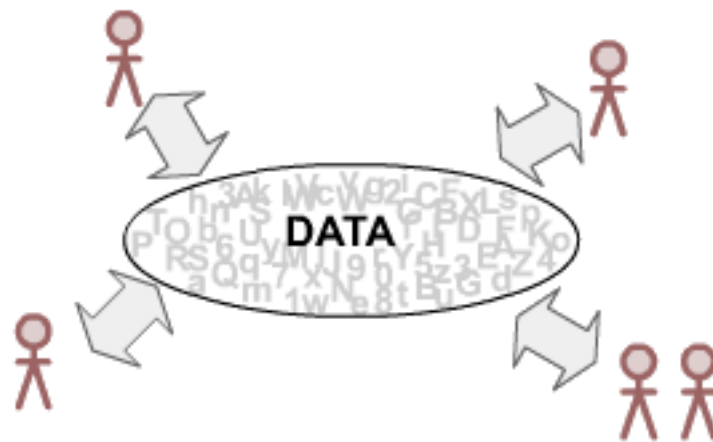
1.2.1. Concurrent Use

A database system allows several users to access the database concurrently. Answering different questions from different users with the same (base) data is a central aspect of an information system.

Such concurrent use of data increases the economy of a system. Data capturing and data storage is not *redundant* [11], the system can be operated from a central control and the data can be updated more efficiently. Additionally, better use of the most often very expensive *(geo) data* [12] can be made.

When using data concurrently the problem of how the system should behave if changes are made simultanteously (e.g. two different users with different applications change the same data simultaneously) needs to be solved. Additionally, there is a serious security risk, for example, in the realms of data protection.

In technical jargon changes to a database are called transactions. This term is explained later in this lesson.

An example for concurrent use is the travel database of a large travel agency. The employees of different branches can access the database concurrently and book journeys for their clients. Each travel agent sees on his interface if there are still seats available for a specific journey or if it is already fully booked.

Concurrent Use of the same Data

---

[11] Redundancy, in general terms, refers to the quality or state of being redundant, that is: exceeding what is necessary or normal, containing an excess. This can have a negative connotation, superfluous, but also positive, serving as a duplicate for preventing failure of an entire system.

[12] Geodata or data with a spatial relation are data about objects that through a position in space directly or indirectly can be referenced. The space is defined through a coordinate system which is in relation to the earth's surface.

1.2.2. Structured and Described Data

A fundamental feature of the database approach is that the database system does not only contain the data but also the complete definition and description of these data. These descriptions are basically details about the extent, the structure, the type and the format of all data and, additionally, the relationship between the data. This kind of stored data is called metadata ("data about data").

Metadata [13] is used by the DBMS software but also by applications like GIS and by users of databases. As DBMS software is not written for one specific database application the *metadata* of a database is used to get information about the extent, the structure, etc. of it.

Structured Data:

   Data is called structured if it can be subdivided systematically and linked.

Following is a simple example how data can be described in a database.

Following is a simple example how data can be described in a database. Below there is a database table. Because of the structure of this table (first column = Forename, second column = Surname, third column = Postcode, forth column = City) it is known that a entry in the first column must be a forename (coded as string) and an entry in the third column must be a postcode (coded as number).

Prename [string]	**Name** [string]	**Postcode** [number]	**City** [string]
Klaus	Meier	5369	Bühlwil
Sabine	Klett	4812	Moordorf
…	…	…	…

*Example of an Database Table*

1.2.3. Separation of Data and Applications

As described in the feature **structured data**, the structure of a database is described through *metadata* which is also stored in the database.

A software application does not need any knowledge about the physical data storage like encoding, format, storage place, etc. It only communicates with the management system of a database (DBMS) via a standardised interface with the help of a standardised language like SQL. The access to the data and the metadata is entirely done by the DBMS.

In this way all the applications can be totally seperated from the data. Therefore database internal reorganisations or improvement of efficiency do not have any influence on the application software.

