Quality of Geographic Information

Development of a System for Quality Labeling of Map Data for Routing Applications

Eva-Lotta Bengtsson
Hanna Tuvesson

Department of Real Estate Management
Lund Institute of Technology
Lund University, Sweden

Avdelningen för Fastighetsvetenskap
Lunds Tekniska Högskola
Lunds Universitet

ISRN LUTVDG/TVLM 01/5058 SE

Preface

This Master of Science Thesis was produced at Itinerary Systems IS AB in Lund, Sweden, during the fall and winter of 2000/2001. The Master of Science Thesis is part of the requirements for a Masters of survey engineering at Lund Institute of Technology. The academic supervisor has been Lars Harrie at the Department of Real Estate Management, Lund Institute of Technology. The examiner is Åsa Knutson at the Department of Real Estate Management, Lund Institute of Technology. The supervisor at Itinerary Systems IS AB has been Jonas Sellergren.

We would like to thank our supervisors Lars Harrie and Jonas Sellergren for their support and assistance, and our examiner Åsa Knutson for her review of the report. We would also like to thank David Svensson and Itinerary Systems IS AB for giving us the opportunity to carry out the work at the company. Finally, we would like to thank the employees at Itinerary Systems IS AB for their help and support.

Lund, March 2001

Eva-Lotta Bengtsson Hanna Tuvesson

Summary

The objective of this Master of Science Thesis is to develop a quality labeling system for map management in the field of location-based services. Itinerary Systems IS AB (ISAB) is a company in Lund, Sweden, providing hardware and software for location-based services. The foundation for the services is a map with objects and attributes enabling routing. In order to provide high quality services the quality of the map must be high.

To implement a quality labeling system at ISAB the guidelines of ISO 9000 *Quality management* must be considered. Quality management is performed by an organization to ensure that its products and services conform to customers' requirements. ISO 9000 focuses e.g. on documentation of work processes in order to build confidence for the final product to fulfil requirements on quality.

The quality of a map is assessed by comparing the map with the real world or another representation of the real world. There are many factors contributing to the quality of a map. Some of the most important ones are methods for data capture, transformations of the data and conversions between data formats. There are several types of quality. Functional quality describes to which degree the map meets requirements of a certain application, while descriptive quality provides an overall quality measure of the map. The map should be labeled with quality information using a set of quality parameters, each corresponding to one aspect of quality, such as positional accuracy, completeness and resolution.

Standards on quality of geographic information are important e.g. to enable quality labeling in a uniform way. For Swedish applications, the most important standards within this area are: ISO 19113 *Geographic Information – Quality Principles*, CEN ENV 12656 *Geographic Information – Data Description – Quality*, and the Swedish Technical Framework. These standards all provide a quality model with a set of appropriate quality parameters. A comparison of these quality models shows that they are very similar and can all be used for quality labeling of maps for routing.

A closer examination of the quality parameters points out which parameters are most important for routing applications. Producers of geographic information quality label their data to various extent. This kind of information is often considered confidential, which makes it difficult for us to state if the most important quality parameters, for this application, are used by the producers. The metadata editor of ArcInfo 8 has been studied to see if it could create quality parameters to improve insufficiently labeled data. It turned out that the editor is not useful for this application.

When developing a system for quality labeling of ISAB's maps, the steps of the map generating process first must be identified. This is done by establishing work instructions for the process. Secondly, every step is evaluated to determine their effect on map quality. Finally, a method for presenting the map quality should be established. The method should provide enough quality information to enable an evaluation of how to improve the maps. The method should also be an aid when evaluating the quality, and thereby, the usefulness of a potential new dataset.

Sammanfattning

Målsättningen med detta examensarbete är att utveckla ett kvalitetsmärkningssystem för karthantering inom området positioneringstjänster. Itinerary Systems IS AB (ISAB) är ett företag i Lund som utvecklar hård- och mjukvara för denna typ av tjänster. Till grund för tjänsterna ligger en digital karta som möjliggör ruttning. För att kunna erbjuda tjänster med hög kvalitet är det viktigt att kartan är bra.

För att implementera ett kvalitetssäkringssystem på ISAB måste man ta hänsyn till de riktlinjer som ges i ISO 9000 *Quality management*. Kvalitetsledning säkerställer att produkter och tjänster håller den kvalitet som kunder kräver. ISO 9000 fokuserar bland annat på dokumentation av arbetsprocessen. Detta ska skapa ett förtroende för att den slutliga produkten uppfyller kvalitetskraven.

Kvaliteten på en karta kan bedömas genom en jämförelse med verkligheten, eller med en annan representation av verkligheten. Några av de faktorer som spelar störst roll för kartans kvalitet är hur data samlats in, de transformationer som har utförts på data samt konverteringar mellan olika dataformat. Det finns många olika sätt att se på kvalitet. Funktionell kvalitet anger om kartan har tillräckligt hög kvalitet i en viss tillämpning, medan deskriptiv kvalitet ger en helhetsbild av kartans kvalitet. Kartan bör märkas med kvalitetsuppgifter med hjälp av kvalitetsparametrar. Dessa representerar olika egenskaper, såsom positionsnoggrannhet, fullständighet och upplösning.

Standarder för kvalitet på geografisk information har stor betydelse när det gäller att bland annat utveckla en enhetlig kvalitetsmärkning. För tillämpningar i Sverige finns det tre standarder som bör beaktas: ISO 19113 *Geographic Information – Quality Principles*, CEN ENV 12656 *Geographic Information – Data Description – Quality*, samt det *Svenska Tekniska Ramverket*. Standarderna innehåller en kvalitetsmodell med kvalitetsparametrar lämpliga att använda för märkning av geografisk information. En jämförelse av de tre kvalitetsmodellerna visar endast små skillnader och att de kan alla användas för kvalitetsmärkning av kartor för ruttning.

En noggrannare undersökning av kvalitetsparametrarna visar vilka parametrar som är av stor vikt för ruttningskartor. Producenter av geografisk information märker sina data i olika utsträckning. Dessutom är den här typen av information ofta konfidentiell, vilket gör att det har varit svårt att ta reda på om producenterna märker sina data med de parametrar som är viktiga för den här typen av kartor. Metadata-editorn i ArcInfo 8 har utvärderats för att se om den kan utnyttjas för att komplettera otillräckligt märkta data. Det visade sig dock att den inte är tillämplig för detta.

För att utveckla ett system för kvalitetsmärkning av ISABs kartor måste först stegen i kartgenereringen identifieras. Detta görs genom att upprätta arbetsinstruktioner för genereringen. Varje steg utvärderas sedan för att se hur kartkvaliteten påverkas. Slutligen utvecklas en metod för att generera och presentera kartans kvalitet. Denna skall ge tillräckligt mycket kvalitetsinformation för att möjliggöra en utvärdering av hur kartan kan förbättras. Metoden skall dessutom fungera som ett hjälpmedel vid utvärdering av kvaliteten på, och därigenom också nyttan av att lägga till, ett nytt dataset.

Abbreviations

CEN	Comité Européen de Normalisation, the European standardization organization
CEN/TC 287	CEN Technical committee 287: Geographic information
CEN ENV 12656	Geographic Information – Data description – Quality, prestandard developed by CEN/TC 287
EN	European standard
ENV	European prestandard
EUREF 89	European Terrestrial Reference Frame 1989
EXPRESS	ISO 10303-11, which is a standardized modeling language
FGDC	US Federal Geographic Data Committee
GDF	Geographic Data File
GIS	Geographic Information System
GPS	Global Positioning System
GRS 80	Geodetic Reference System 1980, ellipsoid
ISAB	Itinerary Systems IS AB
ISO	International Organization for Standardization
ISO/TC 211	ISO Technical committee 211: Geographic Information/Geomatics
ISO 19100	Geographic information, series of new ISO standards
ISO 19113	Geographic Information – Quality Principles, standard developed by ISO/TC 211
ITRF	IERS Terrestrial Reference Frame
NVDB	Swedish National Road Database (Nationella Vägdatabasen)
RH 70	Swedish National Height Reference Network 1970
RN 92	Swedish National Geoid Height System 1992
RR 92	Swedish National Reference System 1992
RT 90	Swedish National Triangulations 1990, for positions in a plane
SIS	Standardization in Sweden
Stanli	Standardisering Landskapsinformation, SIS project working with geographic information standards
STEP	ISO 10303-21, which is the Standard for Exchange of Product Model Data, a file format for transfer of EXPRESS models
STR	Swedish Technical Framework (Svenska Tekniska Ramverket)
SWEREF 93	Swedish Reference Frame 1993
SWEREF 99	Swedish Reference Frame 1999
UML	Unified Modeling Language
WGS 84	World Geodetic System 1984, ellipsoid and reference system
XML	eXtensible Markup Language

English – Swedish glossary

Quality parameter	Kvalitetsparameter
Quality overview element	Övergripande kvalitetselement
Quality element	Kvalitetselement
Quality subelement	Underordnade kvalitetselement

Quality parameters

Accuracy	Noggrannhet
Commission	Övertalighet
Completeness	Fullständighet
Homogeneity	Homogenitet
Last update	Senaste uppdatering
Lineage	Historik
Logical consistency	Logisk konsistens
Omission	Brist
Origin	Ursprung
Positional accuracy	Positionsnoggrannhet
Precision	Precision
Purpose	Syfte
Rate of change	Förändringshastighet
Resolution	Upplösning
Source	Källmaterial
Temporal accuracy	Temporal noggrannhet, aktualitet
Temporal lapse	Ledtid
Temporal validity	Temporal giltighet
Thematic accuracy	Tematisk noggrannhet
Usage	Användning

Components of the Swedish Technical Framework

Spatial model	Rumslig modell
Positional model	Lägesmodell
Quality model	Kvalitetsmodell
Metadata model	Metadatamodell

Table of contents

1 Introduction

1.1 Quality needs

The process of producing maps is, due to modern technology of today, not as difficult as it used to be. Digital geographic data are available from several data suppliers. This data can be processed and compiled in one of the many GIS software on the market to create a nice-looking map. However, it can be difficult to tell how good it is.

Presenting quality is becoming more important, and data producers are increasingly required to quality label their data. It is also important that all producers label data uniformly to enable customers to evaluate if a dataset meets the requirements of a specific application. During the last decades the need for uniform quality labeling of geographic data has been given much attention by standardization organizations. Standards providing guidelines for presenting quality of data are being developed. Also standards for evaluating quality, and achieving quality, by help from a quality system, are being developed.

A quality system is implemented by documenting the work processes creating the data, with the objective to attain knowledge of accessible data and if the data are suitable for a certain application. It is thus a method for creating confidence in the final product, since it gives awareness of the process steps. It might also serve as a competitive tool, since it gives a company the ability to state a specific quality of their products and services.

1.2 Routing

Location-based services inform users about different entities (e.g. points of interest and mobile users) and events (e.g. traffic disturbances) having a specific location. A typical application is to provide information about a mobile user's position. This information may answer questions like:

- How do I get as fast as possible from here to X?
- Where is the closest restaurant?

The location-based services can be accessed via a mobile network or the Internet, and be presented to the user as text or graphics.

Routing is one type of services included in the concept location-based services. Routing includes the techniques and the material required for finding the correct route between two locations. The techniques include calculation of routes, based on network analysis. The optimal route can be found with respect to either time or length, and presented to the user requesting the service, as text or graphics.

The base material for route generation is a map, which merely consists of a road network and a description of its connectivity properties. However, to give correct routes more road related information is required, such as traffic regulations. This map will in this Master of Science Thesis be denoted a routable map.

1.3 Itinerary Systems IS AB

Itinerary Systems IS AB (ISAB) is a privately owned company in Lund, Sweden. ISAB's main objective is to develop software and hardware for location-based services. These types of services are expected to grow rapidly within the next few years as part of the mobile Internet revolution. The company has developed two platforms to address the localized market. One is a route- and positioning server platform for location support for mobile users and the other is a navigation client platform for mobile equipment to be used for navigating or positioning applications.

ISAB receives geographic data from different sources and compiles them to a map. This final map consists of a road network with traffic regulations and additional sets of supplement information, such as companies, road capacities and regions (e.g. municipalities and built-up areas). The final map constitutes a foundation for the services offered by ISAB.

ISAB wants the services to be good, i.e. to have high quality. High quality means that the services should be competitive and meet the customers' demands. The quality depends on both the map data and the software generating routes for the services. The software can be said, assuming perfect map data, to generate a correct route in practically every case. However, the map data are certainly not perfect, which means that there is a need to know the quality of the data in order to give a complete quality measure of the services.

This means that ISAB wishes to develop a quality system for management of the map data. The quality system should help to derive quality measures of the map data, and contain procedures for checking if quality requirements are met with.

1.4 Objectives

The objective of this Master of Science Thesis is to develop a quality labeling system for map management in the field of location-based services. The system will ensure use of appropriate map data as a foundation for the services. The quality system should be based on accepted principles, and included quality labeling should be based on accepted quality models.

Quality of geographic information is in this Master of Science Thesis never discussed in a cartographic perspective, e.g. the visual appearance of a map. Instead it is focused on the functional perspective of the map when used in routing applications.

In the presentation of quality standards concerning geographic information, we have chosen only to study the standards we found most useful for this projects. Concerning the quality system, we have limited our efforts to consider quality labeling. We have not put much effort on quality evaluation.

1.5 Structure

This Master of Science Thesis consists of three main parts followed by a summary of our conclusions and a final discussion. Part I describes fundamental aspects of geographic information, part II describes different aspects of quality, while part III describes how a quality system can be implemented. The discussion is concerned with our own reflections about certain issues arisen when working with this project.

Part I – Fundamentals of geographic information
Basic knowledge of geographic information is required to understand map quality. The first chapters address readers not familiar with basic concepts of surveying. Chapter 2 – *The map* introduces the reader to geographic information by giving an understanding of the importance of maps and different aspects of their usage. In Chapter 3 – *Geodetic principles* some fundamentals of geodesy are presented to provide an understanding of discussions about quality in forthcoming chapters. Chapter 4 – *Geographic information* describes what geographic information means, different aspects of data and how they can be captured, stored and used. Knowledge about these processes is important since they greatly affect the quality of the data.

Part II – Quality of geographic information
This part presents different aspects of quality and how these can be used for labeling geographic datasets with quality information. Chapter 5 – *Quality* gives an introduction to quality of geographic information. It provides definitions of quality and an presentation of the wide variety of elements that are used for labeling geographic data with quality information. Chapter 6 – *Standards* presents standards we find important for describing quality of geographic information, to show what is recommended for quality labeling of data. A comparison between the, in our opinion, most applicable quality models is performed to evaluate which quality parameters are most appropriate to use when quality labeling data for routing applications. In Chapter 7 – *Quality parameters important for routing applications* the quality parameters in the standardized quality models are discussed in relation to location-based services. The discussion points out which quality aspects are important for quality labeling this sort of data. Chapter 8 - *Study of quality labeling among data producers* presents how producers of geographic information label their data with quality information, to see if the information matches the aspects of quality found important in Chapter 7. In Chapter 9 - *Study of how to use metadata management for quality information,* the metadata editor of ArcInfo 8 is studied to find its applicability for determining the quality of a source dataset.

Part III – Implementation of a quality system for map management
The final part of this Master of Science Thesis provides principles for developing a system for quality labeling of ISAB's map data. Chapter 10 - *Quality management systems* provides an introduction to quality management systems and guidelines for establishing documentation. Chapter 11- *Development of a quality system* presents the outcome of this project. The chapter comprises descriptions of how the system for quality labeling of ISAB's map data can be implemented.

1.6 Method

For Part I, chapters 2-4, we studied literature within the area and also searched for information on the Internet.

Knowledge about quality of geographic information, chapters 5-6, we acquired by studying geographic information standards, searching the Internet, and for some issues contacted experts. To be able to develop a system for quality labeling of ISAB's data, we made a comparison between the quality models of different appropriate standards concerning geographic information.

From the foundation, built in chapters 5-6, we studied suitable quality parameters for labeling of data for routing applications. We have been in contact with producers of geographic data to study their labeling of data. We studied the metadata editor of ArcInfo 8 to see if any quality information could be derived from the source datasets. The study was carried out by reading manuals and testing the editor.

To identify the steps of the map generating process on ISAB we established work instructions. The instructions are to be based on guidelines in ISO 9000 for documentation, why we for Chapter 10 studied literature on quality systems. For identification of the steps, we studied internal documents and source code for some of the steps involved, we also discussed the matter with employees at ISAB involved in the map generation, and participated in the process. Based on the outcome of relevant quality parameters in Chapter 7, we discussed how the quality of a source dataset can change in relevant process steps. This in order to find a strategy for evaluating the steps affecting quality. Finally, we have suggested how a description of the quality of the final map can be structured.

PART I

Fundamentals of Geographic Information

2 The map

This chapter aims at giving a brief introduction to the theme of this Master of Science Thesis, the map. It presents requirements on maps and their evolvement from the early days of history, in order to provide an understanding of the importance of maps and different aspects of their usage.

2.1 Introduction

Map is a term that has been given a great number of definitions. The International Cartographic Association (1973, p.7) defines a map as "A representation, normally to scale and on a flat medium, of a selection of material or abstract features on, or in relation to, the surface of the earth or of a celestial body." Robinson et al. (1995, p.?) define a map as being "[…] a graphic representation of the geographical setting […]". According to Dent (1999, p.4) one description that defines all maps is: "A map is a graphic representation of the milieu." Milieu includes all aspects of the cultural and physical environment.

A map handles two fundamental elements in the real world: positions and properties of the positions. These elements can be presented in different ways and maps can therefore be divided into three types (Eklundh 1999, pp.293-94):
- General maps, which show spatial relations between various geographic phenomena.
- Maps intended for special purposes, for example nautical charts and aerial charts.
- Thematic maps, which contain presentations of the geographic distribution of properties from e.g. statistical data and inventories.

A map is required to present certain information to enable proper use. In order for the map to function in a geometrically satisfying way and to be put into a greater context it must be standardized. The user of the map must also be informed to what extent it agrees with reality. The following information is required (Pilesjö and Cederin 1998, chap.3):
- the points of the compass
- the map projection
- the geodetic reference system
- the level of generalization and the generalization technique
- the positional accuracy, and
- the year of production.

2.2 History

The oldest still preserved map dates from about 2500 BC. It originates from Ga Sur, close to Babylon, and is a clay tablet showing the rivers Euphrates and Tigris surrounded by mountains. It also shows the points of the compass. (de Blij et al. 1996, pp.23-24)

The history of maps can be divided into several eras according to the knowledge about the earth and the development of mapping techniques. Palm (1998, pp.10-18) points out seven different eras ranging from the beginning of time to the twentieth century:

- – 500 BC. The world was thought of as a plate. Maps were drawn without mathematical basis. Most maps were made out of stone or clay or just drawn in the sand.
- 500 BC – 200 AD. The world was thought of as a sphere. Earth measurements started with attempts to calculate the perimeter. Crates constructed the first globe. Hipparchos created the conical map projection.
- 200 AD – 1300 AD. According to the view of the church, the maps were once again made schematically. The compass was being used also in the western world.
- 1300 AD – 1475 AD. The modern compass was designed. The mapmakers were once again attempting to draw the maps according to what the real world looks like. This was influenced by the importance of the shipping at this time. The cartographic techniques developed rapidly depending on e.g. the development of printing.
- 1475 AD – 1570 AD. Several maps of the world were printed.
- 1570 AD – 1700 AD. The first extensive atlases of the world were created. Large-scale maps were developed for regional use. The surveying technology was developed when e.g. the binoculars, triangulation, the Cartesian coordinate system and the barometer were invented.
- 1700 AD – 1910 AD. More precise measuring techniques made the earth measurements more accurate. This included the measuring unit "meter", introduction of the prime meridian, refined map projections and the photographic technique.

2.3 Digital maps

The use of digital data has developed quickly during the last decades. Map producers used the digital technology in an early stage to maintain maps in a fast and simple way. This was first performed by storing the maps in files and later in databases. In the beginning the map databases were used to steer the plotters in the map production, but it was soon figured out that they could be used to more than just producing traditional paper maps. The process of converting analog maps into digital form started, and also the process of establishing files containing spatial data. (Pilesjö and Cederin 1998, chap.7) In these files it was possible to store point identification numbers, coordinates and perhaps even a marking type code describing what type of object was aimed at. Nowadays, it is possible to connect an unlimited number of attributes to all objects. The information does not even have to be stored in the same database since connections between different databases are feasible.

