TECHNISCHE UNIVERSITÄT WIEN
VIENNA UNIVERSITY OF TECHNOLOGY

**DIPLOMARBEIT**

# Data Quality in Navigation Systems – A new approach to define User Groups

Ausgeführt am Institut für

**Geoinformation und Kartographie**

der Technischen Universität Wien

unter der Anleitung von
**o. Univ. Prof. DI Dr. Andrew Frank**

durch

**Marianne Herren**
Laudongasse 20/7 1080 Wien

17. September 2007

# Eidesstattliche Erklärung

Ich erkläre hiermit an Eides Statt, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.

Wien, am 17.09.2007

Marianne Herren

# 1. INTRODUCTION

The goal of this thesis is to develop an adaptable system schema for data quality. The question is what degree of quality is acceptable for which user group. This leads to a new approach to categorize user groups by their tasks and the resulting demands. Nowadays data collection and the exponentially increasing amount of data exhaust the actual technical limits *ad nauseam.* Solutions are usually searched in the fields of data compression and memory expansion or data merging. This thesis shows an approach to minimize data necessary to solve a problem satisfactorily to a certain user group. The hypothesis is that it is possible to reduce the amount of data and to degrade data quality to a predefined limit without losing the capability to perform necessary tasks for a certain user group. The reduction of data and data quality leads to a downsizing of storage and eases data handling. The performance of central processing unit and the complete system do not have to be adapted to the increasing data and requirements but rather the data is adapted to the previously discovered user needs and system capabilities. This process enables a reduction of production costs for the developer of a Geo Information (GI) tool. The reduction of costs can be passed over to the price for the user. New users can be attracted by lower prices and new classes of population with different net income can afford GI products. This leads to the conclusion that a degradation of data quality has positive effects on the possible distribution of GI products by attracting new user groups on certain conditions.

## 1. 1 Motivation

The primary goal is to sell GI products using an optimized sale strategy. User requirements have to be analyzed to develop a reproducible decision making process. With this information it is possible to adapt the implementation and determine whether to generalize or to specify. Depending on the usage of the application, usability for the user becomes predictable, and thus influences the choice of used methods in the model. Legal and physical restrictions are defined in the used ontology and must be considered to assess usability, which helps in dealing with uncertainty and for basic measurement. The main aspect in this thesis is to define data quality from the user point of view. The resulting rules should be as simple as possible to ease error specification and evaluation of consequences. Based on the scenario of a street network with one ways represented by driving a car in a city, the impacts of the developed models are verified. The required adaptations to the changed situation can be offered to the customer as an extension to enhance user acceptance of the application. The critical aspect is the definition of usability for the user caused by the decision process, the resulting consequences for data quality, and the methods used in the application.

The basic idea is to merge data (Frank 2002b) with different quality and, depending on the user, leading to different outcomes. Each user has different demands to be met. The subtle distinction between the different demands determines the needed quality. A not necessarily complete presentation of nearly all details can be a solution to gratify a test user since some data are redundant, superfluous, and irrelevant for decision-making. Some data is redundant and not relevant for decision-making. Using cheaper and not so detailed data can reduce costs by finding new possible users and increasing the sales figures. Customers should know the real value of data and appreciate all the efforts behind the product, which can be achieved by showing them the relevant differences between the full and the preview-version.

The final question is about quality. It contains several aspects with varying importance to the user. Quality is a broad expression and should be used carefully. Users have different conceptions of quality; most often they imply only one aspect of quality, their preconception, and neglect others. The intention of this paper is to emphasize the user's demands and the resulting definition of quality.

## 1.2 Hypothesis

The hypothesis shall demonstrate that a deterioration of quality of data permits price differentiation. Data Fusion with noise or a merging of two or more databases is a possible way to degrade quality. A degraded dataset is of lower quality and can be priced lower. Price can be varied; different price levels enable people with different income to afford the product or data with small differences. The produced data is affordable for more people, and finally the sales numbers can increase by supplying more people with a variety of products.

The first goal is to define users and to generalize them into groups. This step is critical. All users are individuals and have different requirements and the decision process is conscious or subconscious and therefore not always traceable or reproducible. Users are also willing to pay different prices for the solutions caused by a different individual decision.

By grouping, the requirements must be assessed and weighted according to their importance. This step includes the analysis of the usage. Navigation system users for example can plan a trip through a city from point A to B, or include several points to visit on this route. This example shows two different possible usages of a navigation system. Each usage requires different metadata and quality of information for the user. These requirements are grouped in several packages and linked to the previous specified usage of the system.

The previously achieved information is used for a data fusion to generate data with deteriorated quality. This is the last step before offering non specialists an affordable solution and to resell the already collected data. Re-usage of data by producing different levels of quality is a challenging and promising marketing strategy to expand the market for GIS-products.

## 1.3 Structure

The Ontology of Navigation describes the used objects in the model. The supported objects are taken from a "Mehrzweckkarte" (multi-purpose-map) from Vienna. Another aspect of ontology is the actors and the problem they want to solve. This model uses a simple pathfinder or route planning model. An agent has to find a way from A to point B. The shortest way is not always the best, depending on the needs of an actor.

The next step is to define users and to classify them into groups. So-called "user-groups" have similar or the same user requirements. Of course not all requirements can be met permanently but it is possible to divide them into important and less important ones. In practice the user himself decides what his requirements to be met are and which can be neglected by choosing the degree of quality, functionality, and therefore the price.

The critical process is to define quality. Depending on the usage and user, quality has different meanings determined by the relevance for the user. Data quality parameters are of qualitative and of quantitative nature. These parameters have to be ranked using an analysis of relevance for the users. The ranking method used here is taken from Naumann (2002). Especially the weighting method is a simple approach to find priorities and to assess also qualitative parameters. The weighted quality parameters are called “Information Quality”.

Based on the previously retrieved information about the quality parameters a fusion between a database and noise is possible to produce deteriorated data. The assumption in this approach is that the resulting database is of different level of quality for the user group.

An important aspect is to maintain the consistency for the example of path finding. The accessibility from the starting point to the endpoint must be kept. In other words point A must have a relation to point B over other several mediator points and the relation must not have any break in between, if there was a connection before. The resulting constraints retrieved from the proven consistency are relevant for the data fusion and describe which relations must not be noised or changed.

After the fusion of data with noise the reliability of the system must be proven as shown in Birolini (1997b). The “pollution” of the data can have unexpected effects on the quality parameters. A comparison between the original database and the produced database is a restricted view because the different user groups are not included in this analysis.

One of the most important aspects of quality is usability. It is a degree of the achieved adaptation of the model for the user or a measurement of the mediator between user and the system: the model, the database, and the application.

“*Usability normally contains the following aspects: Learnability, Efficiency, Satisfaction, Memorability and Errors.*” (Nielson 1993 p. 26) Not only these five aspects affect the handling of a model, but Information Quality of the used data plays also an important part. A comparison between the original user requirements and the possible results provided by the new datasets will prove that the Information Quality of the new database is of high importance for the different user groups. (Nielson 1993)

The result of this fusion with noise will show what the possible rules, limits, and constraints to merge data with noise and to maintain a certain level of “Information

Quality" are. The degradation of the basic database is not principally a degradation of usability and quality for the user but can also help to price quality, usability, and finally data.

# 2. ONTOLOGY OF NAVIGATION

## 2.1 Basic Problem

Driving a car in a city is a common task. The case study used in this thesis is car navigation in the city of Vienna. Driving in a city includes orientation of the user's location, defining the destination and deciding at each intersection which way to choose.

One basic problem for the user is the huge amount and variety of information provided on maps. For example, a street consists of several elements, like sidewalk borders, hydrants, and so on, that are not always necessary to make a decision whereas green areas, sights, bank offices, rivers, and so on are important. The information is not structured and overlapping as shown in the next picture.

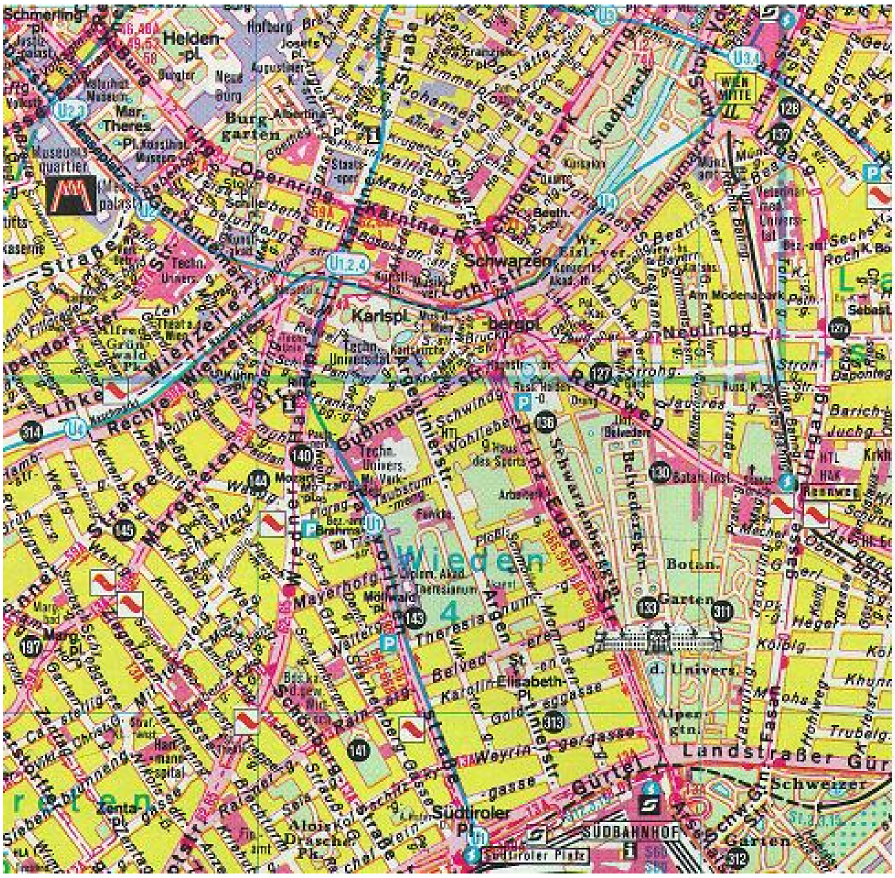


Figure 1: Street Map of Vienna with "unnecessary" information galore

Users might feel confused by the overwhelming amount of information. Most of the information they have is not needed for their decision process and complicates the overview that is necessary to make a decision. A driver only needs a small amount of basic elements to find his way from one point to another, from start to destination. These elements are basically points, lines, and areas. They are surrounding the initial position; describe orientations, and the target destination. It is critical to find a way to evaluate information and to grade according to the value for the user.

## 2.2 Model of Wayfinding

Human activities in real world are described in the spatial cognition research by Gluck (1991). Piaget (1954) and later Siegel and White (1975) showed that the stages in an individual's representation of spatial knowledge are likely to come with increasing age or experience. They set three levels of human spatial knowledge:

1. *Landmarks* are distinct, typically familiar points in the environment.
2. *Route knowledge* is characterized by the knowledge of paths between landmarks (topological information), but lacks general understanding of space (e. g. inability to recall description of the entire route from memory).
3. *Survey knowledge* means the proper understanding of spatial organization, that is, the ability to locate objects in terms of routes between them, using information about distances and directions.

The assumption from Raubal (2002) is that a user navigating in a city can be simulated by an agent. The agent of a system reacts like a real person. An agent perceives its environment through sensors and acts upon that environment through effectors (Russell et al. 1995). An agent has to simulate the user, his decisions and actions. The decision of a user is affected by his environment and leading to an action. The following figure explains the concept of an agent interacting with the environment:

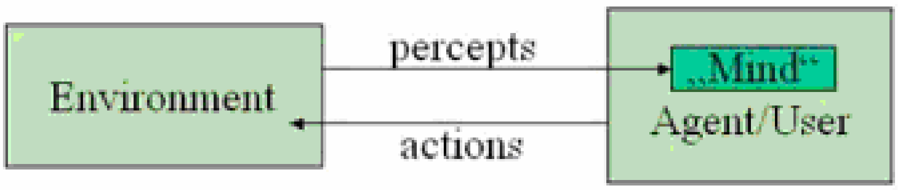


Figure 2: Interaction of the agent with the environment based on Raubal (2002)

The human mind perceives information and data from the environment. An agent reacts similar to a user and takes corresponding actions. These actions change the environment or the current situation. The agent/user perceives again the change in environment and reacts again accordingly. A direct conclusion from perception to action can be observed. The decision-making process of a user or, in this thesis, agent can be described by an Expert System. The user extracts useful information from the perceived input and prices the individual value of each bit of information by already known rules. These rules can be learned from experience or transferred from other users. The agent in the Expert System reacts in a similar way (Mazzetti, 2003). The difference is that the set of rules is predetermined. Expert Systems can be implemented as self-learning and self-incrementing systems using an artificial intelligent algorithm to learn new rules or adapt old ones. Retrieved information has to be evaluated to determine the importance and relevance for the task.

The next question is how to describe the environment. What is necessary for a user or an agent to make a decision in a certain environment? An agent shall behave like a real person and has to consider the same environment and set of rules. Thus the perception of the environment has to be observed and generalized for the resulting ontology.