---

[13] Metadata is literally "data about data", is information that describes another set of data. A common example is a library catalog card, which contains data about the contents and location of a book: It is data about the data in the book referred to by the card. Other common contents of metadata include the source or author of the described dataset, how it should be accessed, and its limitations. Another important type of data about data is the links or relationship among data.
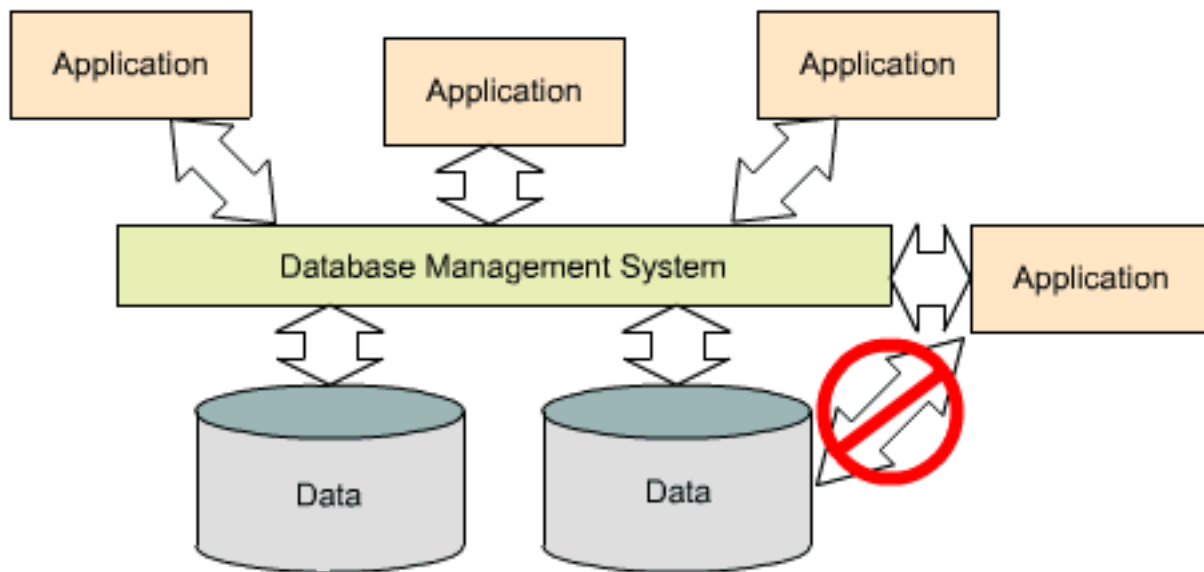
Separation of Data and Applications

1.2.4. Data Integrity

Data integrity is a byword for the quality and the reliability of the data of a database system. In a broader sense data integrity includes also the protection of the database from unauthorised access (confidentiality) and unauthorised changes.

Data reflect facts of the real world. Logically, it is demanded that this reflection is done correctly. A DBMS should support the task to bring only correct and consistent data into the database. Additionally, correct **transactions** ensure that the consistency is maintained during the operation of the system.

An example for inconsistency would be if contradictory statements were saved in the same database.

Wohnstrassen

Strassenname	Fahrverbot
Lindenstrasse	Nein
Rosenstrasse	Ja
Gartenstrasse	19.00–7.00
…	

Strassenplanung

Strassenname	Art	Einschränkungen
Hauptstrasse	Kantonstrasse	keine
Gartenstrasse	Quartierstrasse	keine
Jägerweg	Quartierweg	Fussweg
…		

Two Database Tables with Contradictory Datasets

1.2.5. Transactions

A transaction is a bundle of actions which are done within a database to bring it from one consistent state to a new consistent state. In between the data are inevitable inconsistencies.

A transaction is atomic, which means it cannot be divided up any further. Within a transaction all or none of the actions need to be carried out. Doing only a part of the actions would lead to an inconsistent database state.

One example of a transaction is the transfer of an amount of money from one bank account to another. The debit of the money from one account and the credit of it to another account makes together a consistent transaction. This transaction is also atomic. The debit or credit alone would both lead to an inconsistent state. After finishing the transaction (debit and credit) the changes to both accounts become persistent and the one who gave the money has now less money on his account while the receiver has now a higher balance.