One of the main advantages of using digital geographic data is the possibility to choose the information and how to present it. Digital geographic data can be used just to watch a specific occurrence, as in the case of a paper map. It can also be used in Geographic Information System (GIS) software where it is possible to work with the data previous to presentation. Using GIS software provides the user with a solid foundation for presentation and decision making e.g. in physical planning. It is easy to structure the information to point out what is relevant and present it to those whom it concerns. A new step of the GIS development is to use geographic data interactively, e.g. in navigation systems.

The rapid development of information technology has increased the demand for computer based map information. The available information has become more comprehensive and therefore also more interesting for services based on map information. The increasing amount of available data also makes it easier than before to find external data. Therefore, it is more common to buy data from others than producing them. This exchange of data requires an increased review of the data before usage, in order to avoid future problems caused by low quality data. It is therefore of great interest to label geographic data with quality information.

3 Geodetic principles

This chapter aims at giving an introduction to some fundamental geodetic principles. Since this Master of Science Thesis concerns quality of geographic information it is of importance to be familiar with the fundamentals of coordinate systems, geodetic reference systems and transformations between systems. An incorrect use of these fundamentals may typically cause misunderstandings and introduce errors into a geographic dataset.

3.1 The shape of the earth

The earth has a very complicated shape. It is best described by an equipotential surface of the gravity field, denoted the geoid. The geoid coincides with the undisturbed mean sea level, not affected by sea currents, winds, ebb and tide, and its prolongation below the continents. The gravity potential is the same all over the equipotential surface. Height above the geoid is usually referred to as height above the mean sea level (H), see Figure 3.1.

The shape of the geoid is difficult to describe analytically. It has been considered to be close to spherical. However, since the distances from the center of the earth to the poles are shorter than the distance to the equator this is not an appropriate simplification. The difference in distance is mainly caused by the rotation of the earth, partly from the rotation itself and partly from a redistribution of masses in the inner of the earth due to the rotation. This is why the geoid is best described by a figure called an ellipsoid of revolution, which is a flattened sphere. The geoid differs a little from a well-fitted global ellipsoid, but not more than +/-100 meters. (Eklundh 1999, p.68) This difference is denoted the geoid height (N). Figure 3.1 shows the three fundamental surfaces and their relation to each other considering height.

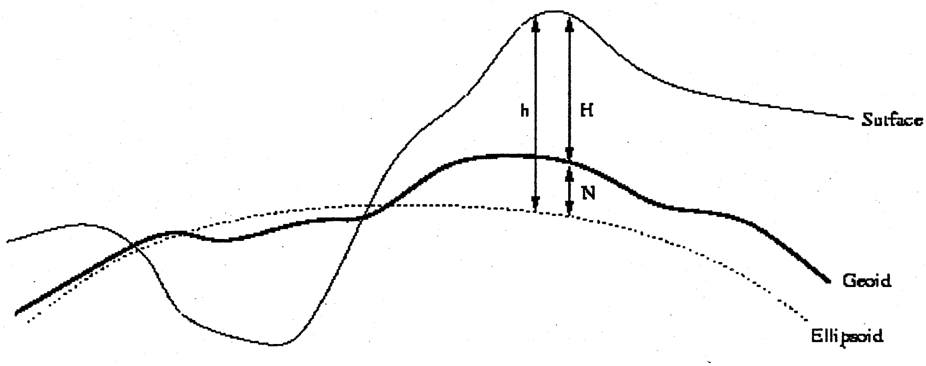

Figure 3.1. Height above the ellipsoid (h), height above the mean sea level (H), and geoid height (N).

3.2 The ellipsoid model

The ellipsoid is, as mentioned above, the geometric figure that best fits the geoid. An ellipsoid is uniquely defined by giving the values of the semi-major axis (a), which corresponds to the distance from the center of the earth to the equator, and the flattening (f). The ellipsoid can, however, also be defined by (a) and the semi-minor axis (b), which corresponds to the distance to the poles. For illustration see Figure 3.2. The relation between these parameters is (Lantmäteriverket 1996a, p.107):

$$f = (a-b)/a. \qquad \textit{(equation 3.1)}$$

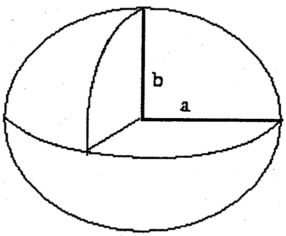

Figure 3.2. The ellipsoid: semi-major axis (a) and semi-minor axis (b).

The internationally adopted earth ellipsoid model is the Geodetic Reference System 1980 (GRS 80) ellipsoid. (Lantmäteriverket 1996a, p.108) Its surface provides an average fit of the geoid. The values of the semi-major axis (a) and the flattening (f) for the GRS 80 ellipsoid are (NIMA 2000, chap.7 p.2):

$$a = 6378137.0 \text{ meters} \qquad \textit{(equation 3.2)}$$
$$f = 1{:}298.257222101.$$

Another global reference ellipsoid is the World Geodetic System 1984 (WGS 84) ellipsoid, which is virtually identical with the GRS 80 ellipsoid. The WGS 84 ellipsoid's parameters are adopted from GRS 80. A change of precision in the definition of GRS 80 has however led to a small difference in the flattening of $16*10^{-12}$, which corresponds to 0.1 millimeters in the semi-minor axis. (Lantmäteriverket 1996a, p.108; Hoffmann-Wellenhof et al. 1997, p.31) Both the GRS 80 ellipsoid and the WGS 84 ellipsoid are derived from satellite measurements performed 1976-1979. (NIMA 2000, chap.3 p.2) The best fitting global ellipsoid known today (January 2000) differ with 54 cm in the semi-major axis from the WGS 84 ellipsoid, and thus also from the GRS 80 ellipsoid. (NIMA 2000, chap.6 p.1)

In Sweden Bessel's ellipsoid from 1841 is used. Its semi-major axis is about 750 meters smaller than the axis of the global ellipsoids. Bessel's ellipsoid is defined by the semi-major axis (a) and the flattening (f) (Lantmäteriverket 1996a, p.107):

$$a = 6377397.155 \text{ meters} \qquad \textit{(equation 3.3)}$$
$$f = 1:299.1528128.$$

The global reference ellipsoids GRS 80 and WGS 84 are geocentric. This means that the ellipsoid center is where the earth has its center of gravity and it has a form and size that agrees as well as possible with the geoid. In addition to the global reference ellipsoids many ellipsoids are defined to fit the geoid shape locally, e.g. Bessel's ellipsoid in Sweden. The reason for this definition is that a global reference ellipsoid may not be fit to use for mapping purposes on national basis. As local ellipsoids are defined to fit the geoid shape in one particular portion of the earth it can differ several hundred meters elsewhere.

3.3 Coordinate systems

There are three common ways to describe the position of a point. The position of a point in space can be described using a three dimensional Cartesian coordinate system (X,Y,Z), with its origin in the center of the ellipsoid and the z-axis coinciding with the rotation axis pointing upwards. The x-axis is located in the intersection line of the equatorial plane and the meridian plane through Greenwich, and the y-axis is oriented orthogonal to both x and z. The second way of describing the position of a point on the surface of the ellipsoid is to use a geodetic coordinate system (lat,lon,h). Latitude (lat) gives the angle between the equatorial plane and the normal to the surface of the ellipsoid in the sought position (P). Longitude (lon) gives the angle in the ellipsoid origin between the meridian plane through Greenwich and the meridian plane through (P). Height (h) is the height above the ellipsoid surface. The Cartesian and geodetic coordinate systems are illustrated in Figure 3.3.

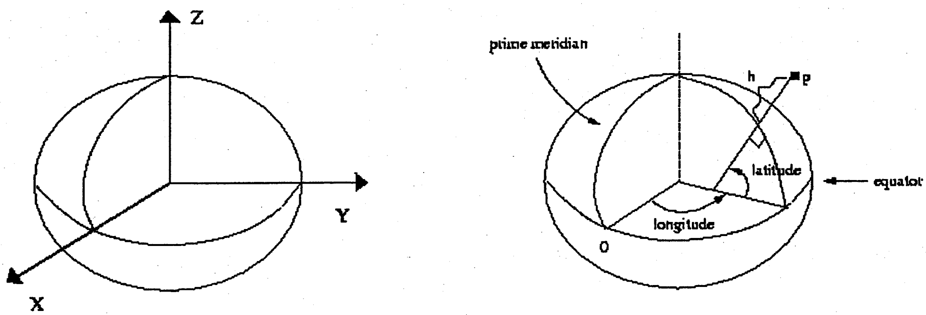

Figure 3.3. The Cartesian and geodetic coordinate systems of an ellipsoid.

In the third type of coordinate system coordinates are given in a map projection plane (x,y). To be able to give a position in three dimensions, height (H) is used to give the altitude above the geoid.

3.4 Map projections

A map projection is (most often) a mathematical representation of the bulging surface of the earth onto a plane. More specifically, a map projection represents the ellipsoid surface or a part of it by a plane surface. Mathematically this projection can be generally defined by:

$$(x,y) = f\,(lat, lon) \qquad \textit{(equation 3.4)}$$

where:
f is a function
x,y are Cartesian coordinates on the map plane, and
lat, lon are the geodetic latitude and longitude.

3.4.1 Properties of map projections

No map projection can project the surface of the earth onto a plane without distortions. Therefore different map projections have been defined, holding different combinations of properties. A map projection can result in maps with e.g. correct areas or correct angles. Area correct maps have the characteristics that every entity is projected with a correct scale in relation to every other entity. Angle correct maps, also called conform maps, enable measurements of relative directions over short distances and the scale is the same in all directions. Area correctness and angle correctness can never be combined in a map projection. The desired property that every part of the earth should be represented with correct areas and that every small figure should retain its geometric shape in the map is thus impossible to achieve. However, in e.g. the Transverse Mercator projection, described in Section 3.4.3, angle correctness is to a certain extent combined with length correctness. The outcome is a map where the central meridian is represented with correct scale.

3.4.2 Types of map projections

Map projections can be categorized as planar, cylindrical or conical projections. (de Blij et al. 1996, p.26) These types are illustrated in Figure 3.4. In planar projections the ellipsoid is projected directly onto a plane. In the two others the projection is made onto a cylinder or a cone, which then is cut and unfolded to a plane. If the plane's normal respectively the axis of the cone and the cylinder are in the ellipsoid's rotation axis the projection is denoted normal. If the plane's normal respectively the axis of the cone and the cylinder are in the equatorial plane the projection is denoted transversal (see Figure 3.5), and in any other way denoted oblique. (Ussisoo 1977, p.24)

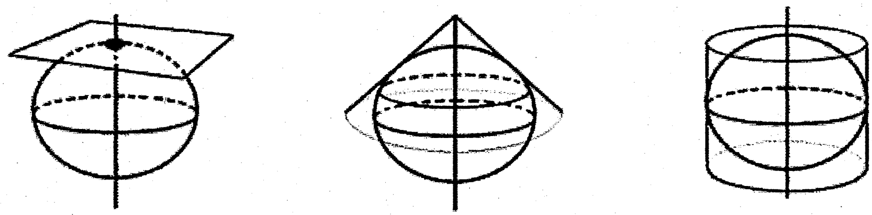

Figure 3.4. Normal planar, conical and cylindrical map projections. (Lantmäteriet 2000a)

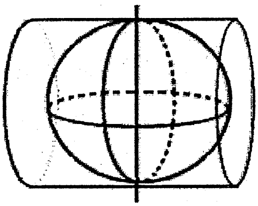

Figure 3.5. Transversal cylindrical map projection. (Lantmäteriet 2000a)

The three categories of projections gain certain properties due to how the earth "touches the map". In planar projections a plane touches the earth ellipsoid at a singe point. This creates a wheel-like symmetry around the point of tangency between the plane and the earth. Planar projections are mostly used when mapping the Polar Regions. The result is a map with the pole in the very center of the map from which the meridians diverge as straight lines. A conical projection produces a map without distortions along the tangent parallel. If the tip of the cone is placed directly above the North Pole this projection is suitable for wide areas (in east west direction) at middle latitudes, such as Europe. A cylindrical projection produces a map where the parallels and meridians appear as straight lines intersecting at right angles. If the projection is normal the cylinder will be tangent to the equator, and if the projection is transversal it will be tangent to the earth along two meridians, one on each side of the earth. One example of the transversal cylindrical projection is the Transverse Mercator projection. The Universal Transverse Mercator (UTM) projection is a system of projections based on Transverse Mercator. (de Blij et al. 1996, pp.26-28; Hofmann-Wellenhof et al. 1997, p.287)

3.4.3 Transverse Mercator – Map projection in Swedish official maps

For official mapping in Sweden the Transversal Mercator projection is used. This projection is also called Gauss conform projection or the Gauss-Krüger projection. For a certain Transverse Mercator projection one of the two tangent meridians is the central meridian. The resulting map will be length correct only along the central meridian with increasing distortions as the distance from this meridian increases. When mapping Sweden, using one single projection, the distance to the central meridian would be up to 400 kilometers. This would cause an error of as much as 2 meters per kilometer in the very eastern and western parts of the country. (Lantmäteriet 2000a)

In particularly large-scaled mapping it is important to reduce the projection distortions. A reduction is achieved by using different central meridians depending on where the area to be mapped is located. Sweden is therefore divided into six different projection zones with a longitudinal difference of 2.5 gon. These zones are named after the central meridian; 7.5 gon V, 5.0 gon V, 2.5 gon V, 0.0 gon, 2.5 gon O and 5.0 gon O.[1] The zones are illustrated in Figure 3.6. The meridian situated through the old observatory in Stockholm is the locally defined prime meridian and corresponds to 0.0 gon. For mapping on national basis the 2.5 gon V projection is used, which is equal to15° 48′ 29.8′′ east of Greenwich. (Lantmäteriverket 1996a, p.111)

[1] V (Väst) means West and O (Ost) means East.

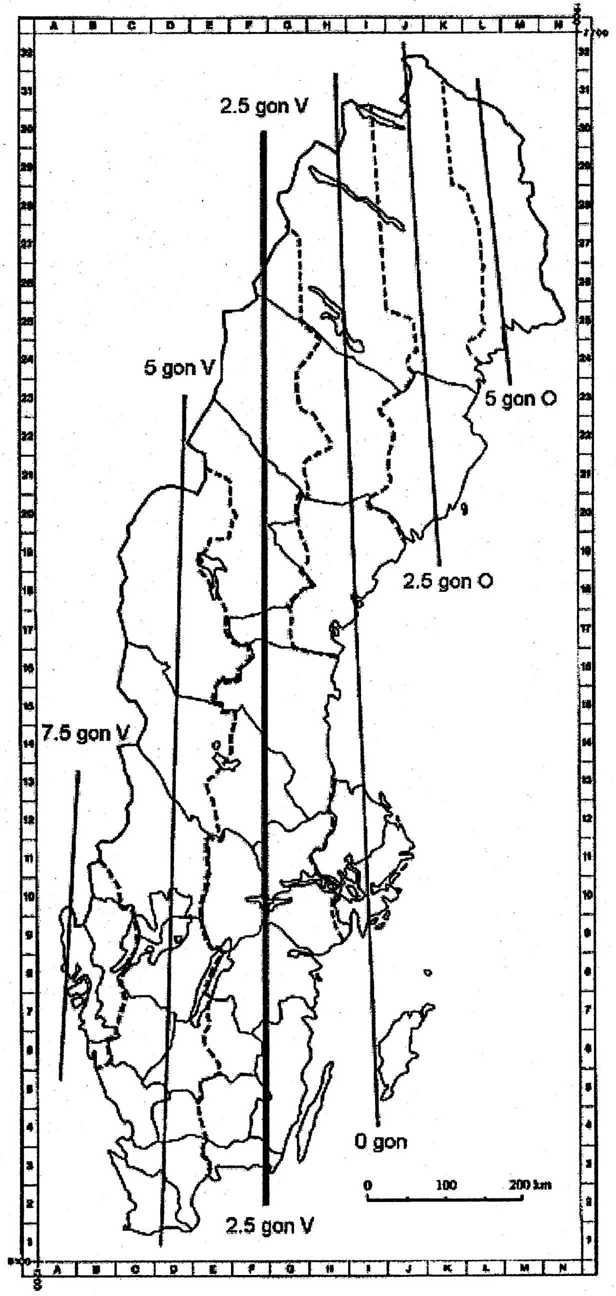

Figure 3.6. Sweden is divided into six projection zones for mapping purposes. (Lantmäteriet 2000a)

3.4.4 Universal Transverse Mercator

The Universal Transverse Mercator (UTM) is a system of map projections. The ellipsoid is divided into 60 zones with a longitudinal width of 6° each. Each of these zones is projected with a central meridian in the center of the zone. A scale factor of 0.9996 is applied to the coordinates to avoid the largest distortions in the outer areas of the zones. The zone numbering shown in Figure 3.7 starts with M1, which has a central meridian with a longitudinal value of 177° West of the prime meridian in Greenwich. Next zone is M2 with a central meridian 171° West of Greenwich. The last zone M60 has its central meridian 177° East of Greenwich. (Hofmann-Wellenhof et al. 1997, p.291)

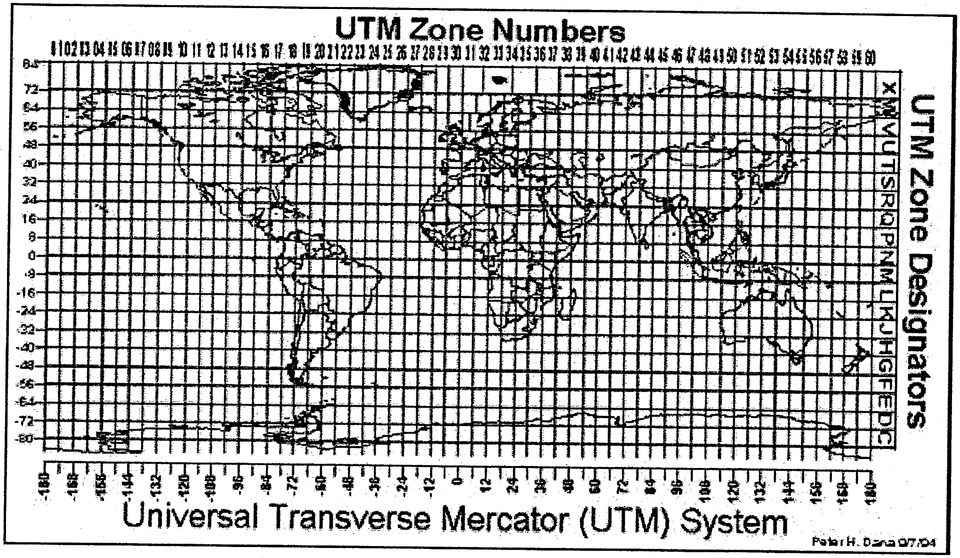

Figure 3.7. The UTM grid zones. (Dana 2000)

Within every zone a planar coordinate system is established. The x value measures the northing and y measures the easting. The x value is counted from the equator, with a 10 000 kilometers false northing on the Southern Hemisphere to ensure positive coordinate values. The y value is counted from the central meridian of the zone with a false easting of 500 kilometers. The Polar Regions are not mapped in the transverse cylindrical projection. Instead a stereographic projection called Universal Polar Stereographic (UPS) is used, which is a planar projection. (Ussisoo 1977, p.96)

The UTM system has in the past been used in Sweden in different applications requiring data exchange with other countries, e.g. when establishing boundaries on land or at sea. The UTM grid system is printed (in blue) on some of the official maps in Sweden in addition to the national reference system RT 90. The basis for this UTM grid system is not commonly stated, but it has traditionally been connected to the European Datum 1950 (ED 50). ED 50 was the result of calculating the first geodetic reference system covering entire Europe. Now this UTM grid is instead usually connected to the WGS 84 reference system. Since this is quite unknown among map users it might result in mistakes due to differences when comparing old and new UTM coordinates. (Lantmäteriet 2000a)

3.5 Geodetic reference systems

To determine a position on the earth it is necessary to refer the position to something. This is why geodetic reference systems are needed. A reference system can be described as a number of points on the ground that has been given coordinates expressed in any coordinate system; geocentric coordinates, geodetic coordinates or map projection coordinates. These known points are called control points. The positions of other points are given in relation to the control points. In many countries a distinction has been made between geodetic reference systems in plane and in height, e.g. the Swedish systems RT 90 and RH 70. However, since three-dimensional measuring techniques are becoming more common, three-dimensional reference systems are also needed.