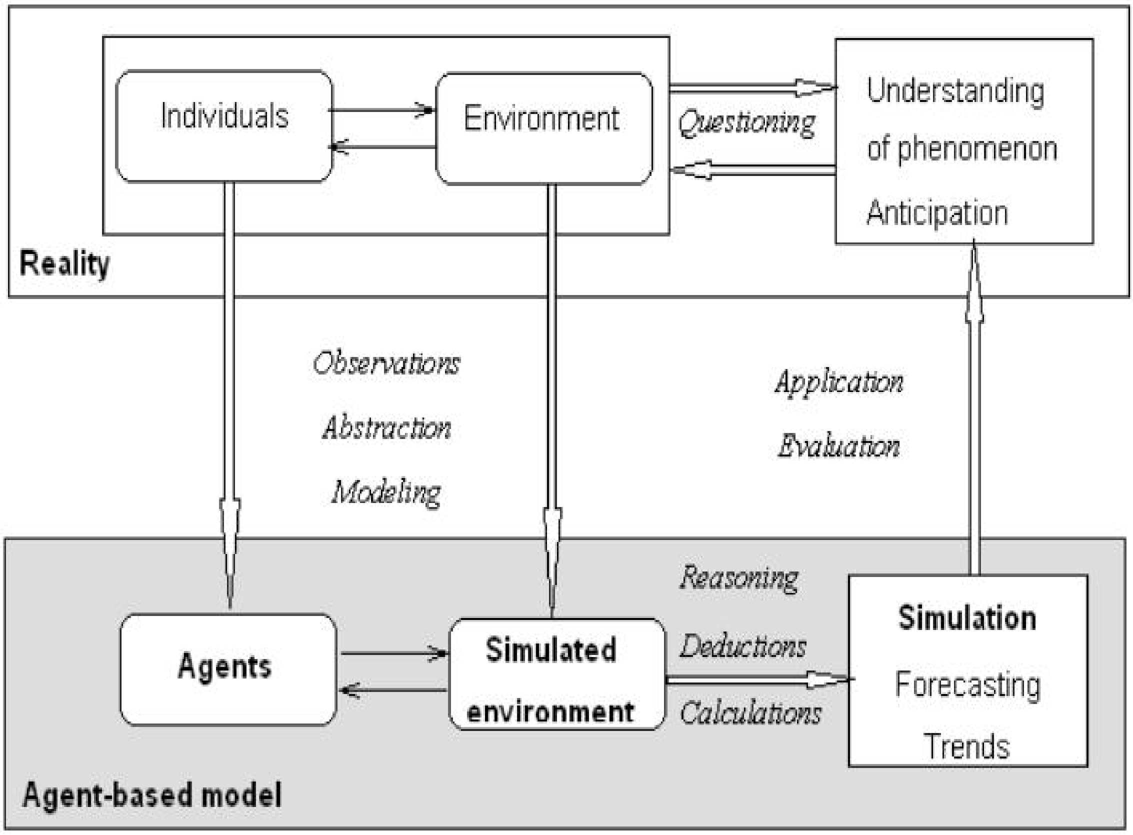


Figure 3: Agent-based model and reality (taken from Krek (2002))

The work of Lynch (1960) was influenced by Piaget and Inhelder and researched sketch maps and path descriptions of people in different cities. He found out that people build a cognitive map or a mental model of a city with five distinct elements. These elements are:

1. *landmarks* as distinct points in a city that serve as reference points to the observer,
2. *paths* describing streets or lanes,
3. *nodes*, located along paths like bridges, intersections, etc.
4. *edges* showing the boundaries of areas and forming physical barriers like rivers, and
5. *districts* describing areas in cities that have a common purpose (residential or industrial areas).

These five elements are sufficient for a person to orientate and to navigate in a cognitive way in cities. Additional information is mostly not requested and only confusing during the navigation process. The elements described above can easily be compared to the spatial objects used on maps.

## 2.3 Spatial objects

The most common definition of geometric primitives depends on the concept of dimension. Dimension is based on a geometric characteristic, that is, length. Classified according to their dimension we can compare the previous elements and order them:

- *Point*: has no length and is of 0-dimension.
- *Line*: has only the property of length and is of 1-dimension.
- *Area*: takes length to the second power and is of 2-dimension.
- *Volume*: takes length to the third power and is of 3-dimension.

These four spatial objects conclude the elements of a map. The n-dimensional geometric primitive has direct physical correspondences to the objects in space. Depending on scale a district in a city can be represented as a point or an area.

## 2.4 User

High quality information is not necessarily an improvement for a user if it is more detailed. It can complicate usage and can cause misunderstanding. Nielsen (1993 pp. 120-123) argues in his chapter about “Less is More” that too much information “…*can distract the user from the primary information. Based on a proper task analysis, it is often possible to identify the information that is truly important to users and which will enable them to perform almost all of their tasks. …Extraneous information not only risks confusing the novice user, but also slows down the expert user.”* This theory is proven by a study of telephone operators who have to find the important information on different screens as fast as possible. Nielsen applies this rule to the information content of screens, features, and interaction mechanisms for a program.

In the last years a race for the “highest quality” information started. Many providers claim to have the best datasets. A user might need parts of data from one provider and from another. Two databases can consist of different metadata. Structure and Content can differ too. The critical factor is to minimize and eliminate incompatibility problems between two different databases (Fournier, 2003). Research is still under way to ease interoperability of metadata with different quality.

Data sharing is a basic principle in our world. The network requires adaptable standards of principles, frameworks, and standards of levels of quality and characteristics. Concluding generalizations are difficult to perform and to justify. Participants of data sharing with a framework are classified as data producers, data integrators, and data users (Dueker and Butler 2000, 13-73). The three groups have definable tasks and use databases in different ways.

A generic customer supplier model is displayed in a paper by Hauser and Clausing (1988, 63-73). Data integrators and data producers are aggregated to the term supplier in the Figure 4. The suppliers and their decision process are excluded in this paper, although the supplier plays an important role in the communication process. The model proposed here focuses externally on the customers and their requirements.

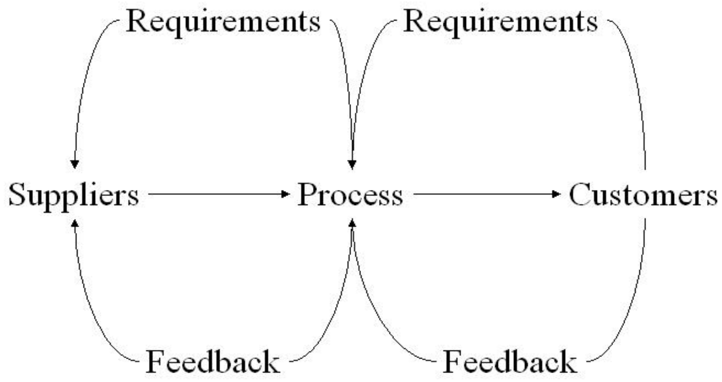


Figure 4: Customer-supplier model taken from Hauser et al. (1988).

The supplier provides data that is critical for the process based on certain requirements; the process produces data for the customer. The customer has requirements that have to be met by the process. If the requirements are not met, the customer gives negative feedback to the process. The process has also certain requirements and the possibility to return feedback to the supplier. This loap can improve the basic system by each feedback cycle if an adaptation increases the satisfaction of the requirements.

High quality defined by a data producer can be useless or even undesirable for the user. Depending on the purpose a large dataset with a large amount of metadata can complicate implementation or usage of an application. Time is also a critical factor in software assessment. Depending on the importance of the metadata and level of quality characteristics a distinction can be difficult. Users have to be classified in more detail and their requirements examined carefully. According to this retrieved information about users, Expert Systems can be developed and adapted. With Expert System user behavior, acceptance and satisfaction can be simulated and new information to improve the System can be integrated. Users have different requirements and rank quality according to their expectation of their needs and how they are met by the system.

## 2.5 Information Centered Approach

Focusing on the user is a new approach to Geo Information System (GIS) technology. A GIS produces information for a user. Information must be suitable for the user and satisfy certain requirements. The first steps towards a successful introduction of the GIS are to answer following questions (Frank 1999):

- Defining the tasks of the involved users and the new additional information the GIS shall produce and display.
- Finding the necessary information for the previously defined tasks.
- Fixing the presentation form of the information that is easiest to understand for the user.
- What information is needed depends on the tasks of the users. The goal of a GIS system is to improve the decision making, to speed up decisions, and to justify the usage to solve a task.

In this hypothesis it is important to mention and to investigate the basic explanation of an organization, where the user is involved. A task a user has to perform is determined by his role in the organization. An organization can be described as (Vliet 1988):

> *"A formalized cooperation of people and means in order to attain a certain goal. An organization also has surrounding, which is everything external which influences the organization or is influenced by the organization."*

The concepts of data information are used interchangeably in the literature, but in this hypothesis they have different meanings. The distinction is taken from Bots and Jansen (1989):

> ***"Data** are the objectively observable expressions of facts or knowledge on a certain medium (e. g. on paper or disk)."*

> ***"Information** is data that have been mutually related and interpreted"*

Data is the basic value used in a task. The datasets are inserted in the application, used, and changed to perform an outcome. This process can be a calculation, an interpretation or transformation of the basic data. The goal of every application or, basically, transformation process is to produce information. Information is in this thesis as Frank states:

> *"Information is an answer to a question by a human. This implies content, relation and interpretation."*

Information is the produced goal of an application using certain data, and represents the outcome of data that can be understood and interpreted by a user. The difficult part is to study users and to identify the tasks they want to perform. This analysis centralizes the choice of functionality of the application and helps to predetermine the degrees of quality needed by the user.

## 2.6 Users and their Decisions

### 2.6.1 User Profiles and their Characteristics

The first step in finding user profiles is to determine who will use the planned product—in this situation using data. Afterwards the description of the whole user population is obtained in terms of user characteristics (Mayhew 2002). These characteristics include:

- *Psychological* characteristics (e. g., attitude, motivation),
- *Knowledge* and experience (e. g., typing skill, task experience),
- *Job and task* characteristics (e. g., frequency of use, task structure), and
- *Physical* characteristics (e. g., color blindness).

It is possible to determine user characteristics by gathering data by interviews and/or user profile questionnaires. Conclusions are drawn from the summarized data high–level regarding further decisions and requirements. Each user profile has to summarize a significant category of users within a task category.

Another approach describes a distinction between environmental and individual influences possible (Czinkota 2001). The environmental influences are most often subconscious, while the individual influences can be directed by the person and every individual determines to himself the degree of consequences. The distinction between conscious and subconscious is not easy because every user or person has its own level of awareness and the decision process is individual.

*Environmental influences:*

- Culture
- Social Class
- Personal influences
- Family
- Situation

*Individual influences:*

- Consumer resources
- Motivation and Involvement
- Knowledge
- Attitudes
- Personality, Values, Lifestyle

One of the most underestimated factor or influence is the experience of the user. Each user can rely on the reputation given from other users but has to decide himself if a tool, application, or solution is “fit for use” for him. Fitness for use is an expression often used in the field of usability and describes the perfect adaptation of a product for a certain usage and person who wants to perform a task.

> *Experience is a much undervalued factor, particularly by those who believe that creative marketing is all it takes to move a product. The consumer usually takes practical experience into account before any other factors. If the product is not liked, no amount of advertising will succeed. Experience will of course change over years. Consumers learn from their consumption decisions and adjust their expectations and behavior (as a result, a certain approach that once worked does not necessarily keep on working.) (Czinkota 2000)*

It will be demonstrated in the fifth chapter that the decision process in navigation is the key factor in this thesis. The division into relevant and non-relevant information leads to relevant and non-relevant sub-decisions. The information offered must be carefully analyzed concerning suitability. The consumer, in this case the user, decides how much attention to devote to the communication, based on the degree of interest or need for the information.

Sometimes users decide to ignore information because they have developed better alternatives in the past. Especially in navigation users proceed in learning and coping with the system and new problems. For car navigation the following experience is described in the following figure.

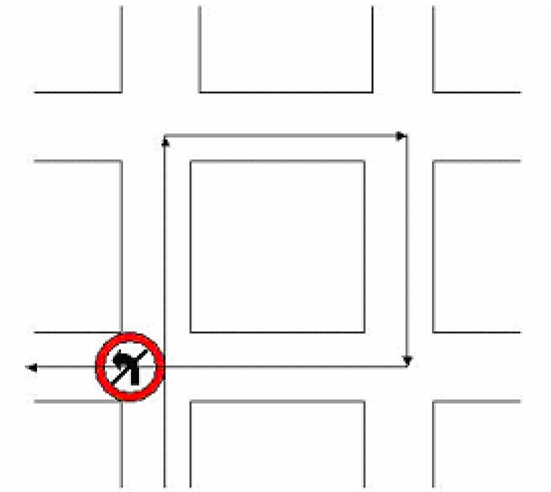

Figure 5: Example of left turn: You want to turn left, but it is prohibited. One possible and usual way is to try right turns to reach the final direction.

Information about user behavior must be retrieved, collected, and analyzed. Computerized databases can handle the complete process of collecting and distributing information as one possibility. That system is called MIS (Market Intelligence System). The collected and organized data are evaluated and divided into different levels of relevance. These categories are helpful for further planning, implementation, and control of the system. The MIS System adds value, transforms data, and facts into a honed tool, which is even able to propose different tactics and simulate different situations.

### 2.6.2 Defining user needs

Five main steps are necessary to identify users and their needs. The following acting steps identify the background with direct observations or discussions. The goal of the resulting definition is a collection of all available information about users to *make up the body of user intelligence* (Chandler and Hyatt 2003, 174):

1. observe and listen to users;
2. review market segmentation and user profiling;
3. analyze current user behavior and customer objectives;
4. define user interface requirements;
5. formulate value proposition and user needs.

After the execution of the five steps users can be divided and grouped by their needs and expectations. This guideline also helps to understand and analyze the decision making process of a user under his personal circumstances. The order of the five points is necessary to optimize the qualitative result of the analysis and to minimize the effort behind. The result is a classification of users, their demands and needs, the expectations, and the system requirements. This analysis is important to maximize the quality expectations of a user and his satisfaction.

The observation and the direct contact with the users are necessary to determine their behavior. The users and their attitudes can be profiled and gathered into user groups. This step eases the analysis part of the users and helps identifying their objectives. After this basic work the next step is to consider the user interface requirements.

In the process of user interface design the main problem is to decide what the customer or user wants to do. This describes the functional base of the interface. The logic area describes the steps a user has to take to accomplish a task. The task can be ordered sequences of steps including signals and messages from the system accessing one or more data bases. In this field it is important to discuss the menus that provide choices and the following user-steps to reach the requested output. The three main factors of user interface requirements are the appearance of a system, the data base behind, and the application. (Chandler and Hyatt 2003) The acceptance of the user is individual and intuitively influenced by the emotions of the user. This effect also influences the non-conscious decision of users in the process of accepting or rejecting a system.

# 3. QUALITY

## 3.1 What does Quality mean to the User?

Quality describes a characteristic property that defines the apparent individual nature of a good or service. It is also a possible measure for value that is expressed in a price. The degree of quality can be reflected in the price level. (Krek 2002)

### 3.1.1. General Hierarchy of Needs

A lot of definitions of quality are known. For this approach the operational definition is most useful: A datum or collection of data X is of higher (or better) quality than a datum or collection of data Y if X meets customer needs better then Y (Redman 1996, p. 5). The needs met represent the level of quality for a certain user group. Every user is an individual and he divides his needs into several levels of priorities. The most relevant and often used concept is that of Maslow (http://www.ship.edu/~cgboeree/maslow.html). The psychologist Abraham Maslow arranged human needs in a hierarchy. Higher levels of needs are dormant until lower level needs are satisfied.