Try it using the buttons at the bottom to the left which allowing navigation through the steps of this example.

Only pictures can be viewed in this version! For Flash, animations, movies etc. see online version. Only screenshots of animations will be displayed. [link]

1.2.6. Data Persistence

Data persistence means that in a DBMS all data is maintained as long as it is not deleted explicitly. The life span of data needs to be determined directly or indirectly be the user and must not be dependent on system features. Additionally data once stored in a database must not be lost.

Changes of a database which are done by a **transaction** are persistent. When a transaction is finished even a system crash cannot put the data in danger.

1.2.7. Data Views

Typically, a database has several users and each of them, depending on access rights and desire, needs an individual view of the data (content and form). Such a data view can consist of a subset of the stored data or from the stored data derived data (not explicitely stored).

A university manages the data about students. Beside matriculation number, name, address, etc. other information, for example in which course the student is registered, if he needs to do a resit, and so on is managed as well.

This extensive database is used by several people all with different needs and rights.

Please click on the four buttons below to see the different data views for different users of this database.

Only pictures can be viewed in this version! For Flash, animations, movies etc. see online version. Only screenshots of animations will be displayed. [link]

The database administrator has a view on the whole database while other users in this example have only a restricted view on the database. The administrator, for example, does not need information about names and matriculation numbers in case they want to create an anonymous statistic about the resits. In the student lists there should not be any sensitive data about the resits or similar. However, the lecturer of a class needs the detailed information about the students in this class including resits and so on.

1.3. Example Applications

Current information technology solutions distinguish themselves by having many distributed users which want to concurrently use constantly updated data. Therefore, database systems are used in different fields and there is rarely an information technology solution that gets by without one.

Following a couple of examples for the use of database systems are discussed.

For each example the application is breifly described first and then the for this application specific database features are commented on. The respective features with their descriptions can be found in the unit **Characteristics of the Database Approach** which is linked from the keywords.

1.3.1. Management of Bank Accounts

The management of bank accounts are a demanding task which has used database systems as aid for some time. Today, it is unthinkable to operate in the very complex world of finances without the help of database systems. The most important characteristics of such database systems:

|	
Transaction	The successfull and correct course of transactions is very important when managing bank accounts. It cannot be true that a credit is made to the wrong account or that a debit is made more than once.
Data Integrity	Data integrity is very important. It needs to be clearly defined what the requirements and rules of consistency are and how these can be followed.
Data Persistence	For an owner of a bank account it is reassuring to know that the persistence of the data is guaranteed. Data are not deleted arbitrarily or are lost mysteriously.

1.3.2. Timetable Informationsystem

The online timetable of the SBB (Swiss Federal Railways) is an example of a web based information system founded on a database system from the bounds of public transport. Its main task is to inform the users with current and correct information about the best connections and the train operation of the SBB at any time. The most important characteristics of such a database system:

Concurrent Use	The database system of the SBB timetable can be used by different users and applications concurrently. While the information are used internally within the SBB, information about the timetables and other specific information can also be used from the public via the internet.
Data Integrity	The users of the SBB timetables need current and correct information at any time. This makes great demands on the data integrity. Therefore, timetable or platform changes or any other changes need to be updated constantly in the database system.