3.5.1 The Swedish national reference system

The geodetic reference system for positioning in Sweden is divided in two components for plane and height. A third component links the height system to the ellipsoid, enabling positioning in three dimensions.

RT 90 is the Swedish national reference frame for positioning in the plane (x,y). Unlike mathematical coordinate systems the x-axis of this system is directed towards north and the y-axis is directed towards east. Coordinates are given in meters from the equator respective the central meridian of the projection. To avoid negative values west of the central meridian there is a standard addition in y of 1 500 000 meters, the so-called false easting.

RH 70 is the Swedish network for determining positions in height (H). The height given in RH 70 is height above the mean sea level (the geoid) and related to the water level in Amsterdam in the 17th century. (Eklundh 1999, p.78)

When the GPS technology was introduced for positioning, height had to be given in relation to the reference ellipsoid. To be able to relate the height in RH 70 to the height above the ellipsoid there is a need to know the geoid height (N), see Figure 3.1. This is defined in the Swedish national geoid height system RN 92, which gives the geoid height in relation to the reference ellipsoid. These three components (RT 90, RH 70, RN 92) together form the national reference system in Sweden called RR 92, which can be used for positioning in three dimensions.

RR 92 uses Bessel's ellipsoid as reference ellipsoid, fitted to the geoid in Sweden. Since Bessel's ellipsoid is smaller than the global, its center is located 730 meters away from the earth's center of gravity and the semi-minor axis 400 meters away from the earth's rotation axis. (Reit 1994, p.10)

3.5.2 Global reference systems

Global reference systems are defined to constitute a network for positioning worldwide. The aim is to provide accurate coordinates regardless where the measuring is performed. The two most well known global reference systems are ITRF and WGS 84.

The IERS Terrestrial Reference Frame (ITRF) is a global reference system of high accuracy, maintained by the International Earth Rotation Service (IERS). It is a geocentric system with its origin very near (decimeter level) the earth's center of gravity. ITRF is based on the global reference ellipsoid GRS 80 and is defined in more than 180 points around the earth. (Hofmann-Wellenhof et al. 1997, p.31) For a system with a high accuracy it is necessary to consider the effects of plate tectonics, since a tectonic plate can drift as much as 5 cm annually. To maintain the accuracy of the defining points, the coordinates are always stated with a time-label, giving e.g. the GPS week number for the measurement. The ITRF system is regularly updated to correct errors caused by effects of the plate tectonics. (Hofmann-Wellenhof et al. 1997, p.31)

In 1989 there was a campaign with GPS measuring in Europe that led to a European reference system called the European Terrestrial Reference Frame 1989 (EUREF 89). EUREF 89 is a densification of ITRF. Four years later, in 1993, there was yet another GPS campaign. The campaign led to a densification of EUREF 89 in Sweden. (Lantmäteriverket 1996a, pp.108-109) This densification is called Swedish Reference Frame 1993 (SWEREF 93) and is defined in 22 points coinciding with the stations in SWEPOS. SWEPOS is a network of permanent reference stations in Sweden. The stations continually receive the satellite message from the GPS satellites and send corrections to roving GPS receivers to improve their position. The National Land Survey of Sweden plans to change reference system to a new version of SWEREF 93 called SWEREF 99. SWEREF 99 is decided to become the new reference system for all official maps in Sweden. (Jivall and Lidberg)

The World Geodetic System 1984 (WGS 84) is another global reference system, based on the WGS 84 ellipsoid. The system has since its creation in the mid-1980s served as reference system for the satellite system GPS. The WGS 84 system is today realized by the coordinates assigned to the 12 permanent US Department of Defense (DoD) GPS tracking stations. (Malys et al. 1997, p.843) Twice, in 1994 and 1996, improvements have been made of this realization. These improvements are called WGS 84 (G730) and WGS 84 (G830) respectively. 'G' means that the improvements were made using GPS technology and the number designates the GPS week number when the improvements were implemented. (NIMA 2000, chap.2 pp.3-4) After these improvements the WGS 84 system is practically consistent with the ITRF system. In the WGS 84 system no specific time is presented with the set of coordinates of the GPS tracking stations to deal with the effects of plate tectonics. (Malys et al. 1997, p.846) This is a severe shortage considering the intended high accuracy, 5 cm in each of the three components (lat,lon,h). There are no points in Sweden that are assigned coordinates in WGS 84 and only one single tracking station is located in Europe. Therefore it is not possible to determine WGS 84 coordinates of high accuracy in Sweden. However, since the difference between ITRF and WGS 84 is insignificant, SWEREF 93 can be seen as the realization of WGS 84 on the ground in Sweden.

Global reference systems are ITRF, EUREF 89, SWEREF 93, WGS 84 and their densifications. These systems are globally adapted, which means that the difference between them in most cases is less than a few decimeters. With accuracy requirements of less than 1 meter this means that the systems can be considered to be one and the same. If higher accuracy is wanted for applications in Sweden, SWEREF 93 should be used. When SWEREF 99 has been established this will be the choice in Sweden.

3.6 Transformations

Geographic information is stored in different coordinates systems. Coordinates given in one system can be expressed in another system using a transformation. There are many different types of transformations to use when changing between reference systems or coordinate systems. This section is primarily based on HMK-Geodesi, GPS (HMK-Ge:GPS).

3.6.1 Transformation between reference systems

One type of transformation is performed between different reference systems. It requires coordinates to be expressed in the same way for both systems, e.g. with Cartesian coordinates. One example of this type of transformation is transforming from a global reference system, e.g. WGS 84, to a national reference system, e.g. RR 92. The transformation is not mathematically defined. It is instead empirically found by common control points in the two reference systems. The transformation is performed by a three dimensional similarity transformation, called a 7-parameter Helmert transformation, which implies a shift of the system origin, three rotations, one around each axis, and a scale constant. For detail see HMK-Ge:GPS appendix C.3.

The Helmert transformation can also be performed in two dimensions. It is used for defining relations between horizontal reference systems, e.g. RT 90 and local reference systems used in Sweden. It results in a shift of the origin (X_0, Y_0), one rotation (β) and one scale constant (m). The two-dimensional Helmert requires the systems to be orthogonal and have the same scale on both coordinate axes. The transformation parameters are derived by solving the following equations for at least two points with coordinates known in both systems.

$$X = X_0 + m \cdot x \cdot \cos(\beta) - m \cdot y \cdot \sin(\beta)$$
$$Y = Y_0 + m \cdot x \cdot \sin(\beta) + m \cdot y \cdot \cos(\beta) \qquad \text{(equation 3.5)}$$

where:

X, Y = coordinates in the to-system
x, y = coordinates in the from-system
X_0, Y_0 = difference in origin between the two systems
m = scaling, and
β = rotation.

It is recommended to have more than two points for deriving the transformation parameters. It is also important that the points are equally distributed over the area, in order to provide a better coverage of the area to be transformed. With more than two points the equation system to be solved will be over-determined, which gives an estimation of the standard error for the transformed coordinates. This is useful since there is never an unambiguous and correct relation between the systems, due to contradictions between the coordinate values of the known points. (Lantmäteriverket, p.9)

3.6.2 Transformation between coordinate systems

Transformation from one coordinate system to another within the same reference system is mathematically defined. This is the kind of transformation to use if e.g. geocentric coordinates (X,Y,Z) are to be expressed in geodetic coordinates (lat,lon,h). Transformation from geodetic coordinates (lat,lon,h) to plane coordinates (x,y) in RT 90 is also defined in mathematical relations.

The principles for the transformation are shown by giving the equations for a position on a sphere, which has a more simple geometry than an ellipsoid. The position given in a polar coordinate system (lat,lon,R) is expressed in a Cartesian coordinate system (X,Y,Z) according to:

$$X = R \cdot cos(lat) \cdot cos(lon)$$
$$Y = R \cdot cos(lat) \cdot sin\,(lon)$$
$$Z = R \cdot sin(lat)$$

(equation 3.6)

where:
R = radius of the sphere.

Corresponding equations for a position on the ellipsoid are quite similar but take into consideration the difference in radius due to the flattening of the ellipsoid. The position is also mostly located either below or above the surface of the ellipsoid, further complicating the equations. Complete equations for transforming between Cartesian, geodetic and plane coordinate systems for an ellipsoid can be found in HMK-Ge:GPS appendixes C.1 and C.2.

3.6.3 Transformation between projection planes

Transformation between projection planes is also mathematically defined. It is used when two reference systems are not in the same projection plane. Used in the Transverse Mercator projection, the transformation results in a change of central meridian and includes several steps. In the first step, the plane coordinates (x,y) have to be expressed in latitude and longitude, where the longitude is given relative the central meridian in the projection. Secondly, the longitude value (λ) is changed according to:

$$\delta\lambda_t = \delta\lambda_f + \lambda_{0f} - \lambda_{0t} \qquad \textit{(equation 3.7)}$$

where:
$\delta\lambda_t$ = the longitude value relative the central meridian of the to-system
$\delta\lambda_f$ = the longitude value relative the central meridian of the from-system
λ_{0t} = the longitude value relative the Greenwich meridian in the to-system
λ_{0f} = the longitude value relative the Greenwich meridian in the from-system

The coordinates are then in the final step recalculated to plane coordinates (x,y). One example of this transformation is expressing RT 90 5.0 gon West coordinates in RT 90 2.5 gon West. For more detail see HMK-Ge:GPS appendix C.2.

4 Geographic information

This chapter aims at describing what geographic information means, different aspects of data and how they can be stored and used. It also describes different methods of data capture in order to give an understanding of how these methods contribute to the quality of a dataset. Modeling helps to structure the dataset in a certain way to enable use in different applications. The chapter ends with an orientation of the most common data formats used by vendors of software for geographic data to show differences and similarities in methods of storing data.

The term information is sometimes confused with data, but there is a big difference between the meanings of the two terms. The term data refers to a collection of single values about a specific occurrence, e.g. the precipitation of each month during a year. Interpreting or processing data yield information, which then can tell e.g. what month is the driest. Geographic data are data concerned with objects or phenomena directly or indirectly associated with a location relative to the surface of the Earth. Geographic information is then the interpretation of this geographic data.

4.1 Basic data types

4.1.1 Attribute data and spatial data

Geographic data can be divided into spatial data and attribute data. Spatial data describe the position, the shape, and the extension of an object, while attribute data describe the properties of the object. The spatial data are mostly collected from digital maps while the attribute data are taken from some sort of register. Spatial data can be stored either as vector data or raster data. Attribute data are normally stored in separate tables. An identifier tied to both the spatial object and to the attribute data enables the connection between them.

4.1.2 Spatial properties

The spatial properties of geographic objects can be described by:

- geometry, and
- topology.

Geometry includes the properties that can be defined using traditional geometrical methods. Examples of geometrical properties are area, length, volume and shape. Topology gives information about the relations between the geographic objects, such as contiguity and connectivity. (Pilesjö and Cederin, pp21-22) Topology thus deals with the characteristics that remain invariant if the space is deformed elastically and continuously, e.g. when geographic information is transformed from one coordinate system to another. One common topological spatial data structure is the so-called link-node structure, which among others requires that lines are intersected in nodes.

4.1.3 Metadata

Metadata are data about data. They describe the content, quality, condition and other characteristics of data, helping both producers and users in understanding the content of geographic datasets. Due to e.g. the growing number of geographic datasets that are easily accessible over the Internet, metadata have gained importance as a prerequisite to facilitate finding appropriate data.

Metadata are used to organize and maintain an organization's investment in data. Complete metadata descriptions of the content and accuracy of a geographic dataset will encourage appropriate use of the data. It is important that the datasets are documented in a proper way to avoid that only one or a few people have knowledge about the dataset. Everyone should be able to make an assessment of how to use the organization's data. Metadata are also used for facilitating exchange of geographic data between organizations. It is not cost efficient for an organization to produce its data if corresponding datasets already have been produced by other organizations. Making metadata available through metadatabases over the Internet enable data producers to inform about their datasets and users can find sets useful for a certain application.

Metadata can express various levels of detail. The simplest form of metadata is the information attached to a map, given in Section 2.1. Other basic metadata are the information attached to a raster layer giving the number of grid cells, the resolution of the cells, the real world coordinates for the corners of the layer, and the legend to the layer. These descriptive data are normally stored in the beginning of the raster file, as a header, but can also be stored in a separate file. A very extensive presentation of metadata can be acquired using one of the many metadata standards that have been established. They provide a set of metadata elements to use for documentation of a geographic dataset. One example of metadata standards is the Content Standard of Geospatial Digital Metadata, which is presented in Section 6.6.

4.2 Raster and vector

Geographic data may be stored in two ways, as raster or as vector. Conversions between raster and vector data can be made, but it is important to be aware that the conversion always results in reduction of accuracy and possibly also consistency.

The raster format can be described as breaking down reality into a grid and then coding the content(s) of each grid cell to a value. The values of the cells are stored in a matrix, where each object is represented by one or more neighboring cells. The attributes can be linked to the objects if e.g. the object's cell value is an identifier linking to the attributes in a separate table. They can also be linked if the cell values in fact are attributes and the raster database consists of a great number of raster layers. Each attribute would then be represented in a unique layer.

The vector format can be described as georeferenced points with lines drawn between them. All coordinates of an object are given the same identifier, which enables linking the attributes to the object. Points are stored as single coordinate pairs. Lines are made up of two or more points linked together, and can be stored in a list with start coordinates, vertices and stop coordinates. Areas can be stored as the line(s) by which it is surrounded. The topological contiguity relations can be expressed more or less efficient depending on the amount of topological data stored.

Rasterization means going from vector data to raster data and is the simpler one of the conversions. It is performed by calculating the coordinate values to the corresponding position in the raster. A line is rasterized by calculating its equation and marking the raster cells that are crossed by it. A polygon is rasterized by rasterizing the boundaries and then marking all cells that are enclosed by it. One problem with rasterization is that topological errors might arise, e.g. that lines in a network are no longer connected. Another problem is that many entities end up in the same cell, which cause a conflict about which one should be represented. This can be corrected by reducing the cell size. (Eklundh 1999, pp.114-115)

Vectorization means going from raster data to vector data and is more difficult than rasterization. It is performed by thinning the lines into a one-pixel width and then transforming them into a collection of chains of pixels, each representing a line. Every chain of pixels is then transformed into a set of vectors, the number of vectors depends on how wiggly the line is. (Worboys 1995, p.230) The major problem when dealing with vectorization is that the objects are not topologically defined in the raster. Therefore, in order to find and define connected objects, a number of searches of the raster must be performed. Vectorization can not be done automatically since errors occur at a regular basis, instead a fairly large manual effort is needed. One of the most demanding procedures is the assignment of id-numbers when the geometric objects have been defined. (Eklundh 1999, p.115)

4.3 Data capture

Data capture is the process of collecting geographic data. Generally data capture simplifies the infinitely detailed reality to be represented by points, arcs and polygons, and their attributes (Pilesjö and Cederin, chap.3). Data capture can be performed by transforming already existing data from analogue to digital form, by collecting data in the field by geodetic methods, or by interpreting aerial photographs or satellite images.

4.3.1 Analogue data to digital form

Data about the landscape have been captured and presented on maps for a long time. Nowadays geographic data are increasingly stored and used in digital form. In order to take advantage of already existing data it is important to transfer old analogue material into digital form. This can be performed in two different ways; digitalization and scanning. Digitalization results in vector structure of data, while scanning results in raster structure. Digitalization can be further divided into manual and on-screen digitalization.

Manual digitalization is performed using a digitizing table. The table is usually connected to a PC or workstation. A coordinate system is established for the material to be digitized (usually a map or an aerial photograph). The analyst then traces the outlines of areas or marks positions of points and lines with the cursor. There are two basic approaches of how to use a digitizing table. In point-digitalization coordinate pairs are registered by pressing a button on the cursor. In continual digitalization coordinate pairs are registered continually, either according to a constant interval of time or when the cursor has moved a specific distance. Sources of error can be e.g. the resolution of the digitizing table and the analyst's lack of skill. Common errors are that polygons are not closed, or that lines are drawn either too far or too short in the process of linking lines together.

On-screen digitalization works in the same way as the manual digitalization, with the difference that the material to digitize is already in digital form. The digitalization is performed on a scanned map or an aerial photograph that is shown on a computer display. The problems for on-screen digitalization are very much similar to those for manual digitalization. However, on-screen digitalization has the advantages of zooming and contrast augment, which may lead to a better result of higher accuracy.

Scanning an analogue map is a fast way of gaining digital data. However, it requires a great deal of editing since everything on the paper is transferred into digital form, including dirt and changes in color tone of the paper. These kinds of errors are not wanted in the final digital product. The resulting raster structure can be transferred to vector by on-screen digitalization or by automatic conversions that are included in many GIS software.

4.3.2 Geodetic methods

Geodetic methods are used to determine the position of geographic objects in two or three dimensions with high accuracy. This can be performed either by traditional geodetic methods or satellite based measuring techniques.

Traditional geodetic measurements – detailed survey

Detailed survey is used for measuring details on the ground. One application is to determine positions of objects such as houses or real estate boundaries for mapping purposes. Another application is to place an object on a known position, e.g. when building a house or marking a boundary. Detailed survey is usually performed with a total station.

Detailed survey is used when smaller areas are to be mapped and high accuracy is important. It is also applicable when positions can not be derived from aerial photographs due to e.g. dense vegetation. There are several advantages of the method. The measurements can be performed on short notice, the result is of high accuracy and currency, and the risk of misinterpretations is minimized. (Lantmäteriverket 1996c, p.18) Measuring with GPS is an alternative becoming more applicable due to constant refinements of the system resulting in higher accuracy.

Traditional geodetic measurements – control survey

Control survey is used for establishing geodetic reference networks. As stated in Section 3.5, geodetic reference systems are needed when determining a position on the ground, e.g. for detailed survey. The reference system is realized on the ground by points that have been given coordinates in the reference system. These points form a network of control points, a so-called geodetic control network. When new control points are needed it is important that they are given an accurate position since the positions of every other point refer to the control network. The technique used for control survey depends on what kind of network the point will be part of. If the new point will be part of the national reference network it is preferred to use GPS measuring techniques. GPS has several advantages compared to the conventional technique, which uses e.g. a total station, especially over large distances. The most important are the financial aspects and that the accuracy is retained when measuring over large distances. For points belonging to a local reference network, however, a conventional technique may be a better choice. For horizontal control networks in small areas, the two techniques are equal in the sense of accuracy, while for networks in height, GPS can not compete with precision leveling. (Lantmäteriverket 1996a, p.33)

The need of dense geodetic control networks is declining with the increasing use of GPS measuring techniques. The GPS measuring technique will be more and more important in detailed survey, which implies that a new kind of control network will be needed. A network of permanent reference stations will be established to distribute corrections to improve the positions of GPS receivers. There is, however, also shortages with this kind of control network. The control points in the network will be sparsely distributed, which causes deteriorated relative positional accuracy between adjacent GPS measured points.

GPS measurements

The Global Positioning System (GPS) is a satellite based navigation and positioning system. It was constructed by the US DoD for military purposes, but can also be used for civilian purposes. The system consists of 24 operational satellites in orbit around the earth at a height of 20 200 km above the earth's surface and with a period of 12 sidereal hours. (Hofmann-Wellenhof et al.1997, p.12) The satellite configuration enables positioning towards at least four satellites 24 hours a day over almost the entire world. The satellites carry atom clocks serving as reference for the signals transmitted from the satellite at different frequencies. The signals contain information about orbital elements, clock behavior, system time and status message. The signals are modulated with two other signals, the coarse/acquisition (C/A) code and the precision (P) code. The C/A-code is available for civilian users and the P-code for military users. Until May 1 2000, the civilian code was degraded by the use of selective availability (SA). This was a denial of full accuracy, accomplished by manipulating orbital data and the satellite clock frequency. (Eklundh 1999, p.153; USNO NAVSTAR Global Positioning System 2000) The discontinuation of this degradation has the effect that "civilian users of GPS will be able to pinpoint locations up to ten times more accurately" (U.S. White House Office of the Press Secretary 2000). This is not entirely true. The improvement is not valid for measurements on the carrier phase, only for code measurement. These techniques are further described in the next paragraphs.