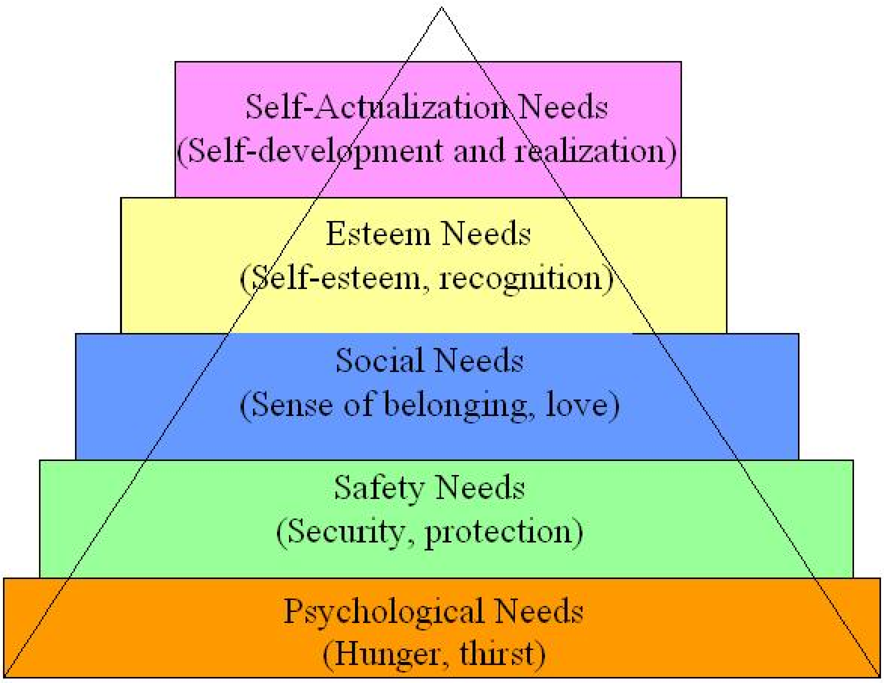


Figure 6: Maslow`s Needs Hierarchy

This pyramid shows the basic order of the human needs. First physiological needs have to be met like hunger and thirst. The safety needs are in second order. Social needs like sense of belonging and love are the third group a human wants to acquire after the other two have been fulfilled. Self-esteem and recognition appears after the previous needs are met. The highest human needs are self-actualization needs like self-development and realization. If one of the lower needs is not met, the higher ones are not important for a human.

### 3.1.2. Good Data

Good data quality plans satisfy the following criteria:

- They should focus on the *most important data*. Here „the most important data" represents those data that are most critical to the user.
- They should be „*customer-driven*". That is data should meet the needs of users or customers.
- The plan should clearly *define management responsibilities*, both for actual improvements and overall program administration.
- Data quality is defined by „*fitness for use*".

The characteristic tree of the most important data quality aspects is shown in Figure 7 taken from a paper by Boehm, Brown et al. (1980). This tree explains the variety of elements influencing quality. Good or sufficient quality can be explained with general utility. Utility is divided into two subgroups in Figure 7: as-is utility and maintainability. An exceptional factor is the portability describing device-independence and completeness. Reliability of a system, efficiency, testability, understandability, and modifiability are also necessary to prove quality of a certain tool. The fourth level shows the quality aspects affecting directly the third level. Interesting are the crossings and the linking arrows can differ depending on the application, the usage of a tool and a user group with its needs.

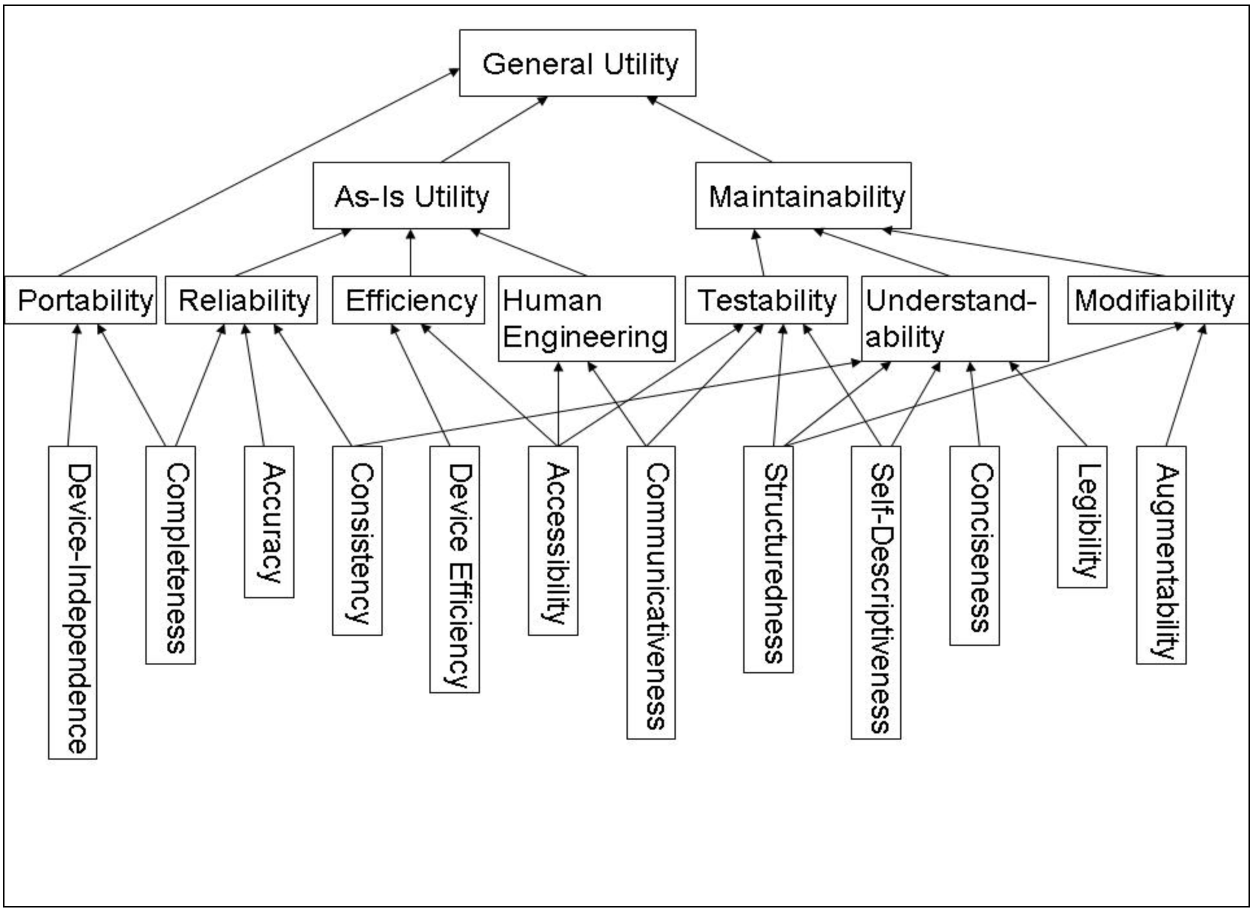


Figure 7: Characteristics Tree for Data Quality and Utility

Quality aspects are graded by a metric system using a weighting algorithm. Completeness and accuracy is often expressed in a percentage scale. Accuracy can be explained by the last update. Device efficiency, device-independence, efficiency, legibility are difficult to describe. Basically all quality aspects are of different importance for the user. Most of the factors are described in a different scale system. To aggregate all metric scales in one system with the same scale, the quality function deployment described below is one possible approach. (Naumann 2002)

## 3.2 Quality Function Deployment

This tool translates subjective user requirements into an objective technical specification. The possible quality aspects are assessed with this tool and adapted to the user needs. Based on the previous section, several quality aspects are mentioned but the priorities of the individual elements differ. After the definition of user groups and their requirements the quality demands can be recognized and assessed according to their priority. The relationship between users, their requirements, the features and the associated processes are described in a matrix. These relationships include:

- the impact of features on customer satisfaction for each requirement (high, medium, or low scale);
- a translation of each customer requirement into technical feature specification; and
- a further translation of each customer requirement and technical feature specification into technical process specification.

The generic, two-dimensional QFD matrix (Quality function Deployment) is shown in the next figure:

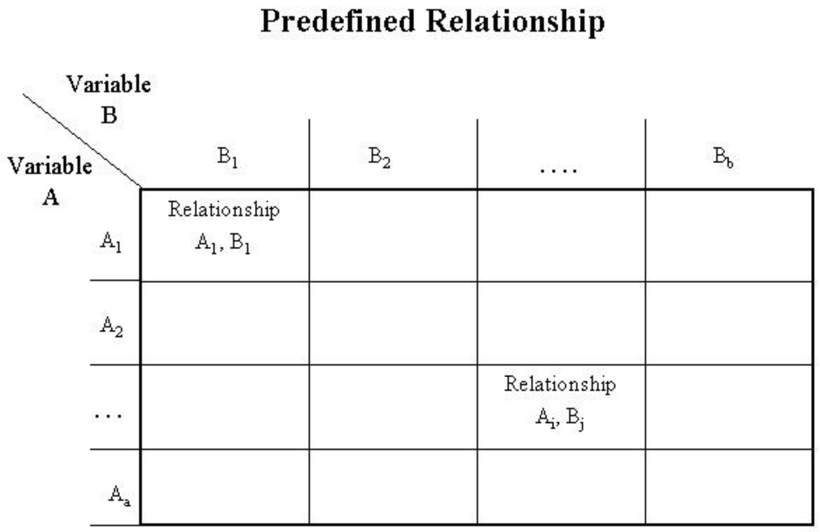


Figure 8: A generic, two-dimensional QFD matrix. (Redman 1996)

Rows and columns represent variables and the entries of the matrix define a predefined relationship between these two variables taken from Redman (1996, 145).

The goal of this matrix is to translate the voice of the customer into the language of the process. This method consists of five steps:

1. Understand the needs of the customer in his terms.
2. Develop a consistent set of data quality requirements.
3. Translate the data quality requirements into technical requirements.
4. Describing and addressing each technical requirement to the different relevant processes of the information chain by budgeting and producing performance specifications on each process.
5. Accumulate performance specifications and give the overall specification for each process.

After understanding user needs, describing processes, and determining specifications a system of measures has to be established to find out whether customer needs are considered completely with their consequences. Measurement Systems are a vehicle to transfer and compare needs with the output of a system. This system is individual and has the task to inform all chain members about the degree of covering customer needs. Participants like data producer, program developer and so on have also to be informed about the degree of meeting needs. Short-term conflicts and problems can easily be identified and solved. It is difficult to implement a system satisfying all criteria and user needs. They are the overall system to perform an optimized output beginning from a task description over a transformation process to the final solution, offered (by display) to the user.

## 3.3 Analysis of Relevance

### 3.3.1 Main crucial aspects in relevance of data

Calculating relevance only via a change of outcome (Frank 2002a) is a time-consuming and therefore expensive way because one has to assess the needs of a user, which is not always easy. A simpler approach is to find out in advance which datasets are redundant and which should not be changed to maintain a certain level of quality.

Another relevant aspect is how to define quality, which aspects in data are important and which can be neglected. Finally, whether there are dependencies between data and how to take them into consideration. Exact investigations about users and their demands lead to a classification of data quality, in order to maximize utility and usability for the customer.

This leads to the consideration which algorithm has to be used for minimizing the previously collected, evaluated, and ranked data by importance for each defined user group limited by their demands and requirements. One easily applicable algorithm is the Minimal Cover Algorithm (Saxena 1988). This algorithm allows multiple solutions and can be easiest changed and adjusted to different problem situations. The easy handling and calculation of the Minimal Cover Algorithm are crucial for this thesis.

### 3.3.2 Application and Example of Minimal Covers and preserving Dependencies

The street network of Vienna is the used example in this thesis. The system is based on directed graphs to include navigating in a network with one way restrictions. Directions are defined on the street segments. Nodes are either an intersection point or an end point of a road and represent an intersection of two lines or a dead end of a road. This classification is taken from Krek (2002).

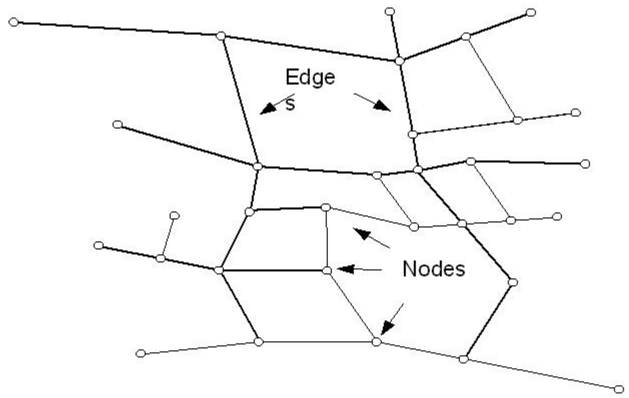


Figure 9: Street network represented as a graph

The next step is to convert the network into a model using directed graphs. A user or agent wants to move from its current position on one node to the next node until reaching the final destination. The direction of the graphs picture the ways he can move. One ways are shown as a single directed graph from A -> B. The two way traffic roads are split into two graphs with opposite directions in this simple graphical model from X -> Y and Y -> X. Finally, the model is a list of relations.

A -> BD, B -> E, C -> ABDG, D -> A, E -> F, F -> EG, G -> CDE.