Such a database system can get more than one call per second in record times. Therefore, perfomance is also a very important characteristic which was not mentioned above.
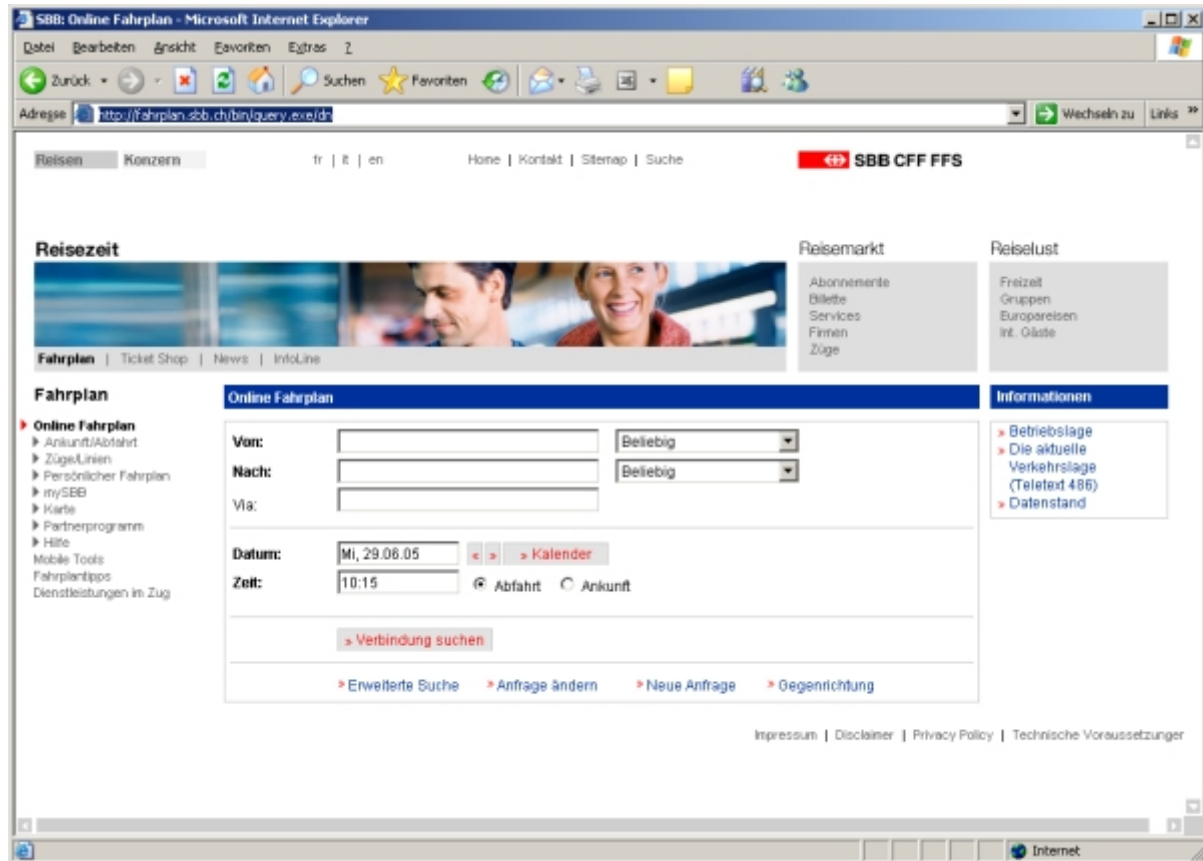
Screenshot SBB Timetable Informationsystem (SBB AG 2005)

Internet access to the SBB timetable: **http://fahrplan.sbb.ch/**

1.3.3. Library Catalogue

The library catalogue NEBIS (Lierz et al. 2004) is the catalogue of a network of libraries and information desks in Switzerland. With the help of NEBIS it is possible to search for specific books or publications in libraries all over Switzerland.

The most important characteristics of such a database system:

Structured and Described Data	It is more than helpful to have a clearly defined structure when recording and updating several thousands of books, magazins and publications. Additionally, the description of the data allows to search selectively for specific objects. Further, a clear structure is also free of redundancy. This saves work as already a very small *redundancy* would lead to a multiple of work.
Data Views	Depending on how detailled and specialised a search should be done, a user needs more or less information from the database. Thus, for a coarse search title and author might already be sufficient. However, when looking for a specific edition of a book, some more information is needed. With this in view the user can choose between different data views. Additionally, there are views for the manager of the catalogue which are not accessible by the public user.