There are different techniques for how to determine a position using GPS. The main principles for measuring are however the same. Any user of the system has a receiver that collects signals from the satellites. Knowing the distance to and the position of four satellites, the three-dimensional position of the receiver can be calculated. The fourth satellite is used for synchronization of the receiver's clock with the satellite clocks. The most common, and most simple, method for measuring the distance from the satellite to the receiver is code measurement. The accuracy of the code measurement is, with no SA degradation, about ten meters. (SWEPOS, 2001) The other method for measuring the distance from the satellite to the receiver is measuring on the carrier phase. This method is always relative, i.e. uses two or more receivers, and uses either one or two frequencies. The signal from the satellite, which is Doppler shifted, is combined with a signal of the same frequency generated in the receiver. The distance between the satellite and the receiver can be expressed as a number of full carrier wavelengths and a fraction. Comparing the two signals, the fraction of the wavelength can be determined very accurately (millimeter level). (Lantmäteriverket 1996a, pp.9-10) The number of full carrier wavelengths is often calculated by post-processing, but for shorter distances (<10km) the calculations can be performed in real-time. The position can be determined with an accuracy of about two centimeters (1σ). (Eklundh 1999, p.155) There was no significant improvement for measuring on the carrier phase when the SA degradation was removed.

One or several receivers can be used for determining a position with the GPS system. When using only one code receiver the technique is called absolute positioning. The position of the receiver is calculated a few times per second, which is adequate for navigation purposes (Lantmäteriverket 1996a, p.10). When two or more receivers are used the measuring technique is called relative positioning. The receivers measure on the same satellites. If they are placed at a close distance from each other, the errors in the distance measurements will be of about the same size for all receivers. When combining the data from all receivers a higher accuracy can be achieved. The relative method can be used both for code- and carrier phase measurement. The accuracy for code measurement is about 2 meters and for carrier phase about 2 centimeters. (SWEPOS 2001) A special kind of relative positioning is called differential GPS, DGPS. One receiver is then placed at a point with known position and becomes a permanent reference station. This receiver at the reference station registers GPS observation data continually from satellites within its range of contact and calculates corrections based on its known position. The corrections are broadcast to receivers within the range of the station. Relative code measuring could be performed in real time and has in 95 % of all cases of this type of measurement an accuracy better than 2 meters. (Eklundh 1999, p.155) This figure is valid for the service in Sweden based on SWEPOS.

GPS measurements in practice

There are numerous ways of how to take advantage of the GPS technology to capture geographic data. It is common to establish temporary reference stations at control points and use the knowledge of the coordinates of these stations to improve measurements performed by roving receivers. The roving receiver is moved around by a person who registers positions at points of interest. This method is appropriate in relatively small areas with short distances between the points, e.g. for detailed survey. A GPS receiver can also be carried by a vehicle to enable coverage over larger distances in a shorter time. This is common for capture of e.g. data of a road network. The receiver is carried by a car and set to register positions at a constant time interval. When the registered points are drawn on a map, the geometry of the road network will appear.

The accuracy measures for GPS given in the previous section are theoretical and valid for favorable conditions only. When measuring with GPS in the countryside, the accuracy can be as high as stated above if no obstacles disturb the GPS signals. However, when performing measurements in cities, the GPS signals bounce on buildings and other constructions causing errors in the determined positions. This phenomenon is particularly noticeable when collecting positions from a receiver in a car (using only a short period of time for every coordinate registration). The signal bouncing causes the position registered by the receiver to "jump". At certain locations the measurements will indicate that the car has driven several hundred meters instead of perhaps just a few decimeters. When the object causing the bouncing has been passed, the position jumps back again.

4.3.3 Interpretation of aerial photographs and satellite images

Aerial photographs have for a long time been the only way to acquire geographic data over large areas at a low cost. The accuracy of data acquired from aerial photographs is, however, not as high as of data acquired with geodetic methods. During the last decades satellite images have improved in resolution and availability. High-resolution satellites provide images that in resolution are comparable to high-altitude aerial photographs.

Aerial photographs

Aerial photographs enable an overview of large areas, constituting a basis for a fast capture of geographic data. The data will be of lower accuracy compared to conventional geodetic methods, such as detailed survey. It is, however, an appropriate method when data over vast regions are needed in a short period of time at a low cost. The aerial photographing is planned very precisely in order to cover the area navigating along exact flight lines. The height of the aircraft is decided based on the purpose of the photographing, i.e. the desired accuracy of the resulting product and the possibility to interpret entities of interest. The photographing is performed from special aircraft equipped with very stable speed, steering control and GPS positioning in order to take the photographs in the correct position. Photographs are usually taken with an overlap of about 60 % in the flight direction and 30 % across the flight direction creating the basis for stereoscopic examination. (Nämnden för skoglig fjärranalys 1993, pp.37-38)

Geographic information can be extracted from aerial photographs using stereoscopic examination. Stereoscopic examination is a method that restores the conditions of the photographing moment for two adjacent aerial photographs. The common area covered by the two photographs can hereby be experienced in three dimensions. The three-dimensional view facilitates extracting geographic information. One example is to derive heights of objects in the photographs using photogrammetric measurements. The height is acquired by measuring the position of the object of interest in both photographs. With knowledge of e.g. flying height of the aircraft and the principal distance of the camera, the height of the object can be calculated. Information about the landscape can be mapped by digitalization in the stereoscopic examination. Another approach of how to extract digital geographic information is scanning and following on-screen digitalization. This requires that the aerial photographs have been converted to an orthophoto.

Satellite images
Satellite images are used for capture of geographic data covering very large areas in one image. The basis for interpretation of satellite images is that different materials reflect different amounts of sunlight throughout the electromagnetic spectrum. This gives a possibility to separate materials by studying their reflection in different wavelengths. Sensors carried by satellites in space capture the reflection from objects on the earth and detectors sensitive to different wavelengths register the amount of reflection. The amount of reflected energy is translated to a numerical value on a predefined scale, for example 0-255, where 0 denotes low reflection and 255 denotes high reflection for a certain wavelength. The area that is registered by one detector gives the resolution (see Section 5.4.2) of the resulting satellite image.

Common satellite sensors are Landsat TM with a resolution of 30 meters (Campbell 1996, p.174) and SPOT MSS with 20 meters resolution for the multispectral sensor and 10 meters for the panchromatic sensor meters (Campbell 1996, pp.181-182). Data from high-resolution satellites like Ikonos have a resolution of 1 meter panchromatic and 4 meters multispectral. High-resolution satellite images can be used in a similar way to aerial photographs for mapping large areas in a short time. Satellite images are delivered in digital raster form; that is, they can be used for mapping with on-screen digitalization. Satellite images can also be used to derive heights in a stereoscopic examination since the SPOT satellites have the quality of registering two "simultaneous" images covering the same area from different angles.

4.4 Modeling

Modeling is a tool for transferring real world entities into an ideal concept in order to build a dataset using computers. It is then important to remember that computers are finite state machines that work on quantified data. In order to identify and structure the geographic objects and describe them, models are used. (Molenaar 1998, p.2)

A model is an artificial construction in which parts of one domain (the source domain) are represented in another domain (the target domain). The purpose of the model is to simplify and abstract from the source domain. Elements of the source domain are translated by the model into the target domain and viewed and analyzed in this new context. (Worboys 1995, p.145) A model identifies and specifies the elementary building blocks and mutual relationships that are used for this translation. These blocks and relationships constitute the rules that decide how objects in the real world should be linked to the discrete view of the world in the dataset. (Molenaar 1998, p.2)

4.4.1 The modeling process

The modeling process describes how the reality is modeled to be represented by a dataset. It is possible to use models at computer readable level only, but these are very abstract and difficult to understand for the human mind. Therefore several models have been developed to overcome these difficulties and link the abstract models closer to a human viewpoint of entities in the real world. (Molenaar 1998, p.7) The modeling process is thus preferably divided into several steps, which might be (Molenaar 1998, pp.7-8; Lantmäteriverket 1997, pp.33-38):

- Spatial modeling. The first step in the modeling process is to decide which objects should be included in the dataset. A spatial model, showing occurrences of the real world, is designed. Lantmäteriverket (1997) does not use spatial modeling.
- Conceptual modeling. The spatial model is transferred into a conceptual model, which expresses the types of the occurrences, their properties and the relations between them.
- Logical data modeling. The conceptual model is transferred into a logical data model and hence becoming more similar to the ones that can be structured by a computer. One type of logical model is the relational model, which organizes data in tables where one row of the table contains an ordered set of attribute values belonging to an object. Another logical model is the object-oriented model, which handles representations of relevant entities as objects.
- Physical data modeling. The last step in the modeling process designs the physical data model or the implementation model. The computer can handle this model since it is stored in bits and organized in bytes, records and pages.

The most important document of the construction phase of the dataset is the dataset specification. It specifies: data capture methods, object type encoding and storage techniques. (Lantmäteriverket 1997, p.33) The dataset specification serves as an ideal way of representing the real world. This view of the real world, which is implied by the specification, is called the nominal ground (CEN 1998a, p.7), universe of discourse (ISO 2000b, p.4) or abstract universe (Veregin 1999, p.183).

4.4.2 Modeling languages

There are several different languages to use for developing and describing the different models in the modeling process. This section will provide a short description of two such languages.

EXPRESS

EXPRESS is a standardized modeling language, which is used e.g. in the Swedish Technical Framework. EXPRESS is standardized as Part 11, *The EXPRESS language reference manual* of the ISO 10303 standards, which are presented in Section 6.3.3. It looks similar to any programming language and has its strengths in the declaration of types. Models described using EXPRESS are easy to transfer between different systems since it is in ASCII format. The standardized way of transferring EXPRESS models is to use STEP files, which also constitute one part of ISO 10303. There is a graphical equivalent to EXPRESS called EXPRESS-G, which aims to visualize the models in order to provide a better understanding. However, everything in an EXPRESS model can not be shown using the graphical EXPRESS-G language. (Celander 1999, pp.22-23)

UML

The Unified Modeling Language (UML) is "a graphical language for visualizing, specifying, constructing and documenting the artifacts of a software-intensive system." (OMG 2000, Foreword) It has been derived from several other object-oriented modeling languages and become the industry standard language for modeling objects and components. (Celigent 2000) The UML specifies a modeling language that integrates the object-oriented community's agreement on core modeling concepts. (OMG 2000, chap.1 p.7) Basically UML defines a number of diagrams, that can be drawn to describe the functionality and behavior of a system or an application, and also what these diagrams mean. (Standardization and the UML 2001; Celander 1999, p.32)

The application area of UML is similar to the one of EXPRESS, but UML is focused on software, not on physical details. UML covers a wider area than EXPRESS, but does not have as detailed possibilities to express data types. (Celander 1999, p.32)

4.4.3 Transfer of models

The models described with EXPRESS or UML can be transferred between systems. The standard way for implementing the models for transfer has traditionally been to develop a data file format. The transfer is also possible to perform via an interface between databases. A new approach for implementing the models is for transfer via the Internet. This section provides one example of transfer via file format and one example for transfer via the Internet.

STEP

A Standard for Exchange of Product Model Data (STEP) file is used for transfer, initially of information about products, between different systems. The file format is standardized as Part 21, the *Clear text encoding of the exchange structure,* of the ISO 10303 standards, which are presented in Section 6.3.3. The main application for STEP files is to transfer EXPRESS models. The format is defined using an algorithm based on any EXPRESS model. The exact file format is thus depending on the appearance of that particular model. The STEP file consists of a header and a data section. The header contains metadata, giving e.g. the EXPRESS model, and the data section contains data according to the model. A severe constraint is that the file can not be interpreted without access to the used EXPRESS model. (Celander 1999, p.24)

XML

The eXtensible Markup Language (XML) is, according to Celander (1999, p.31), a more general solution to the problem that STEP solves for a part of the IT area. Technically speaking, XML can be said to be a simplified and Internet adapted version of SGML (Standard Generalized Markup Language), which has been used several years to describe and represent documents. To describe the contents of a document XML uses syntax, which works as a specification of a file format. Another syntax is used to define rules for the contents of the document. XML is very similar to STEP, also concerning functionality. The interesting part about XML is that it carries the strength of the Internet. The market for XML-based tools is thus much larger than for STEP based tools. (Celander 1999, pp.31-32)

Part 28 of the ISO 10303 standards - *XML representation of EXPRESS-driven data*, is a new method of implementation aiming at making the information in the EXPRESS models accessible via the Internet. XML is the obvious alternative for this and will for these applications replace Part 21. The basic contents of Part 28 will probably be an algorithm for generating an XML file based on an EXPRESS model. (Celander 1999, p.26)

4.5 Data formats

There are several different formats used to store data in files for transfer between software systems, which may be causing difficulties. There have been attempts to establish standardized formats to solve this problem and to limit the number of formats. However, since commercial companies dealing with geographic information have attached certain formats to their respective software, it is a quite impossible task to settle for just a few data formats. It is therefore important that the software can read and write other formats than just the ones originally attached. This section will provide a description of a few data formats frequently used in geographic information systems.

4.5.1 Shapefile

The Shapefile spatial data format is an open format developed by the software company Environmental Systems Research Institute, Inc (ESRI). It is used to transfer vector data and their attributes between different ESRI software. A Shapefile stores non-topological geometry and attribute information for spatial features in a dataset. The geometry of a feature is stored as a shape comprising a set of vector coordinates (ESRI Inc. 1998, p.1) (vector data), and attributes stored in a table with one record per shape. Since Shapefiles do not store topological relationships between different features and feature classes, getting hold of topological relationships requires searching tables for shared coordinates. However, shapefiles are useful for mapmaking and basic analyses.

A Shapefile consists of three files with the same name but with different extensions (ESRI Inc. 1998, p.2):

- the *.shp main file contains the geometry of the shapes
- the *.dbf dBase table contains attribute data, and
- the *.shx index file contains records storing the offset of the corresponding main file record from the beginning of the main file.

A one-to-one relationship between geometry and attribute is based on record number, which means that the attribute records in the dBase table must have the same order as the geometry records in the main file. The geometry information is stored binary, which gives access to the data rapidly and requires less storage capacity compared to ASCII, but the coordinate files are not readable without interpretation of a certain program. Giving the attributes in separate files enable a quick survey of the attribute table, but requires a search using the index file every time attributes for only a few objects are selected.

Shapefiles support points, lines and area features, where the area features are represented as closed polygons. (ESRI Inc. 1998, p.1) One Shapefile can only hold one type of feature; points, lines or polygons. To create an information system comprised of all three types, several layers are added to a project, each layer containing one Shapefile.

4.5.2 MapInfo Interchange File

The MapInfo Interchange File (MIF) format is an ASCII text file format that fully describes the contents of a MapInfo table. ASCII storage gives a quick survey of the data since everything is readable without interpretation of a computer. The simplicity also has the effect that the MIF format easily can be translated to other formats by other software. However, ASCII format does not provide as rapid access to the data as binary storage does. The (geo)graphic objects that can be specified in MIF are numerous; points, lines, polylines, regions, arcs, texts, rectangles, rounded rectangles and ellipses. (MapInfo Corporation 1999, p.7) The format does not store topological relationships between objects.

The data are stored in two files with the same name but with different extensions, geometric data in a *.mif file and attribute data in a *.mid file. The attribute data are delimited and organized with one record of data per row. The mif file has one header section and one data section. The header specifies the names and types of the attribute data and also optional information, such as used coordinate system and if any attribute in the mid file is indexed. (MapInfo Corporation 1999, pp.2-7) Following the header is the data section, which gives the geometry of the geographic objects. The first object in the mif file is associated with the first row in the mid file, the second object associated with the second row etc. For each object, coordinates of the points included in the object are listed, one pair of coordinates per row.

4.5.3 Data eXchange Format

The Data eXchange Format (DXF) was established by AutoDesk Inc. The format is a de facto standard for exchanging digital blueprints and maps between different CAD systems. Thus, it handles spatial data as graphics rather than as objects. DXF supports most types of geographic objects (called entities), such as points, lines, texts, closed polygons and ellipses. It can store geometric properties for geographic objects in two- and three dimensions, but is in its basic appearance not able to store any topological information. Descriptive information can be connected to single geographic objects through a functionality for Extended Entity Data. (AutoDesc Inc. 2000)

The DXF format is based on layers. Every layer can hold objects belonging to different entities. The layer also holds certain properties, such as color, pen style and width of the lines included in the layer. Essentially, DXF files are composed of pairs of codes and associated values. The code indicates the type of value that follows, such as a text string or point coordinates. Using these group codes and value pairs a DXF file is organized into sections, building the entities. The DXF standard is stored in ASCII, but binary storage is also possible. (AutoDesc Inc. 2000)

4.5.4 Geographic Data File

Geographic Data File (GDF) is a data format used for description and transfer of road networks and road related data. The GDF exchange format is one part of the European GDF standard, which is presented in Section 6.4.4.

The GDF exchange format supports geographic objects in a three level structure. At Level 0 the geometry and topology of the objects (called features) are described using the cartographic primitives nodes, edges and faces. Nodes represent real world point features, edges correspond to line features and faces to area features. Level 1 contains simple features, such as road elements, rivers and boundaries. These features can hold attributes specific for each feature, and relations indicating both relations between features (e.g. road A is in city X) and constraints important for car navigation systems (e.g. forbidden turn from road element 1 to road element 2). At level 1 e.g. all road element of an intersection should be represented. At Level 2 the simple features are aggregated to complex features and e.g. intersections are represented with a single point. This level is used when a simplified description of the road network is sufficient. (Geographic Data Files 2000)

The GDF standard provides an extensive catalogue of features, attributes and relations to use when describing road related data. However, there is also a possibility for user-defined features, attributes and their associated values, and relations. This contributes to a diversity of GDF formats if all vendors specify their own format with subtractions from and/or additions to the standardized basis. This in turn might lead to difficulties for companies using this kind of information. If a system is adapted to one specific GDF format, adjusting to the GDF format of another vendor will require a significant effort. (Geographic Data Files 2000)

4.5.5 Optimal format for storing map data

There are differences and similarities between all data formats. The most important similarity is that all formats have been developed for fulfilling the requirements of a certain application. A data format is created for use in certain software or for transfer of data between certain systems. This implies that the formats differ from each other by means of structure and potential use. It is thus important to use an appropriate data format and have software that can read several formats. If not, the need for converting from one format into another will arise, possibly using several formats in between. This may result in information loss due to the differences between the data formats concerned.

An optimal format for storing map data for navigation purposes, should hold certain properties. The format should require a minimum of storage capacity to enable storage of as much data as possible. The data should also enable rapid access on disk and facilitate fast search, which implies binary storage. The GDF file format is used for description and transfer of road related data. It manages relations between features at a detailed level and contains much information needed for routing. However, GDF is practically never used for storing data since it demands large storage capacity. To ensure access to a data format appropriate for a specific application, one solution is to develop a new data format. The advantage of this approach is that the format can be built to fit specific needs and prerequisites. The disadvantage is that yet another conversion might be needed when importing data from external data sources, increasing the risk of information loss. However, with a thorough specification of the format, it can be designed to manage all possible occurrences and attributes from the external data sources independent of their format.

PART II

Quality of Geographic Information

5 Quality

This chapter aims at giving an introduction to quality of geographic information. There are many aspects of quality and since it is such a wide concept it has different meanings to different people. This chapter provides some definitions of quality and some aspects of how to look upon quality. Quality of geographic data originates from the measuring methods, which were used to capture the data. Therefore, a brief survey of errors and accuracy of measurement is provided. The chapter ends with a presentation of the wide variety of parameters that can be used for labeling geographic data with quality information.

5.1 Aspects of quality

Quality is a term with many meanings. Everyone may have his or her idea of what "quality" is. The term has, however, been defined by standardization organizations worldwide. In plain language, the standardized definition of "quality" in ISO 9000 refers to all those features of a product (or service) which are required by the customer. In the two prestandards for geographic information, ISO 19113 *Quality Principles* and CEN ENV 12656 *Data description – Quality*, quality is defined as: "totality of characteristics of a product that bear on its ability to satisfy stated and implied needs". (ISO 2000b, p.4; CEN 1998a, p.7) The definition of quality of geographic databases is, according to Lantmäteriverket (1997, p.51), the degree of agreement between the total condition of the database and the intentions stated in the database specification. This in all states that quality is something, which is assessed when a dataset is put into relation to some kind of "reality". The reality states what is right and what is wrong, and can be e.g. dataset specifications or customers' needs.

5.1.1 Quality management

Quality of geographic information is not an isolated occurrence. The use of geographic datasets is incorporated into many businesses and the quality must therefore be put into a greater context. ISO 9000 is a family of standards devoted to quality management on business level. Quality management is performed by an organization to ensure that its products and services conform to their customers' requirements. The ISO 9000 standards can be used by any organization for developing a quality system within its business. The aim is assurance of all steps in the process of producing the geographic information, but also to provide knowledge of the fact that all steps in the process contribute to the quality of the final product. The point is, to achieve a high quality product, e.g. a geographic dataset, it is vital that quality thinking pervades the whole production line of a business. The ISO 9000 family of standards is further described in Section 6.3.2 and theory on quality management systems is found in Chapter 10.