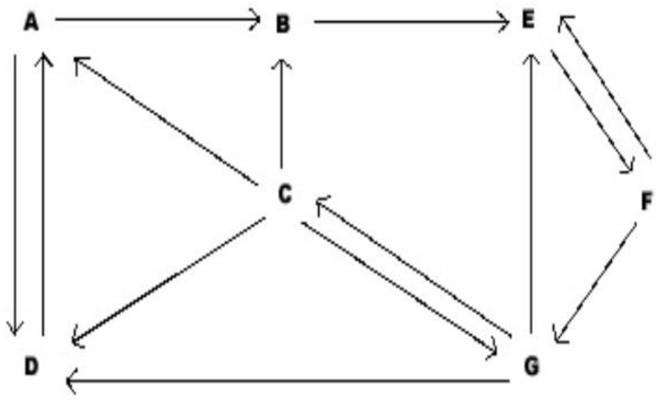


Figure 10: Street network represented with directed graphs

In this example every point on the map is reachable from each starting point. The proof of this statement is given by the cover (Ullmann 1988, chapter 7 "Reasoning about functional dependencies"). Functional dependencies are relations, so the same reasoning applies. Point A has a relation to point B and D. The first step is to divide the assembled relations into individual relations (A -> B, A -> D, C -> A, C -> B, C -> D, C -> G,….). The full list of relations is not shown in this paper but easy to reproduce. To compute the cover, just check this list in the following way to get the cover of A: A -> A, A -> B, A -> D, B -> E, E -> F, F -> G, G -> C. The cover of A is formally: $A^+$ ={A,B,D,E,F,G,C} or in alphabetical order: $A^+$={A,B,C,D,E,F,G}. The remaining covers of $B^+$, $C^+$, $D^+$, $E^+$, $F^+$, $G^+$ can be figured out with this procedure and they evince that each point is reachable from every start.

Some of these relations are unnecessary and can be eliminated, but the previously calculated covers must not change. It is desirable for a decomposition to have a lossless-join property, because it guarantees that any relation can be recovered from its projections. (Ullmann 1988) For a given set of dependencies, an equivalent set with a number of properties can be found. A simple and important property is that the right sides of dependencies are split into single relations. A set of dependencies F is minimal if:

1. every right side of a dependency in F is a single attribute.
2. for no X -> Y in F is the set F- {X -> Y} equivalent to F.
3. for no X -> Y in F and proper subset Z of X is F-{X -> Y} ∪{Z -> Y} equivalent to F.

Intuitively, (2) guarantees that no dependency in F is redundant. Condition (3) guarantees that no attribute is redundant on the left side (3). As each right side has only one attribute by (1) (canonical), no attribute on the right is redundant. This algorithm is also called Minimal Covers Algorithm (Saxena P. C. and Tripathi R. C. 1988, 277-285).

The relation C -> D is redundant. From start C, D is still reachable with C -> A and A -> D. There are several results of this algorithm, which have no effects on reachability but differ in the chosen way and the length of the path. In other words if one relation is eliminated, all constraints are correct and the model is still complete.

This algorithm deals only with the point of view of the data, but the users and their point of view is neglected. Usability is an important aspect and the best model, idea, or application is useless if nobody wants to work with it. More emphasize should be on users, usability, and on the adaptation or improvement of an application in order to guarantee the best output for the user.

## 3.4 Typical Classification: Data Quality

Data quality consists of five basic aspects: the lineage, the positional accuracy, thematic accuracy, logical consistency, and completeness of data. Each parameter of data quality is briefly described in the next section. These parameters describe the metric measurable aspects of quality. They are easy to compare and to work with. More critical are the non-metric quality parameters. They are of greater importance to the user giving a better overview of the user and the user groups.

### 3.4.1 Lineage

Lineage gives information about the origin, the method of data collection and the reality underlying the model. This reality-model includes objects and definitions. Another factor in lineage is the data derivation and all data transformations to produce output "fit for use". In most cases lineage includes also reference to the used control information to complete a data merge between different data-bases successfully.

### 3.4.2 Positional accuracy

The positional accuracy is the exactness of the position of different features using obligations from referencing objects. It is the deviation of the measured or estimated coordinates of a geometric object (using the basic model of points, lines, and areas) compared to reality. (Staudinger et al. 2002) Normally, the accuracy is described by the deviation of data from reality.

### 3.4.3 Attribute accuracy

The attribute accuracy characterizes the thematic distinction, the classification and the assignment of values of geometric data. The definition of attribute accuracy is the same as positional accuracy using the deviation of measured or estimated data from reality.

Values of attributes characterize the definition of objects. In cartography the objects of a map are structured in classes, explaining the behavior of different objects to each other.

Attributes can be of metric or qualitative attitude. Metric attributes are described with descriptive statistics. For qualitative attributes the measurement of nominal or ordinal standard is used. These aspects naturally influence the positional accuracy because substantial properties lead to the appearance and form of an object. The accuracy of attributes determines an object by combining metric and qualitative aspects.

### 3.4.4 Logical Consistency

Logical consistency is the correctness of different topological relations and unmistakable characterization. Formalization has to be complete, traceable, and reconfigurable. The ISO standard talks about the degree of adherence to logical rules of data structure, attribution, and relationships (data structure can be conceptual, logical, or physical). The used obligation and condition is that of the referencing object.

Four types of logical consistency have to be considered. The first is the topological consistency of modeling geometric data. This attitude is important under the aspect of relating and merging different databases. The relation and further usage of the newly created data is only possible if the logical consistency is maintained. It is relevant that every original database uses the same unmistakable characterization of attributes related to the objects. So the database-model has to be consistent to enable a correct merge between databanks, for example merging data from different origin and time. The third consistency concerns the attribute data. If the same attribute data is not used in all databases, a merging of data would lead to a confusing and not useful dataset. The new data would differ in the metadata and could not be compared to the original data. Dependencies and redundancies can disappear or appear and the effects cannot be calculated but have to be estimated.

### 3.4.5 Completeness

The completeness is described by the ISO standard 19115:2003 as the presence and the absence of features, their attributes and their relationships, obligations are used from the referencing object. Completeness requests coverage of the number of all objects in the model with reality. Attributes have to be complete to enable the correct merge of different databases without loss of information. In the ideal case all possible cases should be classified to a certain class or subclass.

Exactness of geometric data and attributes is expressed by the errors of the individual values. Completeness is the over- or underestimation of objects and attributes concerning the complete list of objects: commission and omission.

Commission is a direct effect of overestimation of objects and describes the sum of overestimation of objects minus the number of missing values of variables of the overestimated objects. Omission is the sum of the missing objects and missing values of variables in a database.

### 3.4.6 Time as Quality Parameter

Time is the last factor included in the typical classification of data quality and has direct influence to the accuracy and lineage of data. The timeliness of data can be compared and the data collection is normally periodically done. So time and temporal attributes allow a comparison and combination of different datasets. (Oswalder 1996)

## 3.5 Information Quality

The Information Quality (IQ) is the connector between Data Quality and the user. General definitions for IQ are “fitness for use” (Tayi et al. 1998), “meets information consumers needs”(Redman 1996), or “user satisfaction” (Delone et al. 1992). This implies data that is relevant to their intended use, of sufficient detail and quantity, with a high degree of accuracy and completeness, consistent with other sources, and presented in appropriate ways. Many criteria depend on each other and in this case not all criteria will be used. Information quality is a proposal to describe the relation between application, data, and user. (Wang et al. 1999)

This thesis uses the same assumptions as in a user query for an information quest from web data sources as criteria or information quality. The complete list of IQ-criteria is classified into four sets and their description is from (Naumann 2002) and the ISO/FDIS 19115. Content-related criteria deals with the actual data that is retrieved and the represented properties are intrinsic to the data. Technical criteria concern the aspects determined by soft- and hardware of the source, the network and the user. Intellectual criteria measure subjective aspects of the data source. They depend on the user and the developer and can hardly be measured. Installation-related criteria concern the presentation of the data and are related to usability factors. The relevancy of the different criteria can be adapted slightly using network connections from various data sources.

Table 1 shows the above listed and described quality parameters. They are of great importance for the system and the developer but a user who is not a specialist in GIS-systems is not able to understand and overlook the effects and consequences of the quality parameters. Therefore Table 1 is only the beginning and has to be improved and values added to increase the usability for the users.

Table 1: The complete list of IQ-Criteria (Naumann 2002)

| Category | IQ-Criteria |
|---|---|
| Content-related Criteria | Accuracy<br>Completeness<br>Customer Support<br>Documentation<br>Interpretability<br>Relevancy<br>Value-Added |
| Technical Criteria | Availability<br>Latency<br>Price<br>Quality of service<br>Response time<br>Security<br>Timeliness |
| Intellectual Criteria | Believability<br>Objectivity<br>Reputation |
| Instantiation-related Criteria | Amount of data<br>Representation conciseness<br>Representation consistency<br>Understandability<br>Verifiability |

### 3.5.1 Content-Related Criteria

**Accuracy** formally is the quotient or the number of correct values in a source and the overall number of values in the source. In the field of navigation problems in a street network the positional accuracy is important. Absolute external positional accuracy has to be mentioned, describing the closeness of reported coordinate values to values accepted as or being true. Gridded data positional accuracy is used in gridded fields like clustering of maps. Temporal Accuracy concerns the temporal attributes and temporal relationships of features

**Completeness** requests the coverage of the number of all in the model used objects to real world. Error types are commission and omission. The completeness of attributes is influenced by over- or underestimation of objects. The most important issue in Information Systems is to integrate more than one data source.

**Customer support** is the amount of help for the user via telephone or email and closely related to the documentation criteria. User support is an important factor in the field of usability helping to handle occurring problems and errors.

**Documentation** influences directly the issues of usefulness and understandability. The measurement depends on the application. Sometimes the presentation of data is self-describing and it is not necessary to measure how well a source documents its data. So a more detailed analysis of the documentation is not done in this thesis but further information on this subject can be found in Ossterbauer (2002).

**Interpretability** is the conformation of technical abilities of consumers and the provided information. Interpretability is a critical aspect for the user but has to be judged individually. Each user has different technical and social abilities that can hardly be classified into user groups. Simplicity is one factor that increases interpretability and helps to convert information to an understandable output by interpretation.

**Relevancy** is the most crucial factor showing the satisfaction of user needs by the provided information. In the formal analysis it can be helpful to use ontology to ensure the correct usage and understanding of words (Decker et al. 2000). In this thesis relevancy is reduced to a correctness criterion. Relevancy explains which information is relevant for the user and the categorized user groups.

The **value-added** criterion shows the monetary benefit for the user using the information system. It assesses the most cost producing factors in the information systems. Unnecessary costs are produced by using and providing additional information and data that is not requested and relevant for the user, the user groups, and the specific tasks.

### 3.5.2 Technical Criteria

**Availability** of a data source is the probability that a feasible query is correctly answered in a given time range. In this case it is not important whether the complete response is given or no response at all. Another definition explains availability as the degree of the ability of one unit to be functional at a certain time or time period. In the information system availability is used to measure the ability to provide the requested information adapted to the specific task of the user or user group.

**Latency** describes how long it takes a requested response to reach the user measured in time. In this thesis latency is equal to the response time and influenced by the system combining the hardware and the specific software of the user. Latency is not relevant for this thesis but mentioned to complete the list of technical criteria. Latency cannot be measured for each user and is dependent of the hardware the user owns.

**Price** is another important factor. It is difficult to assess the value of the information and the process behind. Price is influenced by direct costs and transaction costs. It is difficult to measure the utility and the benefit, classed with the process of information production and transformation (Frank et al. 2003).

**Quality of Service** describes the streaming process of a response. It measures the error frequency of the transmission between the request of the user and the source, in other words no interruption occurs during the reception of the requested data. The response time is the delay in seconds a request takes. The influencing factors are sometimes unpredictable and in this thesis irrelevant from the users point of view.

**Security** is an important factor for users. Making a request from a database is a fundamental task. Users want their privacy to be kept and the information about the user should not be available and viewable to other users. Using a public web-tool as database requires anonymization of the user and the authentification of the data source should be guaranteed by a trusted organization. Not all users want their requested queries to be public. Using the street network combining different sources makes security important for users and differs depending on the task and the functionality a user wants to benefit from.

**Timeliness** shows the average age of the data in the sources. There are huge differences in the time of the collection or gathering of data. Timelines critically influence the accuracy of data. Of course, it is not possible to keep all data up to date because of the enormous amount of data, but the timeliness of data should be provided.

### 3.5.3 Intellectual Criteria

The **believability** is the expected accuracy for the user. It is the degree to which the data is accepted as correct by the user. The user has his measure to assess the error rate and to determine satisfaction. So accuracy is technically a correct expression but not from the user's point of view. He himself decides whether information is sufficient or not; therefore, believability can be described as trustworthiness or credibility for the user (Wang et al. 1996).

**Objectivity** is the degree of how unbiased and impartial data is. Mostly the degree of objectivity depends on the affiliation of the information provider, the data source provider. The strong connection to the verifiability criterion can be described as follows:

"The more verifiable a source is, the more objective it is. Again, objectivity is measured by some grade as there is no real unit for this criterion. (Naumann 2002, 35)"

The last intellectual criterion is **reputation**. Users have to make their own experience using a data source or a tool like an application. In normal life a user can pass the positive or negative impression to other users by telling and sharing their experience. This factor can also spread in the market for GIS-products using the informal or formal ways of communication between users and hierarchies in enterprises and consolidated companies.

### 3.5.4 Instantiation-Related Criteria

The **amount of data** is measured in bytes and is the size of the query result requested by a user. The amount criterion is influenced directly by the kind of user request and depends on the application that is used. The amount of data combines the profile of the user, the user query as well as the needed and requested data for the user.

The **representational conciseness** is the degree of matching the structure of original data to the final requested data as output. As already mentioned the original data is not understandable to users and has been transformed and extracted to make it less complex. This process of transformation is critical because of the danger to lose information or to combine data in a contradictionary way leading to conflicts or wrong interpretations of data.

The previous criterion goes hand in hand with **representational consistency**. This factor is critical to maintain homogeneity and value consistency. Representational consistency proves the compatibility of all user queries using different data sources and compares the structure of the current request to the previous requests. This factor is similar to the repetitiveness that data should provide the same attribute values for the same requests (Wang et al. 2003). Users want their requested response to have the same appearance and level of content like former requests.

**Understandability** is the most important factor for the user, describing the data comprehended by the user. This factor is directly responsible to usability of data and an application or system. Understandability should be independent of any representational changes compared to the representational consistency. The semantic value of the processed data as output is the crucial factor to provide understandability.

**Verifiability** is under certain circumstances essential. If more than one source of data is used to process an output for the user, believability of the different data sources is differing. The verifiability is a measure to check the output for correctness for the user. The main keywords in this content include traceability of data and provability of data sources. The assessment of verifiability is mostly done by a third party that should be trustworthy.