Different Data Views (click on the thumbnail for a bigger image)

Result List of the Search (Lierz et al. 2004)

Standard View of a Choosen Entry (Lierz et al. 2004)

*Catalogue View of a Choosen Entry* (Lierz et al. 2004)

'MARC'-View of a Choosen Entry (Lierz et al. 2004)

Screenshot of a Library Catalogue (Lierz et al. 2004)

Online access to the NEBIS library catalogue: **http://www.nebis.ch/**

1.3.4. Central Geodata Warehouse

Recently it has become more common in city and canton governments to manage and update spatial data of all parts of the government in one central data pool called geodata warehouse (a spatial database management system). This makes huge savings as the data is now stored redundance free and needs to be updated only once. Until now, some data needed to be stored and updated in different departements of the government. This made it enormously difficult to have current data in all places. Additionally, when doing bigger projects the data can now simply be taken out of the geodata pool and does not need to be gathered in tedious and lengthy work. The most important characteristics of such database systems:

Concurrent Use	A central geodata warehouse is a nice example of concurrent use of a database system. On the one hand different users get their data from it - the employees of the different departements. On the other hand different application software (e.g. GIS system) are used to access the geodata warehouse. Therefore, it is possible that the forestry departement uses GIS software A for access to the data while the surveyor departement uses GIS software B.
Separation of Data and Applications	As described above, different users' with different applications get access to the data. This is only possible, if the data is separated from the applications. If data were connected to the application it would be much work to process the data in way that other applications could read and use it. This independence is especially important in cases where the DBMS software needs replacement. With the separation of data and applications this is possible without having to re-write all of the application software. Example: A Microsoft Word file (where the data is included in the application format) is quite difficult to open with the Microsoft Excel software even though both applications are made by the same company.
Data Persistence	The capture of *geodata*: and other data is most often a lot of work and very expensive. Therefore, data persistence is a very important characteristic of a geodata warehouse. This way, it is possible to ensure that data is not lost and needs then to be replaced, which is costly.
Data Integrity	Governmental data often give information about legal conditions for example, the cadaster. Therefore, these data need to be thoroughly correct and reliable. That is achieved through the definition and following of specific consistency requirements and rules.
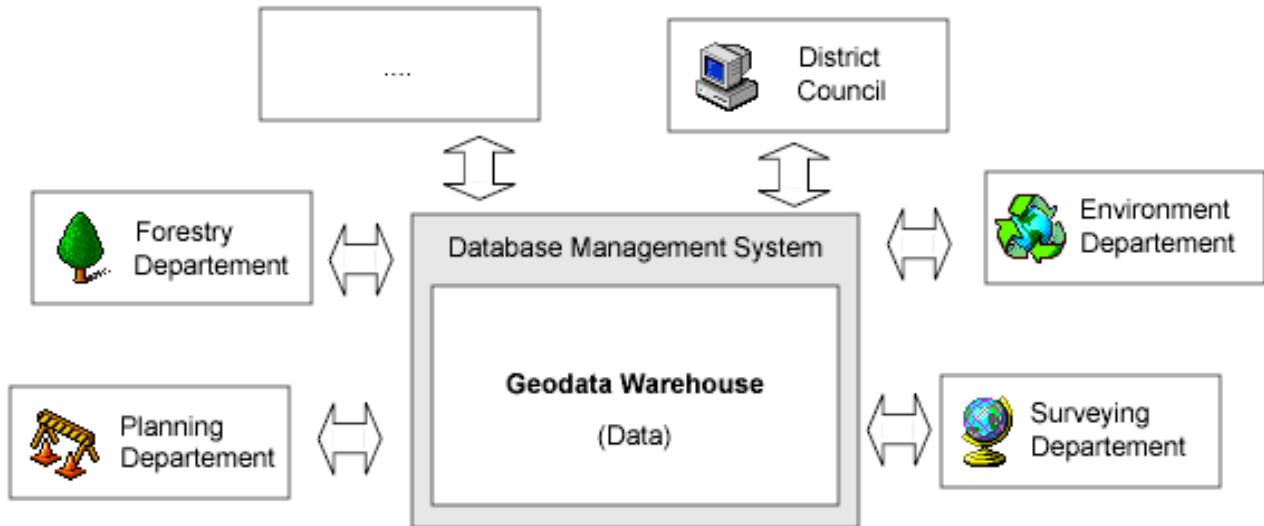
Schematic Representation of a Geodata Warehouse and the possible Access to it from different Offices

1.3.5. Exercise

Find another database application (your work environment, internet,...) and try to figure out which are the most important characteristics of that application.