5.1.2 Abstractive and representative quality

In contrast to geographic objects there is no meaningful concept of data quality in the real world. Data quality comes into existence only as the result of the abstraction and representation processes of the real world, which in turn are the result of the modeling process. Therefore, there can be a distinction made between abstractive data quality and representative data quality. (Duckham, chap.3) Abstractive quality supports or informs about the linkage between the real world and the nominal ground and is aggregated for all objects of the same kind. It can be described by quality parameters such as an abstraction modifier, which gives the degree of diversity between the real world and the nominal ground. Representative quality on the other hand, records the uncertainty introduced when representing the objects of the nominal ground in a dataset. It is usually applied to single or groups of objects and can be described by parameters like lineage, source and usage.

The continuation of this Mater of Science Thesis is not concerned with abstractive quality, only representative quality will be further discussed. Quality parameters used for describing the representative quality aspect are presented in section 5.4.

5.1.3 Descriptive and functional quality

Two other aspects of quality of geographic information are descriptive and functional quality. Descriptive quality provides information of how good a product is from a quantitative point of view. This quality aspect is the basis of objective assessment of the quality of a product and it is therefore important to present as many relevant quality parameters as possible. Descriptive quality can be expressed using combinations of quality parameters such as accuracy, completeness and consistency. Functional quality, on the other hand, provides information of how good a product is when used for a certain application. A dataset can constitute the basis of many different applications and for each of the possible applications only some of the quality parameters may be relevant.

For better understanding of descriptive and functional quality, a road map used for generating optimal routes for car navigation can be considered. The roads in the map have a positional accuracy of 20 meters, which is not very good, but a complete coverage of correct street names. If the resulting route is based on street names the accuracy weakness of the map is not important. In this specific application the functional quality is better than the descriptive quality of the roadmap. But on the other hand, if the application was to navigate in a city center based on coordinates of the roads in the map, the functional quality would not be that good.

5.2 Quality labeling and assessment

For most people quality of geographic information means accuracy of the position of an object. However, quality is not just concerned with accuracy of geometric representation. To assess the total quality there are many aspects that are important to consider, e.g. how the data were initially captured and what operations have been performed on the dataset. These aspects constitute the foundation for deriving quality measures for the dataset, e.g. the accuracy of the position of an object and if all objects of a specific type are included.

These measures of quality are known as quality parameters or quality elements (see section 5.4) and used for labeling geographic datasets with quality information.

Labeling a geographic dataset with quality information is important for many reasons. It will facilitate finding a dataset that meets certain application needs or requirements. A detailed quality description will also encourage sharing, exchange and use of appropriate datasets. Both producers and consumers of geographic information benefit from quality labeling. The producer can evaluate to which degree a dataset meets the requirements set forth in the dataset specification. The user, on the other hand, can determine if the quality can satisfy the requirements of the user's application. From the user's point of view, the word quality can most often be replaced by appropriateness.

The representative quality of datasets with geographic information is most often assessed by a comparison to a dataset specification. The quality is derived from the difference between the description of real world objects in the dataset specification and their representation in the actual dataset. However, sometimes there is no dataset specification available when evaluating a dataset. In this case a specification must be assumed in order to evaluate the quality of the dataset in relation to the real world, but this is not always possible due to economical reasons. Instead an already existing product containing the objects of interest can be used, assuming that the quality of these objects is better than the intended quality of the dataset to be evaluated. This product might be another dataset, an aerial photograph or the result of a detailed survey. (Stenborg 2000, p.29)

5.3 Basic quality terms

Quality of a geographic dataset is closely related to the method used when capturing the dataset. In this section a short introduction is given to the most basic quality terms that are directly related to measurements. Basic quality terms do not refer to the wide variety of quality parameters, which are recommended by standardization organizations for labeling geographic data with quality information. Basic quality terms simply refer to accuracy and errors of a measurement seen from a statistical viewpoint. These terms are closely related to each other. The resulting dataset of a capturing method is of certain accuracy, due to the fact that all measuring methods are contaminated with errors to some extent.

5.3.1 Accuracy and precision

Two quality terms that are often misunderstood and confused are accuracy and precision. However, statistically there is an important difference between them, which is illustrated in Figure 5.1. Accuracy indicates the measured (or calculated or estimated) values' agreement with the true value, and can be represented by a mean error. Precision refers to the variability of the results of repeated measurements of the same quantity. The precision can be represented by the standard deviation of a measurement, indicating the difference between the measured values and their mean value. High accuracy means that estimated values are consistently close to the true value. High precision means that the variability of the estimated values is low.

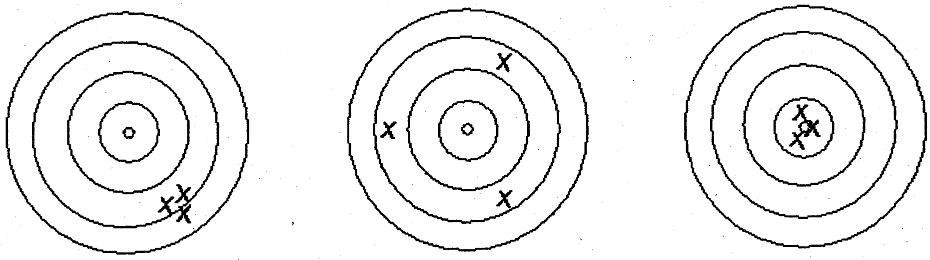

Figure 5.1. The left target shows three shots with low accuracy but high precision, the target in the middle has high accuracy but low precision, and the right target shows both high accuracy and high precision. Redrawn from Eklundh (1999, p.251).

5.3.2 Errors

Whereas accuracy describes the degree of agreement between a measured value and a true value, an error is the discrepancy between these quantities. Usually it is the absolute size of the error that is stated, not whether the discrepancy is positive or negative. The total error consists of three different parts (Lantmäteriverket 1996b, p.14):

- gross errors
- systematic errors, and
- stochastic errors.

Gross errors are related to careless mistakes, insufficient attention or changes in the basic conditions for the measurements when capturing data. By comparing measurements of the same quantities, and analysis during calculations the gross errors can be eliminated to a certain extent. Systematic errors can arise from poorly calibrated measuring instruments or from incorrect coordinate systems when transferring data from one source to another. A dataset affected by systematic errors usually has low accuracy but high precision. All measurements are contaminated with stochastic errors, which are usually small in magnitude. These stochastic errors can be well approximated to belong to the normal distribution, which means that minor errors are most common and that negative and positive errors of the same size are equally common. The size of the stochastic errors gives the quality of the measurement, which is assessed by statistical estimations.

5.4 Data quality parameters

There are many different approaches of what parameters should be used for labeling quality of geographic data. This section provides, in alphabetical order, a presentation and definition of some of the parameters given in geographic information standards (ISO 2000b, CEN 1998a and STG 1998b) and supplementing literature. Some parameters might appear very similar, why it is important to know that they represent different aspects of quality. By combining the parameters, the quality of datasets for different applications can be described.

The data quality parameters are divided into quality overview elements and quality elements. The overview elements provide non-quantitative information of administrative and historical interests. They can be used to derive quantitative information, e.g. may information about how a dataset was captured give an understanding of the positional accuracy. Quality elements on the other hand, are mostly quantitative, expressed as an absolute value with belonging sort or a sort-less ratio. They may, however, also be non-quantitative presented with a boolean variable. The quality elements are often further specified with subelements describing certain aspects of the elements.

5.4.1 Quality overview elements

Homogeneity

Homogeneity supports description of the uniformity of the quality elements in a dataset. It can be either an expected or a tested uniformity. A quality element can have non-homogenous distribution for many reasons, e.g. spatial extent, temporal extent, attribute values. Spatial extent means that the distribution differs depending on geographic location. Temporal extent refers to differing dates of data capture. Finally, attribute values may state that the objects belonging to one type differ from objects of another type. Examples are (CEN 1998a, pp.18-19):

- Horizontal positional accuracy is higher near the coast.
- Logical consistency is highest for the data captured after 1994.
- Completeness is best for cities.

Lineage

Lineage describes the process history of a dataset since its original creation. It includes information about source materials, methods of derivation and what transformations have been applied to the dataset. (Aalders 1997, pp.281-282) All of these processes contribute to the present quality of a dataset. The lineage labeling is intended to be precise enough to identify the sources of individual objects. This means that if the dataset is derived from several sources, lineage information is to be assigned as an additional attribute of the objects or as a spatial overlay to enable this identification. (Veregin 1999, p.184)

Origin

The origin of a dataset is documented by giving the organization responsible for the dataset production and how it was produced. The documentation of how the data were produced must be very detailed in order to constitute a basis for estimation of e.g. the positional accuracy of the geographic data. Thus, it should give information of the entire process – from data capture to digitally stored coordinates, including source material, data capture method, equipment, transformations etc. (Lantmäteriverket 1997, p.55)

Purpose

Purpose includes a description of the reasons and principles for creating a dataset and provides information about its intended use. (ISO 2000b, p.7; Stenborg 2000, p.41)

Source

Source information is the minimum information that should always be available for a dataset and constitute one part of the lineage of the overall dataset. The aim is to provide potential users with information about the general validity of the data. The source information is described by giving the name and the organization, which is responsible for the dataset. It is also required to give the purpose and the date of the original production. (Aalders 1997, pp.280-281)

Usage
Usage is an important quality parameter from the viewpoint of new potential users. Stating any previous use of the dataset for various applications indicates the fitness of the dataset for new intended use. Thus it is of great importance that errors, constraints and limitations for each prior use is stated. (Aalders 1997, p.281)

5.4.2 Quality elements

Completeness
Completeness in the widest sense refers to the agreement between the content in a geographic dataset and the reality. It can be derived on the basis of selection criteria, definition and other mapping rules used to create the dataset. (Veregin 1999, p.183) Completeness is defined for an entire dataset, typically in terms of errors of omission and commission. An error of omission means that a location has been omitted from its actual class. An error of commission means that a location has been assigned to an incorrect class. These errors do not appear independently of each other. If for instance a location of forest has been mapped as agricultural land, it is an error of omission in the sense that the location is not mapped as forest, and an error of commission in the sense that the location has been mapped as agricultural land. (Campbell 1996, p.385) The errors of omission and commission are preferably expressed as percentage of the total dataset.

Different types of completeness can be distinguished (Veregin1999, p.183):

- data completeness
- model completeness
- attribute completeness, and
- value completeness.

Data completeness is assessed when comparing the dataset with the dataset specification and refers to the errors of omission observed between the two quantities. Model completeness refers to the agreement between the specification and the abstract universe of which the dataset is a representation. Attribute completeness documents the degree to which all relevant attributes of an object have been captured, and value completeness refers to if values are present for all attributes.

Logical consistency
Consistency is absence of apparent contradictions in a dataset. For geographic datasets the most common measurement of consistency is topological consistency, which specifies the agreement with topological rules in the dataset. The rules vary with the extent of topology defined in the dataset. If a link-node structure is defined, some rules might be (Veregin 1999, p.182):

- only one point may exist at a given location
- lines must be intersected at nodes, and
- polygons are bounded by lines.

Consistency may also be measured for domain, format, concept, geometry and semantics. (CEN 1998a, pp.25-26; ISO 2000b, pp.6-7) They all are assessed through a comparison with how their representation is specified in the dataset specification. Domain consistency states to which extent the attribute values are within the allowed domain and format consistency states to which extent data are stored in accordance with the specified physical structure of the dataset. The other subelements are stating what their names imply.

The logical consistency and its subelements are preferably expressed as percentage of conformant values. They can also be expressed as boolean values indicating the conformance of the values or the entire file. (CEN 1998a, pp.26-27)

Positional accuracy
Positional accuracy for geographic data, also denoted spatial accuracy, describes how well a given position in a dataset agrees with the real position on the ground. The positional accuracy can be divided into absolute (external) accuracy and relative (internal) accuracy. Absolute accuracy states how accurately the position of an object is determined relative a geodetic reference network. Relative accuracy states how accurately objects are determined in relation to each other. The division of the positional accuracy of a dataset into absolute and relative accuracy is valid for explaining the difference in impact of systematic and stochastic errors. When gross errors have been eliminated, the total positional error of a geographical dataset can be considered as the sum of both systematic and stochastic errors introduced during the production process. Stochastic errors will affect both the absolute and the relative positional accuracy, while systematic errors will only affect the absolute accuracy. (João 1998, p.30)

Metrics for determining the positional accuracy of point objects are well defined. Spatial errors are often given by a mean error relative a geodetic control network, and record the bias of the measurement. Other metrics are confidence level and error ellipses. Spatial errors of point objects in two dimensions are considered to belong to the normal distribution. This enables use of statistical methods in accuracy assessment for points. The errors of more complicated geometric objects, such as lines and polygons, are more difficult to assess. The errors of lines are usually defined using an epsilon band. It is a zone of uncertainty around a line, within which there is a certain probability of observing the actual line. (Veregin 1999, p.180)

Resolution
Resolution refers to the level of detail that can be discerned in a dataset or in a map. It affects to what degree a dataset is suitable for a specific application, in the sense that the resolution must match the level of detail required in the application. All data are of limited resolution since no measuring systems are infinitely precise and because geographic datasets are inevitably generalized to some extent. Resolution is closely related to and dependent on accuracy, in the sense that decreased resolution may result in higher accuracy. (Veregin 1999, p.181) Decreasing the number of categories in a

classification increases the thematic accuracy, and increasing the grid cell size of a raster layer increases the positional accuracy.

There are at least three types of resolution (Veregin 1999, pp.181-182):

- spatial resolution,
- thematic resolution, and
- temporal resolution.

Spatial resolution is related to the minimum size of objects on the ground that can be discerned in a dataset. It is a well developed concept in remote sensing that also can be applied to geographic datasets in raster format. The minimum size refers to the ground dimensions of the pixels in the raster. For vector data the spatial resolution can be given by rules of minimum mapping unit size, which depend on the map scale. Thematic resolution refers to the fineness with which attributes are registered. It is described with e.g. the number of decimals of a value or the number of categories in a classification. Temporal resolution states the time collection interval for each measurement in years, months, days or seconds etc. It is the time required for capturing the data, e.g. to measure the width of a road or register a picture in a motion picture. The temporal resolution for a motion picture is about a thousandth of a second.

Temporal accuracy

The temporal accuracy, or currency, of data states how up-to-date a dataset is. It is usually given by a revision date indicating the last time the data were controlled and found to be correct. Besides the revision date, other dates could be of interest when describing the history and status of an object, such as the date of construction of a road or the date of registration concerning real estates. However, temporal accuracy can also be described by accuracy in time measurements for attributes, rate of change, temporal validity and temporal lapse. Accuracy in time measurement can be expressed as the rmse (root mean square error) of time value of type time unit. A rate of change states how much of the information is target for change during a certain time period, this can be expressed as number of units of change per unit of time. Temporal validity states whether a data record is out of date, valid or not yet valid. Temporal lapse states the time between a change in the real world and corresponding change in the dataset. (CEN 1998a, p.24; ISO 2000b, pp.6-7)

An idea of the expected temporal accuracy of a dataset is given in the plans for maintenance in the specification. However, since users and producers may have different views of the temporal accuracy concept, the maintenance plans may not always provide the wanted answer. A user wants to know if the data are still correct for an intended use at a certain time, but the producer is more likely to understand the currency as an update policy.

Thematic accuracy

Thematic accuracy, also denoted semantic accuracy, refers to the accuracy of the thematic component of the geographic dataset. It is indicated in different ways depending on the measurement scale of the attribute. For quantitative attributes the metrics are quite similar to those used to measure the spatial quality of point features, e.g. mean value. For categorical data on the other hand, quality is assessed using methods developed in the field of classification accuracy in remote sensing.

The classification accuracy assessment is based on a comparison of categories between the classified map and a reference source in selected point locations. The result of the comparison is presented in a so-called confusion matrix, which identifies not only the overall errors for each category but also misclassifications of categories, due to confusion between categories. This confusion results in errors of omission and commission. For further reading see e.g. Campbell (1996).

A basic classification accuracy term is the proportion correctly classified, which provides an estimation of the overall accuracy of the classified map. However, to be able to assess more details of the quality of the map it is necessary to examine the full error matrix. The error matrix is used for deriving different quality measures, such as errors of omission and commission, producer's accuracy and consumer's accuracy. Sometimes a more objective measure is wanted, such as the kappa coefficient. The kappa coefficient expresses to what degree the categories in the point locations in the classified map differ from an expected category assigned at random. Considering an area with 40 % forest and 60 % agricultural land. If a vast number of point locations were divided over the area, about 40 % of them would be located in forest and 60 % in agricultural land. This is a random assignment. Kappa can be assigned values between –1 and 1, where a negative value indicates that a random assignment would be better than the classified map and a positive value indicates that the classification is better than a random assignment would be. (Veregin 1999, p.181; Campbell 1996, chap.13)

5.5 Metaquality elements

Metaquality states the quality of quality. Metaquality elements are used for documenting the reliability of the values of the quality elements and what these values represent. This is important to know in order to take advantage of the given quality. For instance, if the positional accuracy is set to be 0,2 meters, it is an uncertain measure if there is no additional information stating if it is an absolute tolerance value (all points are within that tolerance) or a standard deviation (about one third is not within the tolerance). It is also of importance to know how the quality measure was derived, i.e. whether it originates from a random sample of the dataset or from the entire dataset.

Possible metaquality elements to use are (Aalders 1997, p.280):

- reliability
- confidence
- methodology, and
- abstraction modifier.

Reliability reflects to which degree the quality information represents the whole dataset. The measure of the confidence of the quality information indicates the level of confidence of the dataset, given by a standard error or a confidence interval for a given confidence level (CEN 1998a, p.10). Methodology contains a description of how the quality information was derived, indicating how the result was obtained. The abstraction modifier reports about the abstractive quality of the dataset. It can be a textual record describing the degree of distortion of quality resulting from the process of abstracting the real world to the nominal ground. The degree of abstraction given by the abstraction modifier will be homogenous across all objects of a particular type. (Duckham, chap1.2 and 3.2)

6 Standards

This chapter aims at presenting standards concerned with quality of geographic information. The chapter starts by explaining what standards mean and how they work, in order to provide a foundation for the rest of the chapter. Then the three major standardization organizations (for Swedish applications,) their aims and quality models are presented. Additionally, the metadata standard of the U.S. Federal Geographic Data Committee (FGDC) is presented since the metadata editor studied in Chapter 9 is primarily based on this standard. Also the Open GIS "standard" is presented since this is well known within the field of geographic information. The standards are intentionally presented in a summarized way, not necessarily with the same structure, in order to show how they are designed. In the last part of the chapter different quality models are compared in order to show how quality of geographic information can be presented in different ways.

6.1 Introduction to standardization

There are several definitions of the word standard. The International Organization for Standardization (ISO) describes standards as "[...] documented agreements containing technical specifications or other precise criteria to be used consistently as rules, guidelines, or definitions of characteristics, to ensure that material, products, processes and services are fit for their purpose. Standards thus contribute to making life simpler, and to increasing the reliability and effectiveness of the goods and services we use." (Introduction to ISO 2000) The Swedish standardization organization SIS says that "Standard means a recommendation of doing something, for example design a product or use a test method in a certain way. The use of standards will lead to simplification and will also give security and cost savings in production and trade." (SIS 2000, author's translation) In other words standards can be seen as tools facilitating collaboration between different actors within the same area, by making different products capable of working together. Standards thus facilitate for both producers and users.

The need for standardization becomes obvious when people and organizations must communicate. ISO's homepage lists some reasons why this has become so important (Introduction to ISO 2000):

- Worldwide progress in trade liberalization. The markets are expanding, and in order to achieve a fair competition there must be common references between different regions or countries.
- Interpenetration of sectors. The industries are no longer independent, meaning that they do not only use products they have produced themselves. Instead there are certain industries supplying all other industries with certain products, e.g. GIS software. It is important that these products can function in all environments.
- Worldwide communications systems. This is a sector that is increasing rapidly and needs to be standardized at a global level in order to function worldwide.

- Emerging technologies. New technology needs global standards to define terminology and accumulating databases of quantitative information.
- Developing countries. A standardized infrastructure is needed for the success of economic policies aimed at achieving sustainable development.