## 3.6 Production of Data with degraded Quality

This section gives an overview of possibilities how to derive different grade of quality. Two datasets are normally of different quality and characteristics. Problems may occur during merging if data has been retrieved in different ways. Merging different datasets is a common procedure to integrate metadata in already existent information.

### 3.6.1 Merging Data with different Quality

The basic framework is from Frank (2002a), where a dataset $K_i$ and an additional dataset A are given. These two are merged in a new dataset $K_j$ by an operation. Consider a decision function *d*. When applied to $K_j$ the decision function gives the outcome d ($K_j$) = $o_j$. The dataset A contains relevant information for the decision, if $o_i$ is different from $o_j$.

An example would be merging data collected in different time resulting in an increase in data quality. In Austria data from population statistics 2001 is available, but actual data from 2002 is not. Some private departments have collected samples to approximate the actual numbers, but these are not being published. On one hand it is difficult to gain access to them and on the other hand there can be huge problems in the process during merging.

Depending on the decision function *d* it is necessary to determine precision and quality of data. As mentioned in the previous chapter, exactness is one of the fundamentals of data quality.

### 3.6.2 Merging Data with noise

The same framework as described above is used in merging data with noise, which effectively degrades the dataset. The important aspect here is the relevance of data. Dependencies must be preserved and the inserted noise must not be relevant for the decision at hand, but should be relevant for all other decisions for which the data could be used as well. Consistency is the key for correct merging, where under certain constraints the result of a merge is again consistent.

# 4. DATA AND USABILITY

This chapter connects the previous examined aspects of Data Quality to the common known Usability factors. Usability and the main slogans are used in the next chapter to reinforce the situation whether a user accepts a certain degree of quality or states quality as insufficient for usage.

## 4.1 What is Usability?

Usability is one of the most important factors in the phase of designing up to selling a product. In the last years, usability and the impacts on products have been neglected by huge parts of the computer science industry. It refused to accept usability as major criterion in developing products. Instead of usability, the main goals in the product development have been the attributes and the efficiency of products (Jakob Nielsen 1993; Jakob Nielsen 2000). But the efficiency of a product is influenced by the acceptance of the user. Usability is one basic step to acceptance and finally to efficiency of a product.

„The user is not a designer and the designer is not a user." (Jakob Nielsen 1993, 12) Different users have different needs and the system should provide a usable platform. Most problems occur from the fact that a designer constructs a system from his point of view. Specialists, designers, and programmers work on solutions for users, but they do so from their individual point of view.

A new approach is the "User Centered Design", UCD. Prototyping is described by ISO-standard 13407: "Human centred design process for interactive systems". The main mantras used here are "Know your user!" and "You aren't the user!". Both slogans describe the importance of the user (Fröhlich et al. 2002). Concluding from own experience as a user to other user groups is dangerous and should be avoided. It is only possible to understand the user groups and the context of usage by careful analysis (Hynek 2002). User Centred Design focuses on the users and their requirements from the beginning of the production process.

The critical requirement is to include the user demands and needs to the process of producing a solution. Finally, efficiency of a product or a solution is the satisfaction of the user. If the degree of data quality is sufficient and the representation of the solution is gratifying, the needs of the user are met and the efficiency of a product is optimized.

## 4.2 Aspects of Usability

Usability can be explained as “user friendliness”. The user is satisfied, and his requirements are met. The conclusion for the user is to accept a system but to ignore the preliminary functionality. The system can be an application, a database, or only a single set of data. The main goal of a system is to reach acceptance from the user.

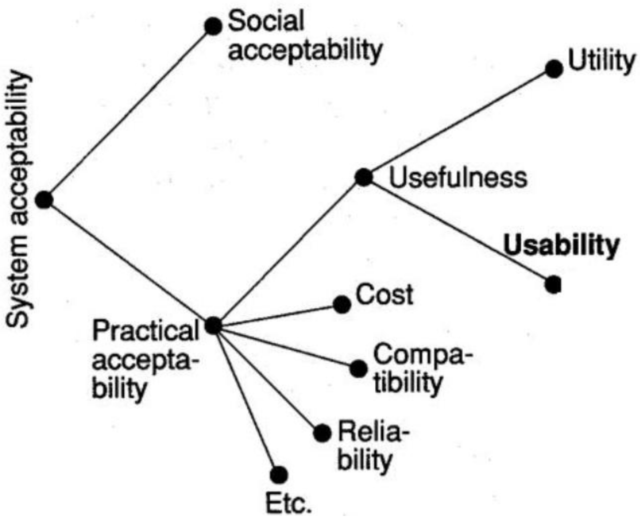


Figure 11: Model of the attributes of system acceptability (Jakob Nielsen 1993).

Usability is one of the main factors that help a user making up his mind about the system or application. Usability is determined by the purpose and by the individual acceptance of the user. Thus its goals should be kept in mind in all stages of developing a new product or system. Even if only limited resources are available or, as in this case, a system is minimized, usability goals should be stressed to adapt the system as optimally as possible for the user.

“Problems related directly to identify goals can be given top priority and problems that do not relate directly to identified goals can be put on the back burner.” (Mayhew 2002)

Therewith usability is the basic factor to achieve acceptability. The major error during developing a system is to insert the usability rules in the last phases of a project.

*"Usability goals should drive design. They can streamline the design process and shorten the design cycle."* (Mayhew 2002)

Factors like reliability, compatibility, cost, and so on affect the user directly. Usability factors influence the decision of the user indirectly and can lead to subconscious decisions that are hardly traceable.

What is affecting or describing usability? The usability of a product is, according to the ISO-standard, *the degree of usage by a certain user to reach certain goals in a certain context efficiently, effectively and satisfyingly* (ISO 9241: Ergonomic requirements for office work with visual display). Usability is not a single-, one-dimensional property of a user interface and has multiple components associated with the five major usability attributes:

### 4.2.1 Learnability

Learnability is the first and fundamental attribute of usability. The first experience of a user leads to acceptance or rejection of a system. It is common in today's software to provide users should with a short period of introduction and to enable them to start working with the application as soon as possible. All systems have certain learning curves, describing the progress of users with an application. The learning curve shows the normal behavior of users and their ability to progress in the usage of a system.

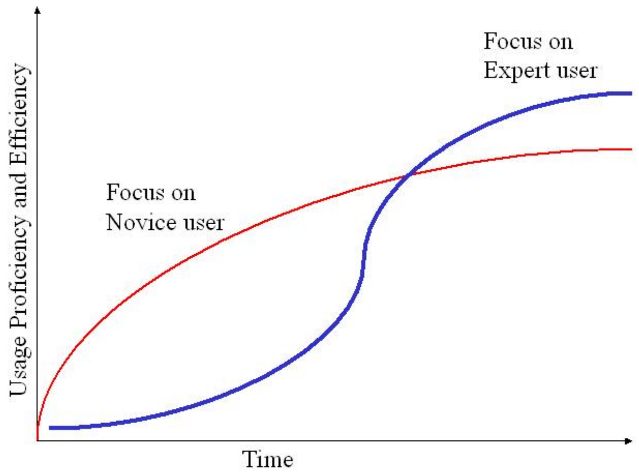


Figure 12: Learning Curve.

### 4.2.2 Memorability

Memorability is directly linked to Learnability. Casual users are defined such that they occasionally make use of the application. They should easily remember the necessary steps after some period during which they have not used the system. These occasional users are the third major category of users besides the expert and the novice users and use the system in low frequency. The casual users indirectly provide the basic information of the Memorability of the system.

### 4.2.3 Efficiency

Efficiency is a subjective attribute of usability. The degree of efficiency depends on the difference between the expected output and the real output in relation to the demands to the user in a certain user group. The formula is:

Efficiency = real output/expected output (user group demands).

The easiest way to increase efficiency is to carefully identify users by observation or direct communication (Chandler et al. 2003) . The important questions are:

- What information/output do the users need?
- Why do users need them?
- How do users use the output after getting it?

### 4.2.4 Satisfaction

The third factor is the users' satisfaction with a certain system. Satisfaction is again closely related to efficiency and the identification of user demands. The degree of user satisfaction is hard to evaluate. Not all users can express their feelings about a system. Some of them refuse to answer directly to questions like: "Are you satisfied and are your expectations you previously had met?" Another possibility to evaluate the user's satisfaction is to assess the real value for the user. Satisfaction depends on the exact identification of users and their objectives.

### 4.2.5 Errors

A system should have a low error rate. An error is defined as any action that does not accomplish the desired goal of the user. The error rate of the system is measured by the occurrence while a user is performing a certain task. Therefore errors also influence the usability of a system by hampering and delaying the desired output for the user.

Errors should be easy to remove or correct. The application should not break down because of errors, which is called "robustness against errors". The user has to be informed about an occurred error, the causality of an error, and the necessary steps to correct an error (Wenk 1996). Errors influence the time of a user working with a system and therefore directly affect usability of a system.

## 4.3. Information Quality and Usability

The crucial factor in an assessment of information quality is the user. His satisfaction is the main goal of IQ-reasoning. The users should participate in the process of selecting the criteria that are used to predict user satisfaction (Chen et al. 1998). Figure 13 shows the connection between Usability and Information Quality. The attributes of Usability influence the acceptance of the user and can help assessing the value for the user.

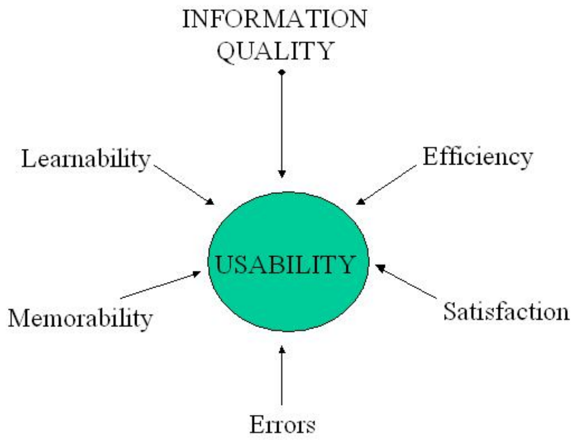


Figure 13: Usability attributes

IQ-assessment is rather difficult, because there are different factors influencing this assessment-process. Most of the criteria are of subjective nature and cannot be automatically measured and categorized. Data sources lack sufficient metadata and background-information about the used data parameters. The amount of data is changing according to the user query (Olken et al. 1990). One of the most important obstacles is multiple sources. They can differ in content and quality and the merge of different data sources can cause problems. Each request by a user group needs other basic data and data quality to produce the degree of information quality.

Three main factors influence IQ-assessment. First, the user as the most important source for IQ-metadata, providing individual input that is collected and refined to improve the quality parameters. The second source of IQ-criteria scores is the data source with all the provided metadata. Third, the query process itself supplies information about IQ-scores and their degree and usage.

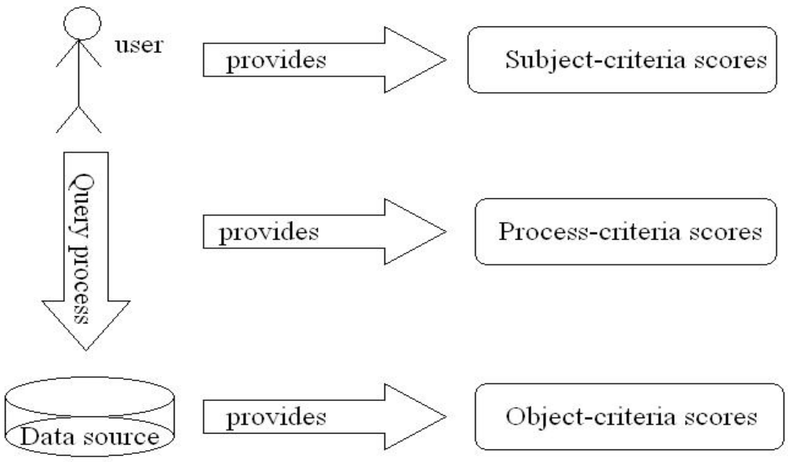


Figure 14: Three sources of IQ-criteria scores.

IQ-assessment is a continuous process that has to be repeatedly adapted and changed. The criteria are subjectively influenced by the different users and of different importance. The used assessment of the IQ-criteria is only a guideline to offer a basic indicator for the users. The measure of assessing these subject-criteria is made by three methods. The technical equipment and the software of the application influence the process-criteria scores. Measurement base are previous statistics and knowledge from former experience with the data sources. The object-criteria are measured automatically only at certain occasions the interference of experts or referencing of the contract is necessary. Table 2 gives an overview of the different criteria, their upper-class identification and how to gather information. As previously mentioned the subject criteria is very complex. It is very difficult to measure believability for example. The easiest way is to ask the users to evaluate the system according to their experience in a ranking system from 1 to 10. Of course that is only one possibility of measuring the IQ-criteria. The permanent adaptation of the IQ-criteria enables to increase usability for the different user groups. The critical fact in IQ-assessment is to translate user requirements into measurable terms.

Table 2: Classification of IQ-metadata criteria taken from (Naumann 2002, 42)

| Assessment Class | IQ-Criteria | Assessment Method |
|---|---|---|
| Subject-Criteria | Believability<br>Concise representation<br>Interpretability<br>Relevancy<br>Reputation<br>Understandability<br>Value-Added | User experience<br>User sampling<br>User sampling<br>Continuous user assessment<br>User experience<br>User sampling<br>Continuous user assessment |
| Object-Criteria | Completeness<br>Customer Support<br>Documentation<br>Objectivity<br>Price<br>Security<br>Timeliness<br>Verifiability | Parsing, sampling<br>Parsing, contract<br>Parsing<br>Expert input<br>Contact<br>Parsing<br>Parsing<br>Expert input |
| Process-Criteria | Accuracy<br>Amount of data<br>Availability<br>Consistent representation<br>Latency<br>Quality of service<br>Response time | Cleansing techniques<br>Continuous assessment<br>Continuous assessment<br>Parsing<br>Continuous assessment<br>Continuous assessment<br>Continuous assessment |

# 5. EXAMPLE OF NAVIGATION IN A CITY

Every person uses transportation data in various situations in life. We have to orientate and to find ways from one point in the world to another. This is a process we are not aware of. Even walking from one room to another includes an orientation process to identify the present location and the destination. The next process is to find a path from the start point to the destination point.  Moving in a city follows the same basic rules. There are only differences in distance and the amount of additional data to process into understandable information. Navigation in a street network consists of the same three basic steps: identification of the present location, determining the destination and finding a way from start to goal.