Write a short summary of this information (like the ones you have seen in this unit) and post your writing to the discussion board under the topic 'Database Applications'. Look at and discuss also at the postings from other students.

A tutor will comment on your summary (also posted to the discussion board).

1.4. Advantages and Disadvantages

From the file-card box to the database

Management and storage of data has changed a great deal over the years - from the file-card box, via the first (file based) digital version, to the modern database systems. The first part of this unit deals with file based systems in comparison to database systems. Afterwards, the advantages and disadvantages of database systems are discussed.

1.4.1. Comparison DBS versus file based

Knowing about the characteristics of a database system (unit **Characteristics of the Database Approach** and unit **Example Applications**) we will have a look at file based systems. With a file based approach each user defines and creates with a specific software the files he needs for a specific application. In comparison to the database system approach this results in some limitations.

Move the mouse over the bold terms to the left and the limitations of the file based approach in comparison to the database approach are explained in writing (right) and graphics (below).

Only pictures can be viewed in this version! For Flash, animations, movies etc. see online version. Only screenshots of animations will be displayed. [link]

Using a file based system it is possible that, for example, the administration of a school maintains information about the registration and fee payment of the students (e.g. based on a spreadsheet software). Meanwhile the lecturers manage data about the students and their grades. Even though both user groups are interested in student data they both might have different files and different software to update and change these. Such *redundancy* in the definition and storage of data wastes storage space. Additionally, the work needed to update data in more than one system is multiplied by the number of systems. In case above a change of basic student data like the address might need an update in the system of the lecturer and in the system of the administration as well.

1.4.2. Advantages of a DBMS

Basically, all in the unit **Characteristics of the Database Approach** listed features can be listed as advantages here too.

- **Concurrent Use**
- **Structured and Described Data**
- **Separation of Data and Applications**
- **Data Integrity**
- **Transactions**
- **Data Persistence**
- **Data Views**

Additionally, there are some advantages which were not yet explicitly mentioned.
Use the blue buttons to navigate.

> **Only pictures can be viewed in this version! For Flash, animations, movies etc. see online version. Only screenshots of animations will be displayed. [link]**

Click on the following link to get a list of the further advantages. (**DOC file of the advantages**)

1.4.3. Disadvantages of a DBMS

Beside the numerous advantages of a database system the disadvantages should not be kept secret.
Use the blue buttons to navigate.

> **Only pictures can be viewed in this version! For Flash, animations, movies etc. see online version. Only screenshots of animations will be displayed. [link]**

Click on the following link to get a list of the disadvantages. (**DOC file of the disadvantages**)

1.5. Tasks

The following task should allow you to test if you have understood the content of this lesson.

> **Only pictures can be viewed in this version! For Flash, animations, movies etc. see online version. Only screenshots of animations will be displayed. [link]**

> **Only pictures can be viewed in this version! For Flash, animations, movies etc. see online version. Only screenshots of animations will be displayed. [link]**

Drag the numbers beside the terms onto the circles in the graphic. Place them as exactly as possible and click '>Check' when you have placed them all. Symbols will show you which numbers are placed correctly and which are on the wrong place. The '>Reset' button clears the graphic from all numbers.
Attention: The '>Check' button only works correctly if all circles in the graphic are covered with a number.

> **Only pictures can be viewed in this version! For Flash, animations, movies etc. see online version. Only screenshots of animations will be displayed. [link]**

1.6. Summary

In this lesson the significance of the data management based on database systems and the role of a database system within a geoinformation system (GIS) was explained. At the beginning, terms like database system and geoinformation system were defined and the terms data and information positioned in context.

Following important reasons for the use of database systems for the management of data in general and geodata in particular were given and illustrated with example applications of different fields of use. Included in the significant characteristics of the database approach are concurrent use as well as structured and described data. Additionally the advantages of the separation of data and applications and the features for high reliablitiy and data security like the concepts of transactions and data views were discussed.

1.7. Recommended Reading

- **ZEHNDER, C.A.**, 1998. *Informationssysteme und Datenbanken*. Zürich: vdf Hochschulverlag AG.