By using standards companies and organizations may become more efficient and thereby increase their productivity. Using standards may also help drawing attention to the company, and may yield new customers since standards build confidence. In the end, using standards will help making money. When participating in the development of standards, national as well as international, it is possible for companies and organizations to have an influence on the standards while they are being developed. This will give the companies and organizations strengthened positions, market benefits and advance information.

A standard can never be compelling to a user but will have impact if many parties join it. The government can also issue constitutions stating that some standards must be complied with. In a growing society new areas to standardize will appear. Also, old standards need to be audited.

In the case of geographic information, data have earlier been used only within a specific organization why there has been no need for standardization. With the rapid development of digital communication and the Internet, the need for exchange of information has increased to a great extent. Standardization within the area is now also increasing.

6.2 Organizations and cooperation

Formal standardization is run by global, European and national standardization organizations. The aim is to bring about an international framework that will support efficient access and usage of data or products, and will enable development and usage of standards for specific applications.

Standardization on global level is handled by the International Organization for Standardization (ISO) and on European level by le Comité Européen de Normalisation (CEN). The Swedish standardization organization SIS (Standardiseringen i Sverige) is the national counterpart. Standardization of geographic information in Sweden is handled by the standardization body STG (Allmänna standardiseringsgruppen), which is authorized by SIS. The work on European and global level must always originate from the national standardization organizations. A difference between CEN and ISO standards in relation to Swedish standard is that Sweden has committed to have national standards that agree with CEN standard, but not necessarily with ISO standard. (STG 1998d, p.3; Peach 1992, p.2)

Besides formal standardization, which is performed by these organizations, interested parties can carry on different types of harmonizing where the result sometimes becomes a de facto standard. The term de facto standard means that it is not a formal standard, but so commonly used that it is almost regarded as a standard.

6.3 ISO

The International Organization for Standardization (ISO) is a non-governmental organization established in 1947. Its mission is to promote the development of standardization and related activities in the world in order to facilitate the international exchange of goods and services, and to develop cooperation in the areas of intellectual, scientific, technological and economic activity. The scope of ISO is not limited to any particular business; it covers all technical fields except electrical and electronic engineering, which is the responsibility of IEC (International Electrotechnical Commission). (Introduction to ISO 2000) The work results in international agreements that are published as International Standards. Until January 1999, ISO has published some 12 000 International Standards. (Introduction to ISO 2000) The standards developed within ISO are voluntary, there are no legal requirements forcing countries to adopt the standards. The countries themselves, however, can chose to attach legal requirements to the standards and thereby make them mandatory. (Peach 1992, p.2)

ISO consists of national standardization bodies from some 130 countries, one from each country. The Swedish representative is SIS. (SIS 2000) The technical work is highly decentralized, carried out in a hierarchy of technical committees (TC), subcommittees (SC) and working groups (WG). The preparation of International Standards is normally carried out through the technical committees. Every member body interested in a specific subject has the right to be represented on that committee. International organizations, governmental and non-governmental, also take part in the work.

Standardization of geographic information is, according to ISO, best served by a set of standards that integrates a detailed description of geographic information concepts with the concepts of information technology. (ISO 2000a, p.9) The development of standards for geographic information therefore must consider adoption or adaptation of information technology standards whenever possible. Only when this can not be done pure geographic information standards need to be developed. (ISO 2000a, p(vi))

In the following of this section standards needed for this project are presented after a brief presentation of the technical committee ISO/TC 211, which develops international standards concerned with geographic information.

6.3.1 ISO/TC 211 Geographic information/Geomatics

ISO/TC 211 *Geographic information/Geomatics* is the technical committee within ISO that deals with standardization in the field of digital geographic information. The aim is to establish a structured set of standards for geographic information. The standards may specify methods, tools and services for data management including definition and description, acquisition, processing, analyzing, accessing, presenting and transfer of such data in digital form between different users, systems and locations. The standards shall provide a framework for the development of sector-specific applications using geographic data. (ISO/TC 211 2000)

ISO/TC 211 has its chairman and secretariat in Norway. It consists of 33 member countries, including Sweden, and 18 observing member countries. The committee is divided into five working groups that each handles specific parts of the standards. (ISO/TC 211 2000) The working groups are listed in Table 6.1.

Working Groups (WG) of ISO/TC 211
WG 1 – Framework and reference model
WG 2 – Geospatial data models and operators
WG 3 – Geospatial data administration
WG 4 – Geospatial services
WG 5 – Profiles and functional standards

Table 6.1. The working groups of ISO/TC 211.

ISO/TC 211 is presently developing a family of standards called ISO 19100 *Geographic information*, which is described in Section 6.3.4. It is the continuation of the old project series 15046. In October 1999 it was decided that all documents within ISO/TC 211 will change number from 15046 to "19100". All numbers between 19100 and 19199 are reserved for standards within ISO/TC 211. (Essén 1999)

6.3.2 ISO 9000 Quality management

ISO 9000 is a family of international standards primarily concerned with quality management. Quality management means what an organization does, throughout the process of producing and delivering, to ensure that its products or services conform to the customers' requirements. The standards are used as an aid when developing quality systems to govern the business of an organization in a standardized and secure way. They specify the requirements of the quality systems, and provide guidelines for how to implement such a system and how to fulfill the requirements. For implementation, evaluation and audition of a quality system documentation is essential. Documenting routines helps the organization to understand the performance of every process and their contribution to the final quality of products and services. (CEN 1994, p.8) The special characteristic of ISO 9000 is that the standards are generic and thus useful to any companies within all types of businesses. Although the requirements and principles are standardized, the applications of these standards will always be unique due to the uniqueness in an organization's business.

The family of ISO 9000 standards is developed within ISO's technical committee ISO/TC 176 – *Quality management and quality assurance*. The first edition of the standards was published in 1987 as the ISO 9000 Standard Series on quality management and assurance. The series then comprised 5 standards; ISO 9000, ISO 9001, ISO 9002, ISO 9003 and ISO 9004. (STG 1994, p.7) The current ISO 9000 family of standards (before the year 2000 revision) contains more than 27 standards and documents. (ISO/TC176/SC2 2000) It comprises the standards belonging to the 9000 series, the 10000 series, which includes e.g. guidelines for auditing quality systems, and the ISO 8402 standard *Quality management and quality assurance - Vocabulary*, which provides the vocabulary to use. These standards can be divided into requirements and guidelines for interpreting and introducing a system for quality management. (STG 1997, p.11) Table 6.2 gives some examples of these requirements and guidelines. (ISO 1998, pp.2-4)

Guidelines	
ISO 8402	Quality management and quality assurance – Vocabulary
ISO 9000-1	Quality management and quality assurance standards – Part 1: Guidelines for selection and use
ISO 9000-2	Quality management and quality assurance standards – Part 2: Generic guidelines for the application of ISO 9001, 9002 and 9003
ISO 9000-3	Quality management and quality assurance standards – Part 3: Guidelines for the application of ISO 9001:1994 to the development, supply, installation and maintenance of computer software.
ISO 9004-1	Quality management and quality system elements – Part 1: Guidelines
ISO 9004-2	Quality management and quality system elements – Part 2: Guidelines for services
ISO 10005	Quality management – Guidelines for quality plans
ISO 10007	Quality management – Guidelines for configuration management
Requirements	
ISO 9001	Quality systems – Model for quality assurance in design, development, production, installation and servicing
ISO 9002	Quality systems – Model for quality assurance in production, installation and servicing

Table 6.2. Some of the standards in the ISO 9000:1994 family.

When an organization has a quality system according to ISO 9000 the risk of misunderstandings and obscurities between the customer and the supplier is minimized. A competent authority can assess a quality system's conformance to the requirements of one or several standards within the ISO 9000 series. If the conformance is satisfactory a certificate may be issued as proof. Thereafter controls are performed twice annually to guarantee that the requirements are met (Lantmäteriverket 1997, p.86). Several thousands of ISO 9000 certificates have been issued to organizations and companies around the world. Perhaps this certification is the most worldwide known approach of the ISO 9000 standards.

A revised edition of the ISO 9000 family of standards was published on December 15 2000. One objective for the Year 2000 revisions was to simplify the structure and reduce the number of standards within the family. The three certification standards, ISO 9001, ISO 9002 and ISO 9003 as well as the quality management guidelines in ISO 9004-1 are being replaced. ISO 9001:2000 provides the requirements and ISO 9004:2000 the guidelines for performance improvements. (ISO/TC176/SC2 2000) The ISO 9001:2000 standard is intended to be "generic and are intended to be applicable to all organizations, regardless of type, size and product category". (ISO/TC 176/SC 2 2000, p.2)

The quality management standards of the revised ISO 9000:2000 series are based on eight universal quality management principles. These principles can be used as a framework to guide organizations towards improved performance. The principles are defined in the new standards ISO 9000:2000 and ISO 9004:2000 and are (ISO/TC 176/SC 2 1997):

- customer focus
- leadership
- involvement of people
- process approach
- system approach to management
- continual improvement
- factual approach to decision making, and
- mutual beneficial supplier relationships.

6.3.3 ISO 10303 Industrial automation systems and integration – Product data representation and exchange

ISO 10303 is a family of standards used for different sorts of product descriptive data. The main idea of the standards is to enable transfer of information between different systems, such as divisions of a company that share the same information but handle it in different ways. (Celander 1999, p.7) Parts of this standard are e.g. used in the Swedish Technical Framework, which is described in Section 6.5.1.

The complete name of the family of standards is *ISO 10303 – Industrial automation systems and integration – Product data representation and exchange*, but it is often called STEP (Standard for Exchange of Product Model Data) after its most widely used component. The development of the parts within the ISO 10303 standards is performed by subcommittee 4 (SC4 Industrial data) in the technical committee ISO/TC 184, which handles Industrial Automation Systems and Integration. (ISO/TC 184 2000) The work was initiated in the mid-1980s and the first part of the standard was established in 1993. (Celander 1999, p.15) The original aim was to exchange CAD information between systems, but now all kinds of product information are handled and the aim is to support this information handling throughout a product's life cycle. The main issues are (Wikström 1998, p.7):

- system independent data description
- system independent data transfer and database access
- system independent long-term storage
- defining interface between systems, programs and databases with different origin and internal structure, and
- facilitating the information's portability between systems, e.g. at system upgrading or system change.

ISO 10303 is an attempt to create a global, uniform, integrated, powerful and reliable standard for information handling. A basic idea is to unambiguously define the concepts used in the standard. (Celander 1999, p.15) These sets of concepts are called information models. An information model thus describes what concepts exist, their meaning and how they are connected to each other.

The ISO 10303 standards are in their original design not directly useful for geographic information. In the geometry model, for instance, the connection to a geodetic reference system is missing and coordinates are supposed to be expressed in an orthogonal Cartesian system. However, the family contains a number of generally useful standards and as Wikström (1998, p.8) states: "The STEP family is suitable as a concept for geographic information, at least to the extent that the European prestandards and the Swedish Technical Framework have accepted" (author's translation). This follows from the fact that CEN as well as the Swedish Technical Framework have chosen to base their geographic information standards on several of the ISO 10303 standards. Also, there has been some collaboration with ISO/TC 211 in order to harmonize geometric and topologic descriptions. (Wikström 1998, p.21)

6.3.4 ISO 19100 Geographic information

ISO 19100 *Geographic information* is a family of geographic information standards developed by ISO/TC 211. The standards establish "[...] a structured set of standards for information concerning objects or phenomena that are directly or indirectly associated with a location relative to the earth." (ISO 2000a, p.7) The standard propositions are being prepared by working groups and each standard corresponds to a work item. When the series is complete it will contain thirty, or so, standards. (Stanli 2000a, p.9) The main parts of the standards are intended to be finished in 2001, and will then replace the old 15046 series. (Essén 2000) The 19100 series covers everything from fundamental data descriptions to transfer and presentation of geographic data. One, for this project, important part of the 19100 series is the quality model of ISO 19113 *Geographic information – Quality principles* (see section 6.3.5).

The 19100 standards are used for defining, describing, and managing geographic information (ISO 2000a, p.6 with focus on the basic semantics and structure of geographic information for data management and data interchange purposes. (ISO 2000a, p.8) The main objectives of the standards are (ISO 2000a, p(vi)):

- Increasing the understanding and usage of geographic information.
- Increasing the availability, access, integration, and sharing of geographic information.
- Promoting the efficient, effective, and economic use of digital geographic information and associated hardware and software systems.
- Contributing to a unified approach to addressing global ecological and humanitarian problems.

In order to achieve these goals, standardization of geographic information in the ISO 19100 series is, as previously described, based on the integration of the concepts of geographic information with those of information technology. (ISO 2000a, p(vi)) The ISO 19100 series uses the modeling language UML and the transfer format XML. (Cederholm 2000)

6.3.5 ISO 19113 Geographic information – Quality principles

ISO 19113 *Geographic information – Quality principles* is an International Standard that provides principles for describing the quality of geographic data and components for reporting quality information. It also provides an approach to organizing information about data quality. (ISO 2000b, pp.(v)-1) The standard can be used when (ISO 2000b, p.4):

- identifying and reporting quality information
- evaluating the quality of a dataset
- developing product specifications and user requirements, and
- specifying application schemas.

The quality description of ISO 19113 can be applied to a dataset series, a dataset or smaller groupings of data located physically within the dataset sharing common characteristics, to evaluate the quality. The quality of the dataset should then be described using data quality elements and data quality overview elements. Data quality elements, together with data quality subelements and the descriptors of a data quality subelement provide quantitative quality information. Data quality overview elements, on the other hand, describe the non-quantitative quality of a dataset. (ISO 2000b, pp.4-8)

In order to make the standard useful for different applications the number of quality parameters are kept minimized. The presented parameters are therefore those predicted to be the most used ones. (Stenborg 2000, p.36) Additional data quality elements, data quality subelements and data quality overview elements may be created if it is suitable for the purpose of the quality description of a specific application. (ISO 2000b, pp.4-8)

According to ISO 19113 the following (presented in Table 6.3) data quality elements and data quality subelements shall be used, where applicable, to describe aspects of the quantitative quality of a dataset. (ISO 2000b, pp.6-7)

Data quality elements	Data quality subelements
Completeness	Commission
	Omission
Logical consistency	Conceptual consistency
	Domain consistency
	Format consistency
	Topological consistency
Positional accuracy	Absolute or external accuracy
	Relative or internal accuracy
	Gridded data position accuracy
Temporal accuracy	Accuracy of a time measurement
	Temporal consistency
	Temporal validity
Thematic accuracy	Classification correctness
	Non-quantitative attribute correctness
	Quantitative attribute accuracy

Table 6.3. Data quality elements and data quality subelements in ISO 19113.

Quality information shall be recorded for every applicable data quality subelement. This information shall be described using seven descriptors, (ISO 2000b, p.7) which are presented in Table 6.4.

Descriptors of data quality subelements
Data quality scope
Data quality measure
Data quality evaluation procedure
Data quality result
Data quality value type
Data quality value unit
Data quality date.

Table 6.4. Descriptors of data quality subelements in ISO 19113.

The non-quantitative quality of a dataset shall where applicable be described using the, data quality overview elements presented in Table 6.5. (ISO 2000b, pp.7-8)

Data quality overview elements
Purpose
Usage
Lineage.

Table 6.5. Data quality overview elements in ISO 19113.

Metaquality, which means quality of the quality information, may include a measure of the confidence or the reliability of the quality information. This type of information is recorded in ISO 19114's *Quality Evaluation Report.*

6.4 CEN

CEN (Comité Européen de Normalisation) was founded in 1961 by the majority of its current members, which mainly are the western European countries. It was first based in Paris but moved to Brussels in 1975. At the same time it acquired formal statutes and was registered as a non profit-making, international, scientific and technical institution, meaning it is an independent organization. The mission of CEN is to "promote voluntary technical harmonization in Europe in conjunction with worldwide bodies and its partners in Europe." (CEN 2000a)

CEN, like ISO, consists of technical committees (TC) which gather the national delegations of experts convened by CEN National Members. The technical committees look at any relevant work falling within its scope and also at data supplied by any relevant organization. The results of this work can then be offered to ISO. (CEN 2000b)

CEN standards can be either European prestandards (ENV) or European standards (EN). A prestandard is a temporary standard waiting for an international (ISO) standard to be established. In many cases CEN has chosen just to develop prestandards, due to the desire of developing ENs identical to International Standards. If the ISO standard in some parts differs from what has been produced as an ENV there are possibilities to make adjustments to the CEN standard. An ENV is initially valid for three years. After two years the members of CEN are requested to submit their comments, including whether the ENV can be converted into a European Standard (EN). (CEN 1998a, p.1) The validity of the ENV may also be prolonged by two years. (Blomberg) Different countries have chosen to take advantage of the prestandards in different ways. Some countries will wait with national introduction until the standard becomes an EN, others perform tests or use the prestandards in implementations. Some countries have gone so far as to base their national standard on European prestandards. (Wikström 1998, pp.5-6) ENVs are mainly used in rapid development industries or when a standards proposal has been brought down in the final vote. In Sweden it is optional whether to use the ENVs or not, it is however desired that ENVs are tested parallel to the national standards. (Blomberg)

The ENVs must be announced, in the same way as for an EN, in all member countries and made available promptly at national level in an appropriate form. The member countries may choose to keep conflicting (to the ENV) national standards until the final decision about the possible conversion of the ENV into an EN is reached. Once an EN is established it will become mandatory in all member countries. (STG 1998d, p.3)

CEN's main objective of standardization in the field of geographic information is to "enable geographic information to be accessed by different users, applications and systems, and from different locations." (CEN 1998b) This requires:

- a standard way of defining and describing this information
- a standard method for structuring and encoding it, and
- a standard way of accessing, transferring and updating it via geographic information processing and communication functions, independent of any particular computer system.

In the following of this section standards needed for this project are presented after a brief presentation of the technical committees CEN/TC 278 and CEN/TC 287, which develop European standards concerned with geographic information.

6.4.1 CEN/TC 278 Road Transport and Traffic Telematics

TC 278 *Road Transport and Traffic Telematics* is the technical committee within CEN that deals with standardization within this area. It shall support e.g. (Ertico 2000):

- communication between vehicles and road infrastructure
- communication between vehicles
- traffic and parking management
- user fee collection
- public transport management, and
- user information.

CEN/TC 278 has developed over 50 work items classified either as application specific, databases, interfaces or basic concepts. One of the items is the geographic data file standard ENV ISO 14825:1996 (see section 6.4.4). The committee cooperates with the ISO counterpart, ISO/TC 204 Transport Information and Control Systems. The two committees cover the same themes and have some working groups joint or linked. (Ertico 2000)

6.4.2 CEN/TC 287 Geographic Information

TC 287 *Geographic Information* is the technical committee within CEN that deals with standardization within this area. The work program is divided into four main parts (CEN 1998b):

- Fundamentals, including reference model, overview and definitions.
- Data description, including rules for application schemas, geometry, quality, metadata and data transfer.
- Referencing, including position, time and indirect positioning systems.
- Processing, including query and update.

The standardization of geographic information within CEN TC 287 started in 1992 while the more extensive package of ISO standards was started in 1994. Europe, in CEN, thus had a two years' start on the rest of the world in deciding the architecture of standards for the market. (Geographic information standards 2000) During 1998 and 1999 some of the prestandards were published, e.g. CEN ENV 12656 *Geographic information – Data description – Quality* (see section 6.4.3) and CEN ENV 12657 *Geographic information – Data description – Metadata.*

The CEN/TC 287 work is an example of user operated standardization. A few system suppliers have participated but mainly as observers. Producer representatives and users developed the standards. (Wikström 1998, p.22)

6.4.3 ENV 12656 Geographic information – Data description - Quality

ENV 12656 *Geographic information – Data description - Quality* is a European prestandard, produced by CEN/TC 287, establishing general principles for describing the quality of geographic information. It also provides a schema for describing data quality and presents information appropriate for assessing the quality of geographic information. The standard was approved by CEN in October 1998 (CEN 1998a, p.1) and is applicable for (CEN 1998a, p.5):

- suppliers of datasets who wish to describe the overall quality of a product
- specific quality information regarding a dataset supplied to a user, and
- users, to describe their requirements.