## 5.1 Identification of Users of a Transportation Data

Who is navigating in a street network and who decides about the path? Dueker and Butler (2000, 13-37) identified two main participants of transportation data: motorists and general public. The two user groups can be further subdivided:

The first user group consists of general services like emergency services, emergency dispatch, and street maintenance. The purpose of public services is to maintain a traffic flow and to help people in case of an emergency to repair damages of the streets and the surrounding and to remove traffic obstacles. Public services have a lot of different tasks and need detailed and accurate data. The required quality of the used data has to be higher than for the general public. High quality data is necessary to calculate risks for people, buildings, streets, and objects. A street segment consists of many different parts. Those users can be classified in one group, having a lot of different tasks but the same requirements. Public service consists of a group of users needing a high quality of data.

The second user group contains users of a vehicle navigation application. This group includes car drives, motorbikes, and bicyclists and so on. The possible purpose and usage is endless. The basic usage is to move in a street network from one point to another using a vehicle.

## 5.2 User groups determined by the way of chosen transport aid

This model assumes that all users have the same possibilities of mode of transportation. A user can choose between car, bicycle, and public transportation. These three possibilities require different information. A bicyclist for example wants information about difference in altitude he has on his way. Other important factors are separate bicycle lanes to avoid the traffic on the streets. Another aspect is: what information does the user already have?

A possible strategy is to decide first, which transport possibility to choose reflecting his membership to a certain group. Depending on the social surrounding and the social resources a person has the possibility to decide on the transport vehicle. In this thesis the distinction is made via the different kind of vehicles in the way-finding problem in a city assuming that a user, or in that specific case a driver, chooses his transportation vehicle according to the purpose. To maximize use and minimize costs, a smaller vehicle would be preferred by the users. If you have to transport heavy freight the choice will fall to a bigger van or a truck. The conclusion is that purpose leads to the chosen way of transport.

Riding a bike is a common way of transport in Vienna because the running costs are low compared to car or truck. Bikes have also other advantages: they are easy to buy, it is easy to park, no parking tickets necessary, and no parking restrictions have to be considered.

Table 3: Three transport vehicles comparing the cost of acquisition, running cost, and parking or storage costs

| Bicycle | Car | Public Transport |
|---|---|---|
| Easy to buy<br>Low acquisition costs | High acquisition costs | No acquisition costs |
| Low running costs | Low running costs | Medium running costs depending on the travel distance |
| Easy to park | Expensive to park | No parking necessary |

The last transportation method in this thesis is public transport. Public transport is not always the easiest way to move from one point to another. A ticket has to be bought and from the start point the access-point to a public transport has to be chosen.

In Vienna the user has to choose between bus, tramway, and metro. The coverage of the three groups of public transport is overlapping. The differences are the time needed and the frequency. The chosen way of public transport can vary in the combination of transport possibilities and can be confusing for the user. Sometimes it is faster to change line three times, but it is more comfortable to stay in one tramway or bus until the destination is reached. The user has to decide between a more comfortable connection with fewer changes or the fastest way. This leads to the following table showing the user groups and their chosen vehicle:

Table 4: Identified user groups for the navigation problem in a city

| **Vehicle** | **Purpose** |
|---|---|
| Motor car | Salesman, Commuter, Tourist, Business Traveler, Standard User |
| Truck | Industrial Supplier, Retail Supplier |
| Public Transport | Inhabitant, Commuter, Tourist, Business Traveler |
| Bicycle | Inhabitant, Tourist |
| Inline Skater | Inhabitant |

## 5.3 Task definition and usability in GIS

What are the tasks of current GIS and how can they be ranked? A comparison between user tasks and task descriptions in different contexts enables a structured overview of GIS use (Davies 1994). Whitefield et al. described the taxonomy (1993) of “work tasks” as tasks to fulfill the user’s work goals and exclude all the extra ‘enabling tasks’ like switching on and off the computer, starting software, etc. So emphasize is put on the usage of the GIS software and not the basic and underlying functionality of the system and hardware underneath. For example, the warm up time of the system can be a noncritical factor influencing the output or the usability.

In this work the usability focuses on the usage of GIS software and his functionality. Generally, a GIS is capable of storing and displaying maps on a screen, zooming in and out to enhance details or to give an overview, and changing the selected section of the map. Most of the GIS systems allow storage of attribute data. Attribute data is additional information about specific points or objects on the maps. Normally, a user can change the colour of schemes, annotations, legends and titles. He can plot or print maps or selected areas. The great advantage of GIS software and solutions is the possibility to manipulate the spatial data and to use spatial statistics. Specialists take advantage of the increasing functionality, but, like in this paper, a regular user, moving from one location to another by a vehicle, is not interested in making calculations and analysis of maps. He only wants to use the provided data and solve tasks to find a way from one point to another under some constraints.

Tasks have to meet a purpose and, depending on the user, this purpose differs. GIS tasks can be subdivided into different levels of detail and the granularity of task description can also vary. The Rasmussen hierarchy containing the purpose, the abstract function, the generic function, the physical function and the physical form can describe each task (Rasmussen 1986). User tasks are the key element for usability of GIS tasks considering the context of use.

In real life a user is not satisfied with the simple solving of the general wayfinding problem. Every user is an individual and has special subtasks to the problem. Of course it is impossible to describe and to meet all individual demands on the system, but it is possible to generalize the tasks. The easiest way is to find out what information could be interesting for a person using brainstorming, questionnaires, and interviews. The following elements were mentioned, depending on the different persons and their underlying task.

## 5.4 A Model of Consumer Behavior

The interesting process is the decision process of a navigator. How do consumers make their decisions? A buying decision is a confusing and very often irrational action. Consumers are overwhelmed by the amount of information and the variety of different choices. Typically, the consumer’s behavior is the result of the influence of a variety of factors and the interaction between them Czinkota (2001, 104). For further analysis a separation between environmental influences and individual differences seems appropriate.

Culture, social classes, family, personal influences, and situations determine the decision process. Individual differences, as consumer resources, motivation, knowledge, attitudes, personality, values, and lifestyle have to be considered. Regarding the variety of factors, the decision process is not simple to understand.

Experience is underestimated for the navigation application. Every user has gathered his set of experiences influenced by different environments, circumstances, and occasions. Every person has his own set of constraints to be met. Some requirements are not so urgent and do not have high priority and other terms can be neglected. To gather and group different experiences and influences for users is a difficult task for the analysis.

Acquisition of knowledge is an investment, which is depreciated or distributed over many uses. Investment costs are influenced by the frequency of movement in a city. If relocation has to be made only once and no transportation of heavy goods is included, the easiest and most economic choice would be public transport. If the frequency of a trip is increased, the costs are also rising and taking a private transport vehicle like a car is preferred. A movement with heavy carriage like moving from one apartment to another can evaluate the decision to rent a truck. It is more cost efficient to pay a single transport with a bigger truck than driving the car many times. Moving by public transport is not practicable. Another example is to commute to the workplace and back every day. These examples explain the influence of frequency in transport.

Driving a car in a city is a common task. The case study used in this paper is car navigation in the city of Vienna. Driving in a city includes orientation, where the user is located, defining the target point and deciding at each intersection which way to choose.

The basic problem for the user is often the large amount and variety of information provided in maps. For example, a street consists of several elements. Additional information is not needed for the user decision process and complicates the overview that is necessary to make a decision. A driver only needs a small amount of basic elements to find his way from one point to another, from one start to a destination. These elements are points, lines, and areas. They are surrounding the initial position; describe orientations, and the target destination.

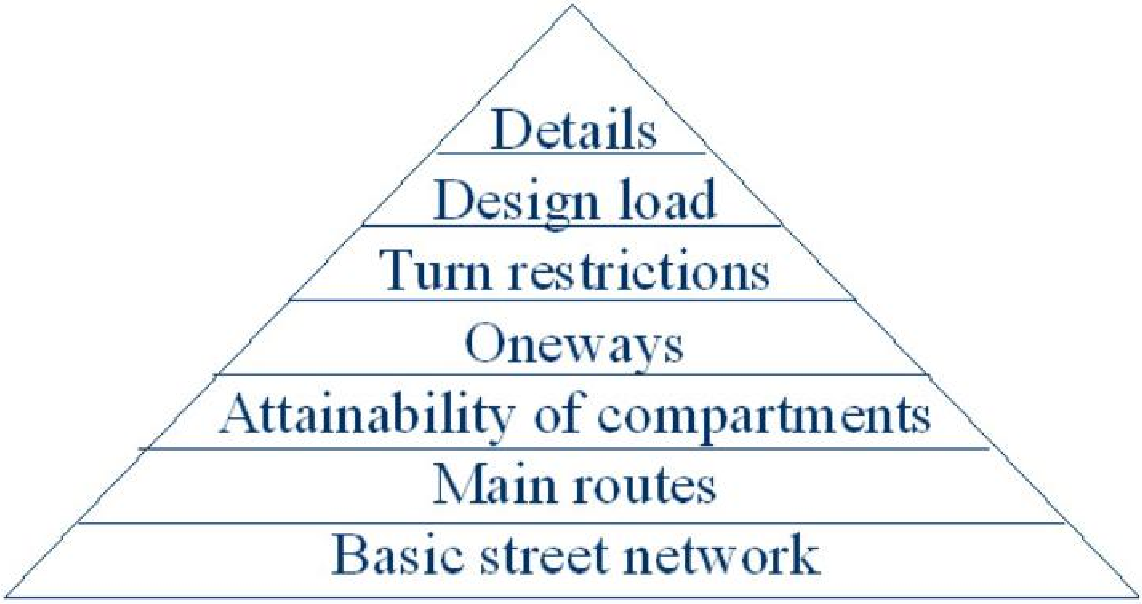


Figure 15: Maslow Pyramid for a truck driver leading from start point A to destination B

This pyramid shows the details needed for a truck driver in a city. The main information is the basic street network. The second level includes the main routes. The third level is the attainability of compartments. Under special circumstances it is impossible for certain vehicles to enter an area because of one ways or turn restrictions. These aspects are listed as additional facts. Their importance depends on the chosen example.

Not considered here are physical requirements like filling the vehicle with gas or resting time for the driver, etc. It is assumed that these needs are ignored in basic transportation applications. The technical and human resources are maximal, in other words the driver is in good health and well rested, the vehicle is fully operational. The only process is to move in a city from point A to point B.

The levels offer a possible gradation and valuation of the different information levels. The higher the levels in the pyramid are the higher the possible price. As possible other variables can be information about speed checks, road works, traffic jams, free or occupied parking places, possibilities to acquire parking tickets, public transport stations, and timetables, location of gas stations, shopping centres, embassies, sights, tourist-info, parks, bicycle routes and so on. This list is not complete and can be extended any time with other parameters according to other tasks and examples.

The advantage of this pyramid is that all levels can be exchanged and newly ordered. The change in priority of user needs can be expressed and re-evaluated simply with this method.

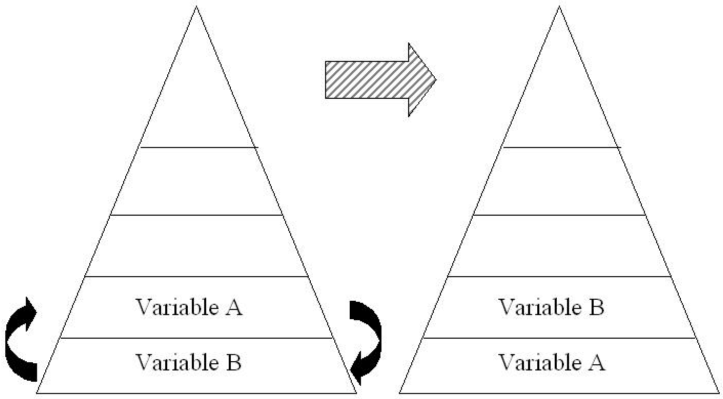


Figure 16: Maslow Pyramid with changeable levels

The observed order for human decision is to meet first physical, then legal levels, and finally, social requirements. Following that order predetermines the Maslow Pyramids and eases exact description of the task to perform.

### 5.4.1 What Information Does the User Need?

Another approach is to classify characteristics instead of users. To classify users is problematic, because sometimes users behave "abnormal". It is impossible to forecast all possible situations, problems and behaviours of the customer. Only during the evaluation phase the correctness and completeness of the chosen models and constraints can be proved.

The next step is to subdivide all the constraints into classes with different priorities. The developer has to predetermine the priority classes, their range and their tolerance. The major problem of this process is to transform nonnumeric information into numbers and to find a sufficient tolerance range. The problem of tolerance ranges is excluded in this thesis, but leaves interesting research aspects for future projects.

## 5.5 Example

### 5.5.1 Collection of data

The user in this thesis is an agent moving from point A to point B in an area of a city. The chosen area of the city is the 8th district in Vienna. The location of this district is next to the city centre and therefore the streets are heavily used. The street network can be compared with a relation system. A street section from crossing AB to BC can be seen as single relation. So in this example the agent can move from point AB to BC (see Figure 17: System of crossing streets). If the other direction is legally forbidden or not possible because of other restraints, this section is called one way. The theory was shown in section 3.3.2.