  Introduction into Information Systems and Databases, in German

1.8. Glossary

Conceptual:

A concept is an abstract, universal idea, notion, or entity that serves to designate a category or class of entities, events, or relations. Concepts are abstract in that they omit the differences of the things in their extension, treating them as if they were identical. They are universal in that they apply equally to everything in their extension. Concepts are also the basic elements of propositions, much the same way a word is the basic semantic element of a sentence.

Data (especially computer data):

The presentation of facts, information or concepts which are created in a computer readable form or are translated into such a form.

Data and Database:

An amount of data which are viewed by the operator as somehow going together plus additional data which is used by the DBMS to work correctly are called a database.

Database Management System:

A database management system (DBMS) is a software product for the persistent, consistent and application independent storage and management of data. But also for the flexible and easy use of big and concurrently used databases.

Database System:

A database system (DBS) consists of a DBMS and one or more databases.

Dimensions of Information:

IIn the information theory three dimensions of information are distinguished: the syntactic, the semantic and the pragmatic. Let us take the example of a traffic light. In the syntactic dimension we differentiate the three colours red, yellow and green. But the traffic light makes more sense in the semantic dimension. In this dimension the colours are linked to meanings. Red means stop, green means go. However, only in the pragmatic dimension does the traffic light become useable for the traffic. Pragmatically, red means that the driver of a car must stop. (Anonymous)

Geodata:

Geodata or data with a spatial relation are data about objects that through a position in space directly or indirectly can be referenced. The space is defined through a coordinate system which is in relation to the earth's surface.

Geoinformation System:

"*A geoinformation system allows capturing, storing, analysing and presenting of all data that describe a part of the earth's surface and all on this part located technical and administrative equipment but also geoscientific, economic and ecologic features. (Translation)*" (BARTELME 1989)

Information:

Information is a useable answer to a concrete question. (ZEHNDER 1998)

Information System:

An information system extends the database with a couple of software tools for querying, presenting, transforming and analysing the data.

Metadata:

Metadata is literally "data about data", is information that describes another set of data. A common example is a library catalog card, which contains data about the contents and location of a book: It is data about the data in the book referred to by the card. Other common contents of metadata include the source or author of the described dataset, how it should be accessed, and its limitations. Another important type of data about data is the links or relationship among data.

Redundancy:

Redundancy, in general terms, refers to the quality or state of being redundant, that is: exceeding what is necessary or normal, containing an excess. This can have a negative connotation, superfluous, but also positive, serving as a duplicate for preventing failure of an entire system.

Structured Data:

Data is called structured if it can be subdivided systematically and linked.

Syntax:

Syntax can in linguistics be described as the study of the rules, or "patterned relations" that govern the way the words in a sentence come together.

1.9. Bibliography

- **Anonymous**. Available from: http://www.foraus.de/download/aktuelles/hl_ausb_m_lernaufg.pdf [Accessed 31. März 2003].
- **BARTELME, N.**, 2000. *Geoinformatik - Modelle, Strukturen, Funktionen*. 3rd. Berlin: Springer.
- **BARTELME, N.**, 1989. *GIS-Technologie. Geoinformationssysteme, Landinformationssysteme und ihre Grundlagen*. Springer.
- **DITTRICH, K.**, 1999. Datenbanksysteme. *In:* **RECHENBERG, P., POMBERGER, G.**, ed. *Informatik-Handbuch*. Wien: Carl Hanser Verlag, 875-908.
- **Lierz, W., Scherer, E.** (2004). *Netzwerk von Bibliotheken und Informationsstellen in der Schweiz* [online]. Zürich. Available from: http://www.nebis.ch/ [Accessed 29. Juni 2005].
- **SBB AG** (2005). *SBB: Online Fahrplan* [online]. Available from: http://www.sbb.ch/ [Accessed 29. Juni 2005].
- **WARTH, D.**. Available from: http://www.fask.uni-mainz.de/inst/iaspk/Linguistik/Syntax/Was.html [Accessed 25. Oktober 2004].
- **ZEHNDER, C.A.**, 1998. *Informationssysteme und Datenbanken*. Zürich: vdf Hochschulverlag AG.