ENV 12656 provides information required to assess the quality of geographic information against its specification. It provides a mechanism for maintaining quality information to be used in quality management systems. It does not, however, provide a mechanism for managing quality itself. It does not, either, specify how measures of quality shall be made or the quality expected of an individual product. Neither does it provide techniques for quality control or assessment, nor concern itself with quality systems or management. The quality schema is designed mainly to be used for digital geographic datasets but the principles are also applicable to describe geographic information in other forms, such as paper maps or lists. The quality description is said to be "suitable for describing the quality of geographic data at any level, from the complete dataset down to individual items of data." (CEN 1998a, p.5)

Four quality elements[2] are used for describing the quality of geographic information. For each geographic dataset, at least one of them shall be given. (CEN 1998a, pp.9-27) The quality elements are further described in Table 6.6.

Quality elements	
Lineage	Production
	Source and processes
Usage	The name of an organization that has used the geographic dataset
	The type of usage by that organization
	Any constraints or limitations that were imposed or found during use
	The date when the geographic dataset was used
Quality parameters	Described in Table 6.7
Homogeneity	Statistics on the quality parameters

Table 6.6. Quality elements in ENV 12656.

When using quality parameters[3] to describe a dataset at least one of the parameters described in Table 6.7, or a user-defined parameter, should be used. The quality parameters will be described using one or more quality indicators[4]. The indicators may also be user-defined. (CEN 1998a, pp.9-27)

[2] CEN's quality elements correspond to the quality overview elements as described in chapter 5.
[3] CEN's quality parameters correspond to the quality elements as described in chapter 5.
[4] CEN's quality indicators correspond to quality subelements as described in chapter 5.

Quality parameters	**Quality indicators**
Positional accuracy (using at least one of the following:)	Relative horizontal accuracy
	Relative vertical accuracy
	User-defined quality indicator
Semantic accuracy (using at least one of the following:)	The accuracy of classification
	The agreement for an attribute
	User-defined quality indicator
Temporal accuracy (using, wherever possible, for example:)	Accuracy in time measurement
	Last-update
	Rate of change
	Temporal lapse
	Temporal validity
Completeness (using at least one of the following:)	Omission
	Commission
	User-defined quality indicator
Logical consistency(using, wherever possible, for example:)	Geometric consistency
	Topological consistency
	Semantic consistency
	Format consistency
	Validity of values

Table 6.7. Quality parameters and quality indicators in ENV 12656.

Each quality indicator shall be defined by three descriptors. (CEN 1998a, p.17) These are presented in Table 6.8.

Descriptor of quality indicator	**Constituents**
Name of quality indicator	
Description of the aspect of quality the quality indicator describes	
Definition of quality measures[5] defining what the quality indicator measures	Name of quality measure
	Description of test to be performed to obtain a result for the quality measure
	Type of values representing the result of the quality measures

Table 6.8. Descriptors of quality indicators in ENV 12656.

[5] A quality measure, according to CEN ENV 12656, defines a specific test to be performed to obtain a result on a geographic dataset.

Metaquality may be presented at each specific instance. It should be recorded in accordance with ENV 12657 *Geographic Information – Data Description - Metadata* and shall comprise a measure of the confidence of the quality information. Additionally, it may also comprise (CEN 1998a, pp.8-10):

- a measure of the reliability of the quality information
- a description of the methodology used to derive the quality information, including if it has been assigned a theoretical estimation or a quality parameter measure, and the specification of the procedure followed in both cases, and
- an abstraction modifier (see Section 5.5).

6.4.4 ENV ISO 14825:1996 Geographic Data Files

ENV ISO 14825:1996 *Geographic Data Files* (GDF) is a European Standard used to describe and transfer road related geographic data. The standard was developed by CEN/TC 278 WG 7 Geographic Data Files with the initial intention to "specify the data content, means of data representation and structure of data supply for vehicle navigation systems" (CEN 1995, p.20). The standard states rules for data capture and how geographic features, attributes and relations have been defined. One of the components of the standard is the GDF exchange format.

The GDF standard contains catalogs describing and classifying significant features, their attributes and relations, and also how to model meta information such as geodetic references and data sources. It contains a schema for representing a feature, and a specification that describes rules and methods on measuring the quality of a dataset. Furthermore, it contains data structures describing how the modeled information can be represented by a set of data types, and finally also information needed to supply the GDF data as physical data records. (CEN 1995, p.22)

Quality is in the GDF standard described in the section *Quality Description Specifications*. The specification includes generic definitions of terms and methods related to the description and measurement of quality. Based on the specification it is possible to define individually determined quality levels. (CEN 1995, p.215) The quality concept is approached by a number of defined checks that must be performed in order to acquire quality measures of the data. The quality evaluation of a dataset consists of two different parts (CEN 1995, p.217):

- Format quality checking, which describes to what extent the information is stored correctly. Several error groups are considered, e.g. syntax errors, value errors and topological errors.
- Semantical quality checking, which describes to what extent the information corresponds to reality.

The semantical quality checking can be performed indirectly or directly. The indirect semantic quality checking defines the quality of a dataset according to if the aspects of its production meet criteria as defined in ISO 9000. The direct semantic quality checking, on the other hand, compares data with the real world situation or defined reference sources. This is done in accordance with another International Standard, ISO 2859 *Sampling Procedures and Tables for Inspection by Attributes*. (CEN 1995, pp.218-220) Several direct checks are performed, each yielding one or two quality aspects of the following (CEN 1995, p.221):

- positional accuracy
- completeness
- logical consistency
- attribute accuracy
- attribute completeness
- relationship completeness, and
- relationship correctness.

6.5 Stanli

The Stanli project (Standardisering Landskapsinformation) develops standards for geographic information and data, within Sweden. The project was initiated in 1988 by ULI (Utvecklingsrådet för Landskapsinformation) and has participants from about 35 government agencies, companies and special interest organizations. Stanli became a part of Swedish standardization in 1990 when ULI assigned STG (Standardiseringsgruppen) to run the project. (Stanli 2000b; Eklundh 1999, p.326) Stanli has its own project secretariat, and a project committee has the overall responsibility. A technical committee (SIS-STG/TK 80 *Framework for geographic information*) within the STG is responsible for appointing Swedish representatives to international working groups and delegations, making decisions of technical matter and preparing proposals of framework for national standardization. Various applications have different technical committees. The project also includes a working group handling terminology issues, and a forum responsible for providing guidance, education, literature and aid in applying the standards. (Eklundh 1999, p.326; Lantmäteriverket 1997, p.67)

Stanli's main objective is to produce useful standards that will be applied by leading producers and users of geographic data and systems, because it is of benefit to their business. (Stanli 1997, p.4) Stanli also aims at helping new lines of business discover the possibility to use geographic data as a support to their business. The vision is to make the use of geographic data simple. This means that (Stanli 2000b):

- a joint regulatory framework supports production and usage
- searches, orders and delivery work
- data have the meaning, structure and quality agreed upon
- data can be integrated into other data and systems
- data function is independent of technical systems
- the usage of data is secured over time, and
- producers and users apply the regulatory framework agreed upon.

Stanli is taking active part in the development of international standards, both within ISO and CEN, in order to develop a framework standard. Stanli is also an active part in building national standards based on the international framework of general standards established within the area.

In the end of year 2000 six Swedish standardization bodies together with SIS (Standardiseringen i Sverige) were fused together as a new organization; Swedish

Standards Institute SIS. Both STG and Stanli are part of this new organization. (Stanli 2000a, p.7)

6.5.1 The Swedish Technical Framework

Stanli decided, in 1997, to develop an interim technical framework for geographic information. One reason for doing this was that several standards were being developed in Sweden, and since an international framework tended to be delayed, they needed a national base in order to proceed. Another reason was that experiences from a Swedish framework would contribute to the development of the international framework (by ISO). (Eklundh 1999, p.329) The Swedish Technical Framework (STR, Svenska Tekniska Ramverket) is described in the manual STG Hb 171.

The framework is based on European prestandards, the ISO 10303 "STEP" standards and on some methods and specifications developed in Sweden. The principles follow the ones developed within ISO 10303 and CEN/TC 287, and also ISO/TC 211 proposals. (Wikström 1998, p.8; STG 1998d, p.3) Since an established EN must become a Swedish standard the corresponding part of the Swedish framework will then no longer be valid. The remaining parts of the framework must then be modified to secure the adaptation to Swedish standard. When the new ISO 19100 series is established the Swedish Technical Framework will be adjusted to it. (Cederholm 2000)

The framework mainly includes general principles, but also special parts concerning geographic and geometrical aspects. In order to simplify the process of integrating geographic data with other business related data, the framework creates a fundamental platform for developing standards by specifying methods for data description and data transfer. This will decrease the number of interfaces each actor needs to create and maintain in order to import and export data. The framework may also be used by producers and users of geographic data even when standards for a specific area are not yet developed.

The framework can, in short, be described as documents of how geographic data should be constructed. It is governing when developing a standard (STG 1998c, p.2) and includes models, rules and directions for doing this. Description methods for the contents of geographic datasets e.g. in the form of database catalogs or metadata are also included. The basic components for data description and data transfer of geographic information in the framework handbook are (STG 1998a):

- summary and description of basic concepts
- rules for development of application models
- routines for quality assurance (of a standard)
- ISO 10303-11 EXPRESS
- ISO 10303-21 STEP
- spatial model
- positional model
- quality model (see section 6.5.2), and
- metadata model.

6.5.2 The Swedish Technical Framework, quality model

The quality model of the Swedish Technical Framework is the national counterpart to ISO's 19113 and CEN's ENV 12656. It is based on CEN's quality model but since it is adapted to Swedish needs (Stenborg 2000, p.23) it differs from CEN in the following aspects (STG 1998d, pp.15-16):

- the quality information that can be presented has been limited
- for the quality information that can be presented, the quality measures have been standardized, and
- by using predefined measures the amount of data required for the quality information can be reduced.

The quality information in the STR is not supposed to meet all possible needs of all producers. It will only cover the most frequently used quality parameters. The STR, however, also includes guidelines for creating and present additional quality parameters in a standardized way, if the need occurs. (Stenborg 2000, p.23)

The quality model contains data quality overview elements and quality concepts[6] as presented in Tables 6.9 and 6.10. The quality concepts are described using quality parameters. (STG 1998b, pp.5-9) It is also possible for the user to define additional quality concepts and quality parameters. (Stenborg 2000, p.44)

Data quality overview elements	
Administrating organization	
Lineage	Original production
	Process history

Table 6.9. Data quality overview elements in the Swedish Technical Framework.

Quality concepts	**Quality parameters**
Positional accuracy	External positional accuracy in plane
	External positional accuracy in height
	Internal positional accuracy in plane
	Internal positional accuracy in height
Completeness	Omission
	Commission
Thematic accuracy	Classification
	Quantitative attributes
	Qualitative attributes
Logical consistency	Topology
	Domain

Table 6.10. Quality concepts and quality parameters in the Swedish Technical Framework.

[6] Quality concepts, according to the quality model of the Swedish Technical Framework, correspond to quality elements as described in chapter 5.

For each of the quality parameters one or more quality measures can be defined. Some of the quality parameters can be applied to both objects and attributes while some of them can only be applied to either one. (Stenborg 2000, p.47) Five components, as presented in Table 6.11, will be used to describe user-defined quality parameters and their measures. They may also be used to describe the standardized quality parameters. (STG 1998b, p.23)

Components
Term, quality concept
Term, quality parameter
Measure
Unit
Value

Table 6.11. Components of quality parameters in the Swedish Technical Framework.

Metaquality should be given for every quality parameter value in order to define the reliability of the quality information. The metaquality should contain information about (STG 1998b, p.14):

- confidence
- representativity, and
- test method.

6.5.3 SS 63 70 04 Geografisk information – Väg- och järnvägsnät – Begrepps- och tillämpningsmodell

Stanli has also developed the Swedish standard SS 63 70 04 Geografisk information – Väg- och järnvägsnät – Begrepps- och tillämpningsmodell (Geographic information – Road and railroad network – Concept and application model). The standard defines fundamentals for describing a road or railroad network, as well as occurrences linked to the network. (SIS 1999, p.6)

The standard consists of a conceptual model and an application model. The conceptual model defines the used concepts and their internal relations. The application model consists of a formal description of the information used in a certain application, and is intended for application developers. (SIS 1999, p.7) The application model is described using the modeling language EXPRESS (see section 4.2.2) and uses the transfer format STEP (see section 4.2.3). (SIS 1999, p.16) The standard also contains recommendations for representing geometry and topology of the network.

The SS 63 70 04 standard has been of limited use since it, among other things, does not cover all required functionality. Therefore, a decision has been made to make further developments of the standard. Since Stanli is also investigating the consequences of adapting the new standards of the ISO 19100 series (see section 6.3.4) some fundamental parts of the SS 63 70 04 need to be replaced. This means e.g. that the modeling language UML should be used instead of EXPRESS, and that the transfer format XML should be used instead of STEP. There was in March 2001 no decision on when the revision of the standard should be finished. (NVDB 2000)

6.6 The U.S. Federal Geographic Data Committee

The U.S. Federal Geographic Data Committee (FGDC) was formed in 1990. Their main objective is to enable sharing and efficient transfer of spatial data between producers and users in the US in order to avoid duplication of efforts when collecting and maintaining spatial data. In April 1994 the FGDC was made responsible for the coordination of the establishment of the National Spatial Data Infrastructure (NSDI). NSDI is defined as "the technologies, policies, standards and human resources necessary to acquire, process, store, distribute and improve utilization of geospatial data" (U.S. *Executive Order 12906* 1994) The responsibility was accompanied by an obligation to develop standards for implementing the NSDI. The standardization is performed in cooperation with federal, state and local governments as well as other interested participants. As part of the NSDI the FGDC was also assigned to develop a National Geospatial Data Clearinghouse, which is a network of geographic data producers, managers and users for facilitating exchange of metadata. (U.S. *Executive Order 12906* 1994)

The FGDC has established a documentation standard called *Content Standard for Digital Geospatial Metadata* (CSDGM). The purpose of the standard is to provide consistency on the type of metadata to capture for documentation of geographic data in digital form. This means that it will provide a standardized set of terminology and definitions, primarily based on the four main roles of metadata (Federal Geographic Data Committee 1998, p(iv)):

- availability, which includes information about what datasets that exist for a particular region
- fitness for use, which includes the metadata needed to determine if a certain dataset can be used for a certain application
- access, which informs about how to acquire the dataset of interest, and
- transfer, which provides information about how to process and use a dataset.

The first version of this standard was approved in June 1994. It was implemented by many federal agencies as well as by other organizations within the US. The CSDGM standard was also used as reference for development of other metadata standards, e.g. the first draft of the global metadata standard ISO 15046-15. Due to later changes to the ISO metadata standard they do not fully agree anymore, but are supposed to be compliant to some extent. (Federal Geographic Data Committee 1998, pp(v)-(vi)) The ISO standard is now being revised and will according to plans be established as international standard ISO 19115 in July 2001. (ISO/TC 211 2000)

The CSDGM standard provides standard metadata elements for documenting several aspects of a geographic dataset. This includes (Federal Geographic Data Committee 1998, pp.1-52):

- Identification information provides basic information about the dataset.
- Data quality information provides a general assessment of the quality of the dataset with lineage, positional accuracy, attribute accuracy, logical consistency and completeness.
- Spatial data organization information describes how the geographic information is Represented, i.e. vector or raster.
- Spatial reference information describes the reference system used for coordinates in the dataset.
- Entity and attribute information provides details about entity types, their attributes and the domains from which attribute values may be assigned.
- Distribution information provides information about the distributor of and the options for acquiring the dataset.
- Metadata reference information gives information about the temporal validity of the metadata information.

It is possible to extend the CSDGM standard by defining additional metadata elements. Any implementation of the standard can thus either use the standard set of elements or be extended with several user-defined elements to enable a more detailed documentation of a dataset. (Federal Geographic Data Committee 1998, p.74)

6.7 Open GIS Consortium

The organization Open GIS Consortium (OGC), where GIS stands for Geographic Interoperability Specification, was founded in 1994 by the US major private GIS software companies. Later also software vendors from Europe and other countries, universities, authorities and cooperative organizations joined. The reason for starting the organization was to address the issue of incompatible standards in geographic information technology and developing a specification for this. (Open GIS Consortium)

Development of the specification is done through the OpenGIS Project. The aim of the project is to "[...] provide a comprehensive suite of open interface specifications that enable developers to write interoperating components that provide these capabilities." (Open GIS Consortium) The *OpenGIS Specification* (OGIS) is a software specification for geoprocessing interoperability. The OGC Technical Committee (TC) is the primary operational unit of the OpenGIS, since it creates OpenGIS interface specifications enabling interoperation and componentization of geoprocessing services. The task forces and their working groups do most of the work. However, also technology providers with well-developed technology can submit requests for having their solutions adopted as part of the OpenGIS.

The difference between OGC and the standardization organizations described previously is the way of addressing the problem of incompatible "standards". While other organizations try to normalize data format, data transfer, data dictionaries or metadata, OGC tries to create open, common interfaces between software system components so that the systems can use any data format internally. (Open GIS Consortium) In this way, any developer can produce software for any component, and it will work with other components in the market, for the same functionality (if they are also conformant to the interface specifications). (Kottman) Additionally, OGC tries to develop open software approaches that address inconsistent data dictionaries and metadata schemas. The first, by OGC accepted product, is the database interface ArcSDE 8 (ESRI 2000), which is used in ArcInfo 8 to enable communication between clients and databases connected in a network.

The *OpenGIS Specification* supports fundamental descriptions, e.g. of geometry, topology and reference systems, but does not present a standard schema model or set of schema models. Nor does the specification contain methods for defining the significance of geographic information. The Open Geodata Model (OGM) of the Specification instead contains a set of well-known types and structures, which the Technical Committee has agreed on to use for expressing any geographic schema model. These types and structures are defined in the form of interfaces, describing the behavior of the OGM primitives. One useful behavior is to expose information in well-known structures. These OGM building blocks can represent both low level (software engine) and high level (Information Community) schema models. (Open GIS Consortium)

The OGC is not a *de jure* standardization organization; the developed solutions therefore do not become standards. OGC instead adds implementation detail to the standards of *de jure* standardization organizations and thereby extends these standards whenever needed. OGC cooperates with ISO/TC 211 and other related organizations, with the aim that the general standards, in the parts where they overlap, should be mutual. The organizations have an agreement to maintain the technical configuration of their respective developments. OGC can be viewed as implementing (or adding an implementation layer on top of) ISO/TC211's "abstract standards". ISO will adopt an OGC specification when it meets certain requirements. The cooperation brings along a number of benefits, including (Open GIS Consortium):

- Assurance that OpenGIS conformant products will also conform closely with ISO TC/211 standards, which supports global product acceptance and global interoperability.
- Better OpenGIS Specifications and better ISO TC/211 standards.

The OGC has not carried out much work on quality interfaces. They are though taking a big step when adopting *ISO 19115 Geographic information – Metadata*. Eventually, OGC also plans to embrace *ISO 19113 Geographic information – Quality principles* wholesale, and to extend it with implementation level interface specifications. (Kottman)

So far the OGC specifications are relatively simple. They focus on spatial aspects, while it is uncertain whether they will consider semantic meaning of data and integration with other business related data. The work is, however, thought to have a great impact on the GIS market and lead to a "supplier standard" for the area. (Eklundh 1999, p.331)

6.8 Comparison of the quality models

The presentation of the geographic information standards previously in this chapter tells us that there are different ways to present quality of geographic information. Which one to use must be decided bearing in mind the specific application of the dataset. It is also important to consider future as well as present applications, in order to avoid having to reorganize data.

The three standardization organizations ISO, CEN and Stanli organize data in a quite similar way in their quality models ISO 19113, CEN ENV 12656 and the quality model of the Swedish Technical Framework. The data are structured into overview elements describing administrative and historical information of the dataset, and elements carrying quantitative information of the quality of the dataset. The CEN GDF standard differs from the three quality models, as it does not present a list of quality elements to use. Instead it deals with quality by defining different so-called checks to be performed on the data, which then will yield the quality elements' values. The Swedish standard SS 63 70 04 is similar to the GDF standard in the sense that they both present recommendations for the structure of road networks. The Swedish standard, however, is not at all concerned with quality of the defined network and will therefore not be discussed in this section. Open GIS, finally, can not be said to develop standards for geographic information, but rather implementing them and thereby facilitating the use of geographic data. Since the Open GIS Specification is not so much concerned with quality, it will neither be discussed in this section.