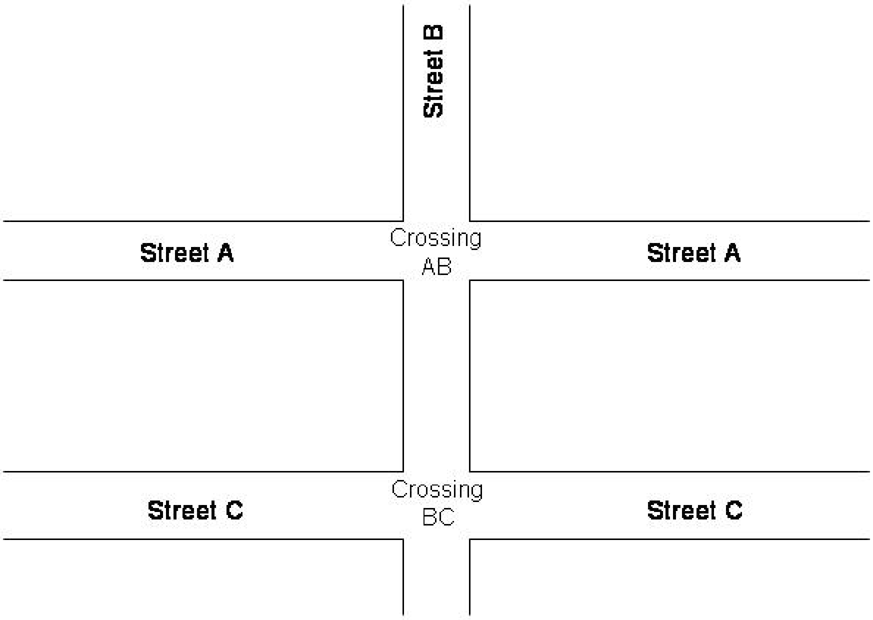


Figure 17: System of crossing streets

The first task in the chosen example is the simple naming of the crossings according to the streets they contact. The result is a table with empty and indefinitely extendable number of columns. Only in the fist column the name of the street section is entered. Each street section is unique. Section AB-BC is not the same section as BC-BA. If the section AB-BC is a one way, as previously mentioned, BC-AB does not exist and a line in the table with this name is not generated. The logical conclusion is to use the generated column with the street section names as primary key for the following functions and analysis.

Considering the one way situation assumes that this restriction is valid for the agent or the user of the street network. If the agent simulates a pedestrian, one ways are of no importance. The users simulated in this thesis are primary grouped by the way of chosen transport aid by definition. Pedestrians are disregarded here. One way restrictions can be not valid for other user groups under special circumstances. For example, some one ways are not in force for bicycle riders (if specially marked). Another example is an emergency case. If necessary or given danger the situation can lead to the necessity to violate legal restrictions. Fire brigades sometimes break the rule not to drive against a one way under the condition that traffic is not endangered by this normally illegal and strictly forbidden action. Such situations are not really allowed but tolerated if there is danger for the life of citizens. The decision to consider one way restrictions simplifies the possible case scenarios. This thesis focuses on possible movements of an agent using a vehicle and excludes illegal possibilities.

This assumption already degrades quality of the produced map. Completeness is one important criterion as described in chapter 3.5.1. In this thesis a simple solution is used to increase completeness. All street sections are listed in the table, disregarding any restrictions. An additional column is inserted in the created table showing the permission to move from section start to section end for a vehicle. If the value of the field is 1 moving from AB to BC is legally allowed. If the value BC AB is 0, then this particular section is a one way. Finally, the created table can be also used for pedestrian behaviour. The purpose of the table can be easily extended for future possible analysis.

The next step preparing the table for the final simulation is to turn the street section system into a relation system. The condition for the conversion is the column called “legal restriction”. The value in “legal restriction” is either zero, then no relation is created, or one, then the algorithm generates a line with the street section in the relation system. So the lines of the relation system contain only directed street sections. If the movement from start to end of the street section is illegal, no entry in the table is created.

Table 5: Basic example of entries in the moving table

| From | To | Table Entry | |
|---|---|---|---|
| AB | BC | 1 | Connection |
| AB | BC | 0 | No Connection |

Moving from a start point to an end point means covering distance. This distance has to be measured in metric units, collected and added to the table. Each street segment in the table has to contain information about the distance from AB to BC (end to goal). This metadata is necessary to perform the shortest path algorithm in this street system.

### 5.5.2 Weighting data

The previously created table contains all legal possible movements for a later simulated agent moving from point AB to BC. The next step is to weight the priority of the individual entries. The algorithm for the used user groups in this thesis weights multi-lane streets highest. Main routes through the analyzed section of Vienna are also weighted higher. Streets with two-way traffic are higher weighted than one ways. The higher the weight of a relation the more important are itemized streets for a certain user group.

The user needs have to be observed, collected and categorized first. As explained in section 2.6 user needs have to be analyzed and formalized. Grouped by similar needs, user groups are defined. Within one user group the requirements can be sequenced according to the relevancy. The importance of an entry can depend on a decision of a certain user group. The data suppliers can integrate feedback or additional requirements of the user groups in this step of the process. As explained in section 3.5 the user and the supplier have different aspects what criteria are more important and in this special case which street section is preferred and which not.

The simple example of a bicycle rider compared to a car driver explains the situation. The biker prefers streets without tramway. The tracks are dangerous for bikers; they can slip on the tracks or get stuck in the tracks and topple over blocked wheels. In Vienna most of the streetcar drive on busy routes through the city. The urban management tries to separate public traffic with underground and tramways from car traffic, but those efforts are still in progress. Regarding danger of tracks, the bicycle driver weights one ways as relatively safe and higher than busy streets. The car driver prefers direct routes with few traffic lights and having priority to smaller streets merging into the main routes. Truck drivers and bus drivers often have problems turning into small one ways, so they prefer also larger streets with multiple lanes because their turn radius is larger. Many aspects can be listed and compared under the point of view of the user group. The result can show different importance and finally weighting.

Returning to the chosen user groups for this thesis, the weighting of street segments can be similar for different groups. So the common weighting for street segments is raised for segments with multiple lane ways whether there is oncoming traffic or not. Traffic flows faster and for the individual driver getting across is easier and faster. Recapitulating the weighting aspects show high importance for the following street segments:

- street segments with more than one lane,
- main routes through the city or city areas,
- streets with two-way traffic, and
- streets with right of way.

The relation table is now enlarged with one further column, displaying high priority 1 and low priority 0. A more detailed weighting is also possible by using more priority levels. This thesis sets more detailed distinction aside to focus on the explanation and analysis of the procedural method. The consequences for high priority are that the relations are relevant and must not be changed or eliminated. Relations with low priority can be excluded, or eliminated in further processes. The resulting table contains all individual relations, naming street segments, their drivable direction and the priority. The last efforts enhanced the table by adding information and by attaching aspects for relevancy.

### 5.5.3 Applying the Minimal Cover Algorithm

The relation table contains all relations of the street network. The Minimal Cover Algorithm is an algorithm to minimize the table, but to keep the reachability of every street crossing. Chapter 3.3.2 explains the mathematical steps. The cover of the individual relations has to be computed. Each cover shows all destinations that are reachable from one particular start point. In theory every crossing can be a beginning. So the cover of each relation start has to be calculated. This step is necessary to compare the result with the minimized table. The highest priority is to maintain the reachability of the original reachable destinations. Otherwise the usage of the map is limited and unexpected effects can take place. The Result of the Cover Algorithm shows all crossings, which can be reached from one start moving from one crossing to another.

The Minimal Cover Algorithm eliminates all not necessary relations from the table, but does not change the previous calculated cover. Every relation can be recovered by its projection in the minimized table. Further tests could analyze degenerated tables, not meeting the basic requirement of maintaining the reachability. But these experiments would require increased embedding of user groups to analyze the effects on usability. Extensive tests and interviews would be necessary to recognize the fine borderlines and to define a scale mechanism. In this thesis it is sufficient to eliminate not necessary connections of street elements, in other words relations. The mathematical procedure can be described as:

If A -> B and B->C then A->C is not necessary and eliminated.

The basic idea to use the Minimal Cover Algorithm is the possibility to neglect unexpected circumstances and user reactions to arbitrary manipulations of the basic street map. All users or user groups had one common demand on the street maps: the user wanted no loss of information and coverage concerning the reachability of the crossings. If the reachability is limited by the degrading process some users acted completely not expected by simply refusing to use the degraded map. Therefore, the Minimal Cover Algorithm is a simple method to meet this basic user demand and to avoid unexpected user behaviour.

After implementing the Minimal Cover Algorithm multiple outcomes are possible. The more relations exist, the more possible results of the algorithm are possible. The maximal number of possible minimized relation tables is “n relations to the power of (n-1)”. Only one resulting map is used in this thesis to ease the analysis and to comparison with the original map. This resulting map is created under consideration of the previous weighting. So deterioration of one ways is preferred to main routes and bigger streets. Street segments with a protection flag contain the basic points; each one can be used as possible initiation for the Minimal Cover Algorithm.

Other algorithms can also be used to minimize the table. Another solution could be the algorithm showed by Robbins theorem (Robbins, 1939). Nash-William has generalized this theorem further (Nash-Williams, 1960). The goal was to convert a street network into a one way system to optimize traffic flow. This thesis focuses on one algorithm, the Minimal Cover Algorithm, to ease the analysis and comparison of the results afterwards.

### 5.5.4 Calculating the shortest path

Two tables are given after the previous step: the original street map converted into a relation schema and the minimized relation table. Both tables have the same structure. The column expressing the priority of the relation is not necessary anymore and can be ignored for further steps in this thesis.

The short excursion in the graph theory is necessary to explain the coherences and the following steps for the Dijkstra Algorithm (Dorninger, 1996). The edges *e* of a graph *G* is called AB for example and the graph is the entire street. Each edge is already associated with a real number w(e). Those graphs are called weighted graphs. The weighted graphs are inserted in a matrix. The single weighted rows and columns display the edges of the graph. The so-called street-corners (edges) of the street network (graph) are renamed and numbered to easy the idendification. The weight is the minimum distance of the between two edges. This weight is equated with the costs given by the matrix C = [$c_{ij}$].The Shortest Path Problem is the problem to find a way through the graph $G = (X, \Gamma)$ with minimum costs $C = \lfloor c_{ij} \rfloor$ starting from vertex s $\in$ X to a specific ending vertex $t \in X$, provided that such a path exists. $\Gamma$ includes the set of correspondences. The elements $c_{ij}$ of the cost matrix C can be positive, negative or zero.

If a circuit of G exists whose total cost is negative and $x_i$ is a vertex in this circuit, then proceeding from s to $x_i$, leads to result, where no best path can be uniquely defined (Christofides, 1986).

For clarity of exposition the weight of a path in a weighted graph is referred as length and the minimum weight of a $(u,v)$-path will be called the distance between u and v and denoted by $d(u,v)$. All weights are non-negative; otherwise the Hamiltonian Algorithm has to be used instead for the Dijkstra Algorithm. It was discovered by Dijkstra (1959) and independently by Whiting and Hiller (1960) and finds the shortest path from $u_0$ to all other vertices of the graph G. In general, the method is based on assigning temporary labels to vertices, the label on a vertex being an upper bound on the path length from *s* to that vertex. These labels are then continuously reduced by an iterative procedure. Exactly one temporary label becomes permanent at each iteration step. At each stage, these shortest paths together form a connected graph without cycles; such a graph is called tree. A permanent label indicates exact length of the shortest path from *s* to the vertex in question. The Dijkstra Algorithm is described by five steps:

Let $l(x_i)$ be the label on vertex $x_i$.

*Initialization:*

*Step 1*. Set l(s)=0 and mark the label as permanent. Set $l(x_i) = \infty$ for all $x_i \neq s$ and mark these labels temporary. Set p=s.

*Updating of labels:*

*Step 2*. For all $x_i \in \Gamma(p)$ and which have temporary labels, update the labels according to:

$$l(x_i) = \min[l(x_i), l(p) + c(p, x_i)].$$

*Fixing a label as permanent:*

*Step3*. Of all temporarily labelled vertices find $x_i^*$ for which $l(x_i^*) = \min[l(x_i)]$.

*Step 4*. Mark the label of $x_i^*$ permanent and set $p = x_i^*$.

*Step 5*. (If the path from *s* to every other vertex is required). If all the vertices are permanently labelled, then the labels are the lengths of the shortest paths. Stop.

If some labels are temporary go to step 2.

If the shortest path from s to any $x_i$ is unique, then the arcs $(x_i', x_i)$ on the shortest path form a directed tree with s as is root. It is also possible that there is more than on "shortest" path from s to any other vertex. The vertexes of the shortest path can be found recursively starting with $p = x_i$. $p = x_i$. If $x'_i$ is the vertex just before $x_i$ in the shortest path from s to $x_i$, then for any given vertex $x_i, x'_i$ can be found as that one of the remaining vertices for which:

$$l(x'_i) + c(x'_i, x_i) = l(x_i)$$

The shortest path for all possible start points and destinations can be easily calculated and compared with the original calculation. This thesis refers to the work of Krek (2002), where the shortest path algorithm was implemented in Haskell and can be easily used to prove the conclusions at the end of this thesis. This algorithm is the basic tool to compare the later described city area Figure 18 in Vienna.


Figure 18: Original map of the 8th district in Vienna with street names, green areas, public transport, etc.

### 5.5.5 Analysis of the effects

The next two pictures show the original map, the blank street system and the minimized map. The section between describes the decisions leading from the original map to the reduced. The final conclusions are referring to the last chapter.

To increased usability the second map is not minimized but the level of detail is reduced. It is not necessary and for a user not understandable if street sections with multiple lanes or main routes are reduced. Those connections in a city are preferred, because street signs are very often better viewable there and navigation through a city via the main routes for a foreign user is simpler and easier. The user gets multiple confirmations by recognizing street signs leading the direction (in case he is on the right way). This situation is optimal but very rare. Most of the street sign systems in cities are either redundant or leading a way in circles around. The easiest way is to reduce the system not to the minimal limit but to keep some connections available as data for the user.

To increased usability the second map is not minimized but the level of detail is reduced. It is not necessary and for a user not understandable if street sections with multiple lanes or main routes are reduced. Those connections in a city are preferred, because street signs are very often better viewable there and navigation through a city via the main routes for a foreign user is simpler and easier. The user gets multiple confirmations by recognizing street signs leading the direction (in case he is on the right way). This situation is optimal but very rare. Most of the street sign systems in cities are either redundant or leading a way in circles around. The easiest way is to reduce the system not to the minimal limit but to keep some connections available as data for the user.

After considering all necessary facts concerning user and requirements one possible solution of minimizing a map can look like Figure 20. All street segments are accessible; all points can be reached by a user or an agent moving around. The more segments are eliminated the more the danger of reducing important facts for the user group simulated by the agent increases. The difference between the two maps can be presented in distance or in time needed. This thesis does not aim at the concrete simulation of an agent but on the theoretical model behind and the argumentation how to reduce data and limit access to data.

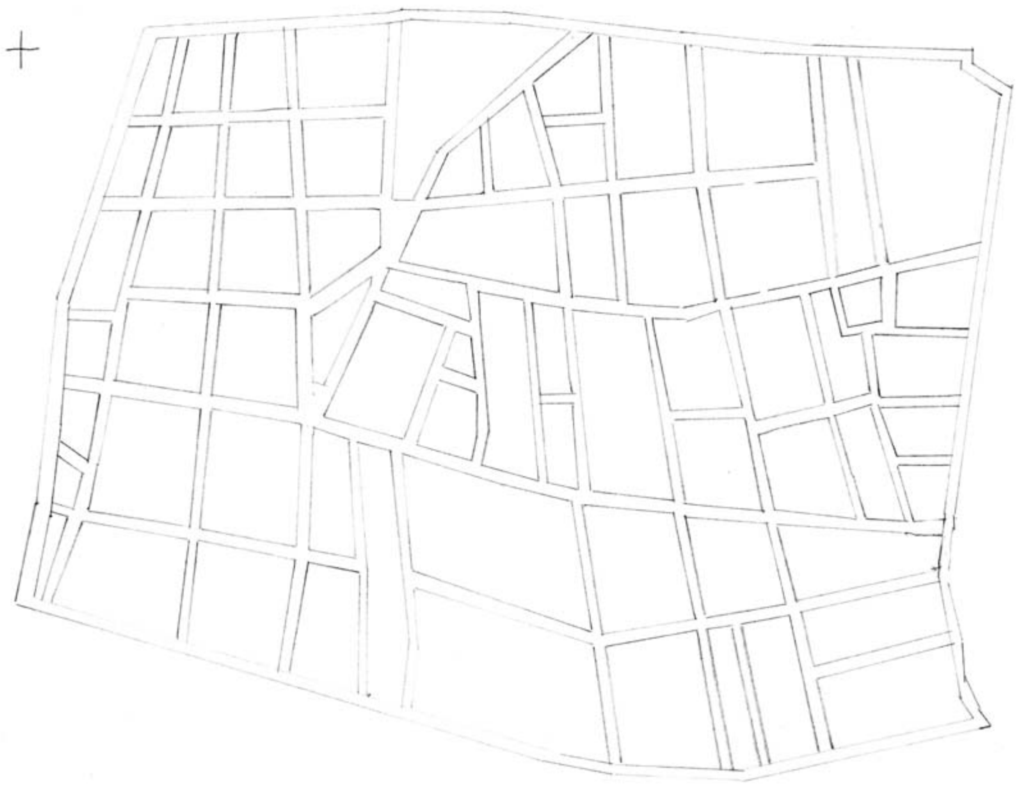

Figure 19: Basic street system of Vienna 8th district

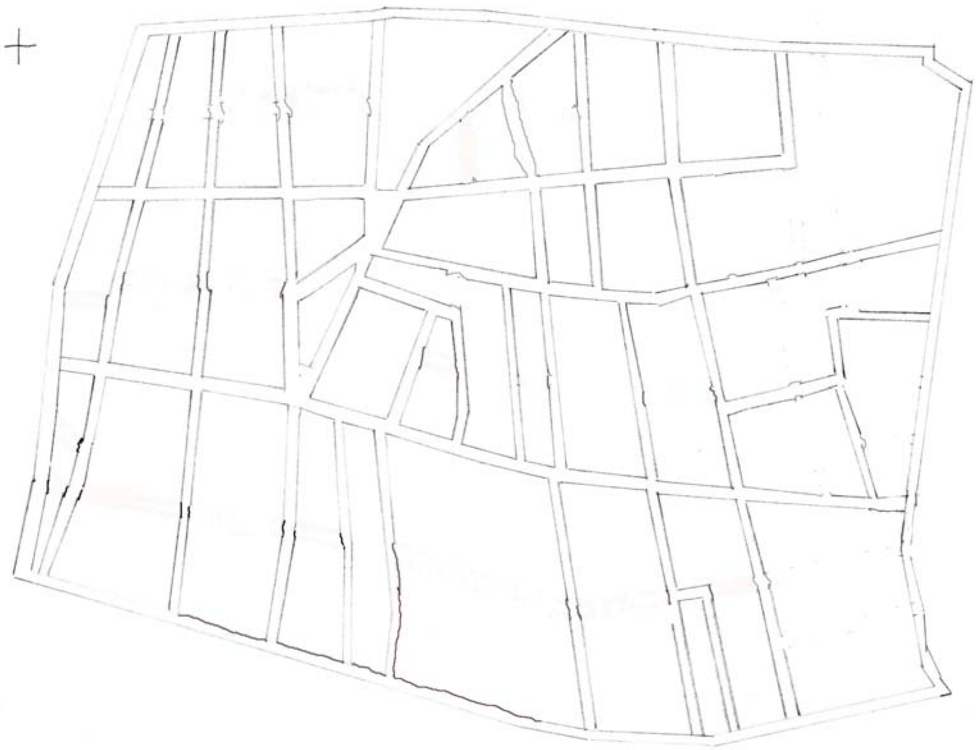

Figure 20: Reduced street map of the 8th district

(missing some not essential street segments compared to Figure 19)

# 6. CONCLUSION AND FUTURE WORK

The first section of this chapter deals with the major results and findings during the work on this thesis. It gives an overview of the most important facts and leads to the final conclusions and the focuses of future research. The example used in this thesis is very simple to ease the argumentation for the results, but an increasing of the level of detail is desirable and leaves room for future challenges.

## 6.1 Results and Major Findings

The crucial process is to find a link between a system, the data, and the user. The user is interested in a broad variety of aspects not only concerning data and data quality. A user wants an exact solution for his problem and does not care about the solution finding process. The critical aspect is to gather information about users and to group them. Each user group has certain requirements and different aspects of usability that have to be considered. The decision function can be easily determined if the exact circumstances of a user, his activity, and the environment are known. A price differentiation between the user groups is finally possible by knowing and analyzing the user group preferences.

After understanding user needs, describing processes, and determining specifications a system of measurements has to be established to find whether customer needs are considered completely with their consequences. Measurement Systems are a vehicle to transfer and compare needs with the output of a system. This system is individual and has the task to inform all chain members of the degree of covering customer needs. Short-term conflicts and problems can easily be identified and solved. It is difficult to implement a system satisfying all criteria and user needs. They are the overall system to perform an optimized output beginning from a task description over a transformation process to the final solution, offered (by display) to the user.

The pyramid of the psychologist Abraham Maslow is applied to show graphically the grading of the requirements of a user group. Maslow arranged human needs in a hierarchy. This concept represents possible hierarchies, showing the result of the analysis of the different meaning of quality from the point of view of the user.

The main factor that can be observed is the connection of data reduction and time. The more important time is for the user the more sensitive is a user group to reductions of the street network by eliminating street segments or other facts. The more data is available to more exact the calculations can be carried out to find the way from point A to point B.

The reduction of data concerning street segments is one possible way of data reduction. A street map consists of several elements including information and location of green areas, bus and tramway stops, routes of public transport, locations of garages, ATMs', post offices, buildings of public services, police stations, entertainment facilities. Varying the number of information accessible to the predefined user groups is less time sensitive and could lead to future tests of user behaviour.

## 6.2 Future Work

Future projects would be a development of user specifications. Users have to be questioned and grouped by identical tasks and requirements. With a sufficient number of users, the determination of different degrees of "user quality" and Information Quality can be proven. Measurement rules for Usability can be derived and adapted to the different user groups. According to the previous found user specifications, marketing strategies can be worked out in more detail.

One very important aspect is to prove the reliability of systems with minimized characteristics as shown in a publication by Birolini (1997a). The reliability is a huge factor influencing the acceptance or refusal of a tool or data set. Therefore the reliability must not be neglected for future projects.

Future work would be the development of the Maslow's Pyramids for different user groups and situations. These pyramids are simple graphic solutions to represent the individual rankings and priorities. Each user group has his own ranking system where the different levels can easily be adapted or changed. Depending on which level is required, additional information can be requested from a database and so a "user-optimized" dataset can be created. Extensive tests could give additional evidence and improve the adaptation to each User Group separately, leading to a possible expansion of attractiveness for future customers.

## REFERENCES:


Birolini (1997a). "Zuverlässigkeit von Geräten und Systemen." 370.

Birolini, A. (1997b). *Zuverlässigkeit von Geräten und Systemen*. Berlin Heidelberg, Springer Verlag.

Bots and Jansen (1989). *Organisatie en Informatie*. the Netherlands.

Chandler, K. and K. Hyatt (2003). *Customer-centered design a new approach to web-usability*. New-Jersey, Hewlett-Packard.

Chen, Y., Q. Zhu and N. Wang (1998). Query processing with quality control in the World Wide Web. *World Wide Web*: 241-255.

Czinkota, M. R. (2000). *Marketing Best Practices*, The Dryden Press Harcourt College Publishers.

Czinkota, M. R. (2001). *Marketing Management*.

Davies, C. (1994). *Cognitive Aspects of Human-Computer Interaction for Geographic Information Systems*. NATO Advanced Research Workshop on Cognitive Aspects of Human Computer Interaction for GIS, Palma de Mallorca, Spain, Kluwer Academic Publisher.

Decker, S., D. Fensel, F. v. Hamrelen, I. Horrocks, S. Melnik, M. R. Klein and J. Broekstra (2000). *Knowledge representation on the web*. Proceedings of the International Workshop on Description Logics, Aachen, Germany.

Delone, W. H. and E. R. McLean (1992). "Information systems success: the quest for the dependent variable." *Information Systems Research* **3**(1): 60-95.

Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. Numerische Mathematik, 1, 269-271.

Frank, A. (1999). "Material for a Post-Graduate Course." *Geographic Information Systems* Vol.3: GIS Organization.

Frank, A. and A. Krek (2003). *Transaction Costs*. AGILE Conference, Lyons France.

Frank, A. a. G. A. (2002a). "What is relevant in a dataset." *5th AGILE Conference, Palma*.

Frank, A. a. G. A. (2002b). *What is relevant in a dataset*. 5th AGILE Conference, Palma.

Fröhlich, M. and J. Mühlig (2002). Usability in der Konzeption. *Usability Nutzerfreundiches Web-Design*. V. v. G. Markus Beier. Berlin.

Gluck, M. (1991). Making Sense of Human Wayfinding: Review of Cognitive and Linguistic Knowledge for Personal Navigation with a New Research Direction. *Cognitive and Linguistic Aspects of Geographic Space*. D. M. Mark and A. U. Frank. Dordrecht, The Netherlands, Kluwer Academic Publishers**:** 117-135.

Hauser, J. R. and D. Clausing (1988). "The House of Quality." *Harvard Business Review* Vol 66(No. 3): pp. 63-73.

Hynek, T. (2002). User experience research- treibende Kraft der Designstrategie. *Usability Nutzerfreundliches Webdesign*. Berlin, Springer Verlag**:** 43-59.

Jakob Nielsen, N. N. G. (1993). *Usability Engineering*. Cambridge, Academic press limited.

Jakob Nielsen, N. N. G. (2000). *Designing Web  Usability: The Practice of Simplicity*, New riders Publoishing.

Krek, A. (2002). *An agent-based model for quantifying the economic value of geographic information*. Vienna, Technical University Vienna.

Lynch, K. (1960). *The Image of the City*. Cambridge, Mass., The MIT Press.

Maslow A. (1943)A Theory of Human Motivation.Psychological Review, 50, 370-396

Mayhew, D. J. (2002). *The Usability Engineering Lifecycle - A practitioner's handbook for user interface design*. San Francisco, California, Morgan Kaufmann Publlishers Inc.

Naumann, F. (2002). *Quality-Driven Query Answering for Integrated Information Systems*. Berlin, Springer Verlag.

Olken, F. and D. Rotem (1990). *Radom sampling from database files: A survey*. International Conference on Scientific and Statistical Database Management (SSDBM), Charlotte, NC.

Ossterbauer, C. (2002). *CAWUT Computer-Aided-Usability-Testing*. Vienna, University of Vienna.

Oswalder, G. (1996). *Datenqualität - Merkmale zur Beurteilung der Brauchbarkeit von Geodaten für Geographische Informationssysteme*. Wien, University of Vienna.

Piaget, J. (1954). *The construction of reality in the child*. New York, Basic Books.

Rasmussen, J. (1986). *Information processing and human-machine interaction*. New York, Elsevier Science Publishing Co. Inc.

Raubal, M. (2002). "Wayfinding in Built Environments." *IfGIprints* **14**(26): 87-105.

Redman, T. (1996). *Data Quality for the Information Age*. Boston, London, Artech House.

Russell, S. J. and P. Norvig (1995). *Artificial Intelligence*. Englewood Cliffs, NJ, Prentice Hall.

Siegel, A. W. and S. H. White (1975). The development of spatial representations of large-scale environments. *Advances in child development and behaviour: Vol. 10.* H. W. Reese. New York: Academic Press**:** 9-55.

Staudinger, M. and B. Haselgrüber (2002). *Die Genauigkeit der Ortsbestimmung mit Mobilfunkgeräten bei der automatischen Standortbestimmung in Notfällen*. AGIT, Salzburg.

Tayi, G. K. and D. p. Ballou (1998). Examing data quality. *Communications of the ACM*. **41 (2):** 54-57.

Ullmann, J. D. (1988). *Principles of Database and Knowledge Base Systems, Vol. 1 and 2,* Computer Science Press.

Vliet, V. (1988). "Organisatie, informatie en systeemontwerp." *Wolters- Noordhoff B. V. Groningen.*

Wang, R. Y. and D. M. Strong (1996). "Beyond accuracy: What data quality means to data consumers." *Journal on Mangement of Information Systems* **12**(4): 5-34.

Wang, R. Y., D. M. Strong, B. K. Kahn and Y. W. Lee (1999). *An information quality assessment methodoligy*. Proceedings of the International Conference on Information Quality, Cambridge MA.

Wang, S., W. Shi, D. Li, D. Li and X. Wang (2003). *A Method of Spatial Data Mining Dealing with Randomness and Fuzziness*. 2nd international Symposium on Spatial Data Quality, Honkong, Advanced Research Centre for Spatial Information Technology.

Wenk, U. (1996). *Usability von Computer-Programmen und der Zusammenhang zwischen Usability und Einstellung von Anwendern zur verwendeten Software*. Wien, Universität Wien.

Whitefield, A., A. Esgate, I. Denley and P. Byerley (1993). On distinguishing work tasks and enabling tasks. *Interacting with computers 5*. **3:** 333-347.