Regarding the fact that we want to quality label road related map data, the GDF standard would at first seem favorable to use. This standard deals, however, mainly with description and transfer of the data, and quality is only a small part of the standard. Therefore we might as well study ISO 19113, CEN ENV 12656 and the quality model of the Swedish Technical Framework. They actually use the same quality elements as the GDF standard, but present several levels of quality and more details, which facilitates the discussion of which quality parameters would be beneficial for routing applications (this discussion is presented in Chapter 7). In order to prepare for this discussion a comparison of ISO 19113, CEN ENV 12656 and the quality model of the Swedish Technical Framework is performed, and presented in Table 6.12. It will then be easier to find differences between the models and decide how to perform quality labeling of ISAB's map data.

ISO 19113	CEN ENV 12656	Swedish Technical Framework – Quality model
Quality overview elements	**Quality elements**	
		Administrating organization
Usage	Usage	
Purpose		
Lineage	Lineage	Lineage
Source information	Production	Original production
Process step or history information	Source and processes	Process history
	Homogeneity	
	Quality parameters	Quality concepts
Quality elements	**Quality parameters**	**Quality concepts**
Positional accuracy	Positional accuracy	Positional accuracy
Absolute or external accuracy	Relative horizontal accuracy	External positional accuracy, in plane
Relative or internal accuracy	Relative vertical accuracy	External positional accuracy, in height
Gridded data position accuracy		Internal positional accuracy, in plane
		Internal positional accuracy, in height
Completeness	Completeness	Completeness
Omission	Omission	Omission
Commission	Commission	Commission
Thematic accuracy	Semantic accuracy	Thematic accuracy
Classification correctness	Classification accuracy	Classification
Quantitative attribute accuracy	Agreement of an attribute	Quantitative attributes
Non-quantitative attribute correctness		Qualitative attributes
Logical consistency	Logical consistency	Logical consistency
Domain consistency	Validity of values	Domain
Topological consistency	Topological consistency	Topology
Format consistency	Format consistency	
Conceptual consistency	Geometric consistency	
	Semantic consistency	
Temporal accuracy	Temporal accuracy	
Accuracy in time measurement	Accuracy in time measurement	
Temporal validity	Temporal validity	
Temporal consistency	Last-update	
	Rate of change	
	Temporal lapse	

Table 6.12. The structures of ISO 19113, CEN ENV 12656 and the quality model of the Swedish Technical Framework.

In the continuation of this section, to avoid confusion, ISO's choice of words will be used when referring to data quality overview elements, data quality elements and data quality subelements. The term parameters refers to all of these.

Table 6.12 shows that the three quality models are quite similarly structured, although they have their own specific characteristics. All models have chosen to keep the number of parameters minimized. The standards therefore do not contain all possible parameters, but only those predicted to be the most useful ones. Additionally, all parameters do not have to be presented, but those that are applicable to the dataset. For users needing additional quality information, all of the standards allow creation of user-defined parameters.

A significant difference between the national standard (the quality model of the STR) and the international standards (ISO 19113 and CEN ENV 12656) is that the measures of the quality parameters are standardized in the STR. In 19113 and 12656 only incomplete lists of examples of measures are presented. This is due to the fact that national standards are developed for a more or less homogeneous group. International standards, on the other hand, are developed for a more heterogeneous group, where different producers are thought to have so different desires the measures should not be standardized. (Stenborg 2000, p.36) The STR being adjusted to Swedish needs has also caused a limitation of the quality information that can be presented. (STG 1998d, pp.15-16) As shown in Table 6.12, logical consistency in the STR does not present such a large variety of subelements, as do the international standards. Also temporal accuracy as presented in the international standards is not included in the STR at all.

Quality overview elements
Concerning the quality overview elements the difference between the models may at first seem very large. In reality, however, the same information probably will be documented in all three models. The information missing in one quality model, compared to the others, will be documented at another place, e.g. in the database specification or in the enclosed metadata description. (Stenborg 2000, p.39)

Quality elements
Concerning the quality elements all three models are equivalent. The only exception is that the STR does not use the element temporal accuracy, as do the international standards. Thematic accuracy (ISO and STR) is by CEN called semantic accuracy.

Quality subelements
Concerning the quality subelements there are some differences between the quality models:

- Positional accuracy. The models use different ways to define the accuracy; absolute (external) or relative (internal) accuracy in one, two or three dimensions. ISO additionally presents gridded data position accuracy.
- Thematic/semantic accuracy. A major difference between ISO's and the other models is that ISO classifies both objects and attributes, while the others only classify objects. Another difference between the models is that CEN looks at the number of objects correctly classified, the STR looks at the number of objects incorrectly classified while ISO does not define how to present the subelements.
- Logical consistency. All models present domain and topological consistency, ISO and CEN also present additional subelements.
- Temporal accuracy. ISO and CEN present several subelements concerning different aspects of time in the dataset. The STR does not use temporal accuracy at all.

The differences may seem large, but one should bear in mind that the models do not require all subelements to be used. Also, user-defined subelements may be used in all models. This enables all models to be adjusted to a certain application.

Descriptors of quality subelements

On the level of design or description of quality subelements it is more difficult to make a comparison between the models. Both ISO and CEN have defined sets of descriptors to describe the data quality subelements. The STR, on the other hand, does not present such a list for the standardized subelements, but do have a list of components to use when describing a user-defined quality subelement. The components of this list can also be used when describing the standardized parameters, but it is not necessary. The difference between the descriptors in the three models is not as large as it may seem, though. What is missing in one standard's quality model is found somewhere else or is expressed implicitly. In reality the quality models work in the same way.

Section conclusions

Despite the differences between the three quality models the resemblance is compelling. Since all models can be adjusted to fit any application, it makes no difference which quality model is complied with when discussing which quality parameters to use for quality labeling map data. It may even be beneficial to study the parameters of all three models in our discussion, since they show different aspects of the parameters and their usefulness. This will provide us with ideas of how to apply the different parameters on ISAB's data.

7 Quality parameters important for routing applications

This chapter aims at discussing different aspects of the quality parameters described in previous chapters, in order to find an appropriate quality labeling system for geographic data in routing applications.

The final quality of a map depends on the quality of the source datasets and the quality of the methods used to compile these datasets. In this chapter we discuss which quality parameters would be preferable to use when quality labeling maps used in routing applications. For definitions of the quality parameters see Section 5.4.

Generally, the more relevant quality parameters presented for a dataset the better. We feel that quality overview elements, in some sort of combination, must be enclosed with the dataset. The important information contained in the quality overview elements consists of the age and producer of the dataset, as well as the purpose and method for production. The quality elements can be considered to be more or less important, and we want to evaluate which are most appropriate when labeling data for routing applications. The quality elements are examined in the same order as they were described in Section 5.4.2.

For the different elements, data are divided into two categories:

- Map data, which include all data necessary for routing, i.e. the road network and important attributes.
- Extra data, which are not necessary for routing but have the ability to improve the routing services. The extra data can be divided into three sub categories: point data such as companies, segment data such as road capacities, and polygon data such as regions.

7.1 Completeness

For all types of objects it is important to state if the source dataset contains all objects of the real world or not. This information is included in data completeness. The other aspects of completeness, attribute completeness and value completeness, are closely linked. If an attribute is captured for an object it is of interest that this attribute has a value.

Map data

Completeness is an important element for the map data. Data completeness states how many of all existing roads are in a dataset, which is essential for estimating the correctness of the routes. Perhaps it would be useful to state this completeness in relation to the importance of the roads by dividing the roads into categories. The completeness would then be stated for each category, e.g. "the dataset contains 100% of all major roads, and 80% of all minor roads". This would provide us with a better idea of on which level the routing may be incorrect. Attribute completeness and value completeness are also important, although not as important as data completeness. It is more important that the roads are included in the dataset than that they have all attributes (with values). While we have the ability to add the values ourselves (e.g. by estimation), it is much more difficult to add missing roads since we most likely do not know where they are located. Thus the requirements on the data completeness would be higher than the requirements on attribute and value completeness.

Extra data
Completeness is an important element also for the extra data, e.g. stating the proportion of companies included in the data file, but it is not as essential as for the map data. Many of the datasets can be divided into categories, e.g. may companies be divided into large companies and small companies, in order to give a better picture of the completeness. Attribute completeness and value completeness here also play an important role. The attribute providing the location, which for a company may be coordinates or an address, is essential for being able to add the extra data to the map.

7.2 Logical consistency

Logical consistency, which means absence of apparent contradictions in a dataset, is an element difficult to measure. The assessment can be performed with a sample of data, or from knowledge about the most usual outcome of a certain procedure. Generally it can be said that logical consistency is very important since the data must follow all different aspects of established specifications in order to be used correctly.

Map data
Topological consistency is very important for the map data. All relation must be represented correctly in order for correct route generation. Domain consistency is also important when wanting to add data to the map data, where knowledge of the domain consistency will provide an estimation of the most probable result of the addition. Format consistency states a requirement on the source data file to be stored according to the specification of the format. It is important, as the information in that source file otherwise can not be used. All other aspects of consistency are also important, since it is only possible to add data and route correctly if the data are structured and specified according to established specifications.

Extra data
Extra data have the same requirements on consistency as do map data, except for topological consistency since the extra data topology is not important in this context. Data must follow established specifications in order to be added correctly or added at all.

7.3 Positional accuracy

Positional accuracy has various levels of importance depending on the particular application for which data shall be used. When routing in real-time, on coordinates, positional accuracy is very important in order to find the correct road to route on. In this case the density of the road network is an important factor, since it in a dense network is more difficult to connect a coordinate to the correct road. The requirements of positional accuracy for the densest area must therefore set the requirements for the entire dataset. A more specified aspect of positional accuracy is absolute or relative accuracy. The importance of these is also depending on the particular application. Absolute accuracy is important when finding the correct road starting from a coordinate determined e.g. with GPS. Relative accuracy is important when the starting point is marked in the map and routing is performed "internally" on map coordinates, without using GPS. Routing can also be based only on road names, which makes positional accuracy much less important, since coordinates are never involved. If routing is only performed in the horizontal plane, positional accuracy in height is not relevant.

Map data
Positional accuracy of the road network is, as described above, more or less important depending on the application. The requirement on positional accuracy has to be high enough to enable finding the correct location in the map with GPS determined coordinates, e.g. the starting point on the correct road segment.

Extra data
For point files positional accuracy is important if the point location is based on coordinates, they then have the same requirements as the map data described above. If the locations in the point files are address based, positional accuracy is not important. For other extra data files positional accuracy is important since they often are based on coordinates. Hence, it is important that the accuracy of the extra data matches the accuracy of the map data to make it possible to connect the extra data with the correct road segment.

7.4 Resolution

The quality element resolution is not being presented by any of the standardization organizations. Its contents are considered to appear in other elements or in other documents enclosed with the dataset. An example of this is thematic resolution that for attributes with values in categories is strongly linked to the domain, which is defined in the specification. If the domain defines the allowed values for an attribute to be e.g. forest, field, water and city, the number of categories is four, which is the thematic resolution. Another example is that information about spatial resolution may be presented in the specification.

Map data
Spatial resolution is important for knowing to which detail roads have been classified and registered. Private roads e.g. may be classified in the map data, while bicycle paths are not. A measure of the spatial resolution states the lowest rank category registered. This spatial resolution may be presented also in the specification or in the quality element completeness (if the data completeness is presented per category).

Thematic resolution is quite important when the domain for an attribute is presented as an interval, where it defines how many decimals should be presented. This may, however, also be implied by the thematic accuracy. Where the domain has predefined values, such as categories, the thematic resolution is indirectly stated in the domain as the number of categories. Hence, it must not be separately specified.

Extra data
Spatial resolution is important also for extra data in order to show to which detail an object has been registered. It can, however, also in this case be presented somewhere else. Considering thematic resolution the same applies as for map data.

7.5 Temporal accuracy

Temporal accuracy is an important quality element that most often states how old the information in a dataset is. Generally, this information provides an overall understanding of how well the dataset will suit specific applications. Some aspects of temporal accuracy may also be found in the overview elements lineage and usage.

Map data, extra data
Last update is an important subelement, stating if the dataset is old or new, which implies if it is valid. This enables an assessment of whether the dataset is appropriate to use.

Rate of change can be a useful subelement when wanting to update data at a regular basis. If e.g. a dataset has been used a certain time period, it is possible with help of the changing rate to estimate to which extent the data have changed. With a requirement on how much of the data that must be up to date, it is possible to determine if it is time to update the data.

Temporal lapse can be used for finding the age of the information in the file. However, if the subelement last update states the date for control, temporal lapse is unessential.

Temporal validity may be important depending on the contents of the dataset. If there is a possibility that not valid objects are present this should be stated. An example is the Øresund Bridge, which for a period of time was constructed, but not yet used for traffic.

7.6 Thematic accuracy

Thematic accuracy is an important quality element when evaluating the constituents of a map, since it states the proportion incorrectly described data. The subelement classification correctness is important when working with a map showing different objects. The parameter then indicates if the map "looks right" compared to the real world. When working with a map showing only one type of object (e.g. roads) thematic accuracy is not needed. If, however, the intention of the map is to show cartographic features other than roads, or roads divided into different categories, classification correctness becomes important.

Map data, extra data
Quantitative accuracy is presented for attributes with an interval domain, e.g. width, length and height. The attributes are considered right or wrong depending on whether the values are within a specified error margin or not. This error margin may also imply the thematic resolution; i.e. how many decimals should be presented.

Qualitative correctness is presented for attributes with categories and names. Depending on whether the attributes have been assigned correct "value" they are considered right or wrong.

Errors of major or minor importance can be found, but are all regarded as errors. The minor ones may, however, be corrected with the help of the software, e.g. slightly misspelled street names can be considered to be correct if the software is able to compare words.

8 Study of quality labeling among data producers

This chapter aims at giving an orientation on how some producers of geographic information label their datasets with quality information. Also the matter of confidentiality on this kind of information will be discussed.

In Chapter 7 the quality parameters important for routing applications were pointed out. To provide guidance to further discussions on development of a system for labeling ISAB's data with quality information, we are interested in the usage of these quality parameters by different producers of geographic datasets.

Our intentions were to study to what extent producers of digital road related data present quality of their data. It would also be interesting to find out how widespread quality labeling is among producers, and what quality parameters are used for this purpose. We decided to contact a few public authorities and private companies to study their quality labeling.

8.1 Examples of source quality

8.1.1 Swedish National Road Association

One example of quality labeling of source data is found at Vägverket, the Swedish National Road Association. Vägverket is presently developing a new dataset with roads covering the entire country, called the NVDB (nationell vägdatabas; the national road database) and has for this dataset specified a quality model. Since the NVDB is a new product, principles for quality labeling were already established when the NVDB quality model was specified. The NVDB therefore has a more extensive quality model compared to older products. Vägverket also requires the suppliers to quality label the datasets that are to be delivered to the NVDB.

The quality model of the NVDB is based on the model of the Swedish Technical Framework (STR) (Vägverket 1999b, p.14), described in Section 6.5.1, and states how quality of the road network should be presented. Almost all quality parameters in the STR are used for quality labeling the NVDB. The NVDB quality model, however, differs from the STR in the use of positional accuracy and thematic accuracy. For positional accuracy only external accuracy is presented and then separated for points and lines. (Vägverket 1999c, pp.7-9) Thematic accuracy is just described using accuracy of quantitative and qualitative attributes (Vägverket 1999c, pp.15-17). The NVDB quality model has furthermore specified some additional parameters (Vägverket 1999c, p.19-21):

- time for update
- temporal lapse
- accuracy in the length of a link in the road network, and
- internal accuracy of the linear expansion of an occurrence.

These additional parameters are needed since the quality model of the STR is developed for general applications concerning geographic information and thus do not fully match the requirements of the NVDB. Time for update and temporal lapse are aspects of temporal accuracy, which is not included in the STR, but present in CEN's quality model and ISO's quality principles.

Requirements on the measures of the quality parameters are explicitly given in the NVDB quality model, and expressed as a maximum rate of errors. (Vägverket 1999c, p.5) There are different requirements based on the purpose of the roads in the road network. National roads have the highest priority, meaning that e.g. the accuracy must be high, only a very small portion of the roads can be omitted, and the temporal lapse must be short. Hierarchically following national roads, are roads in densely populated areas that do not belong to the national roads. These are in their turn followed by roads outside densely populated areas. Private roads with minor importance have the lowest rank.

8.1.2 National Land Survey of Sweden

Another example of quality labeling is found at Lantmäteriet, the National Land Survey of Sweden. Lantmäteriet provides several analog and digital map products, which are labeled with basic quality information. Several of the products, e.g. the green map (terrain map) and the roads of the blue map, have a quality description presenting (Lantmäteriet 2000b):

- source material
- method of data capture
- positional accuracy
- completeness
- logical structure, and
- temporal accuracy.

The products of Lantmäteriet are older than the NVDB and therefore the quality parameters differ from the ones defined in the quality models presented in Chapter 6. The equivalence to the overview elements is not as exhaustive as in any of the quality models, but provide the most essential information enabling evaluation of the products. The first two parameters, source material and method of data capture, are excellent tools for deriving other quality measures, such as positional accuracy. The other parameters correspond to the quality elements in Chapter 5. They are, however, not presented with subelements, and several aspects of quality are missing compared to e.g. the STR. Measures of the quality parameters are not officially presented.

8.2 Problems with the study

Our original intention was to make more examples of how quality labeling is performed by different organizations and authorities. However, acquiring this information, particularly from private companies, turned out difficult. Navigation Technologies B.V. (NavTech) and Tele Atlas are two companies producing road data for use in navigation applications. Both companies market their products as high quality databases, but are very cautious when informing about their quality models and requirements on the measures of quality. The information is considered confidential and is not willingly made accessible for unauthorized. NavTech, through Quality Assurance Specialist Mr. Roel Morel, justifies the confidentiality with being a company policy. The main reason for the confidentiality is that NavTech itself develops the principles and tools for quality labeling, investing a lot of time and money. Since the business for navigable maps is new, existing tools for geographic data do not fulfill the special needs. Sharing information about this, without agreements on confidentiality, would not be wise considering the competition in the business. This policy is valid for several companies and authorities, making it difficult to acquire information on quality and impossible to present any information in this study.

Our thoughts on the confidentiality concerning quality labeling of geographic information are discussed in Chapter 12.

8.3 Results of the study

Due to the issue of confidentiality this study did not turn out as we had expected. We wanted to find out how quality labeling is used among data producers, but we only found two producers labeling their datasets. Although the NVDB quality model matches the important quality parameters pointed out in Chapter 7, this is not enough to make any conclusions about quality labeling among producers in general.

Since quality of geographic data has been given more attention by standardization organizations during the last years, we assume that new products are more likely to be labeled in a satisfying way. We also assume that companies, such as NavTech, do provide a detailed description of the quality of their databases, since those databases are their main product. Therefore, we believe that producers of new products are able to meet with ISAB's needs for having the important quality parameters enclosed with the dataset.

9 Study of how to use metadata management for quality information

This chapter aims at presenting the metadata editor of ArcInfo 8 and our thoughts on whether this metadata editor could be beneficial as a model for the quality management of map data.

ArcInfo 8, developed by Environmental Systems Research Institute Inc (ESRI), is software for dealing with geographic data. It has functionality for managing metadata of a geographic dataset. When working with quality of geographic information, metadata management is an interesting aspect that might yield a different approach to quality modeling since metadata includes quality information. Therefore ArcInfo 8 was studied for quality labeling of map data. Our understanding was that the metadata editor could read a dataset and thereby extract metadata from it, e.g. temporal aspects. We wanted to see what the editor could extract from the source dataset, and study the changes in the metadata as different operations were performed on the dataset.

9.1 Presentation of ArcInfo 8 and its metadata editor

ArcInfo Version 8 and Spatial Database Engine (ArcSDE) 8 were released in February 1999 after more than two years of development. (ESRI 2000) The new part of Version 8 is *Desktop ArcInfo*, which consists of three new applications; *ArcMap*, *ArcCatalog* and *ArcToolbox*. *ArcMap* is a map-centered application used to present, query and analyze vector and raster data. It contains an object-based editor with tools for creation and maintenance of geographic databases. *ArcCatalog* is a data-centered application enabling users to find, browse through and handle all types of data. *ArcToolbox* is a menu-based application used for geoprocessing operations such as data conversion, geographical analyzes, buffer analyzes and map transformations. The difference between Version 8 and the old version, Version 7, is that Version 8 in addition to the georelated model, which was used in the old version, also uses an object-oriented model. To store and work with geographically related objects the database interface ArcSDE 8 is used. This was the first product accepted by the Open GIS Consortium (OGC). ArcSDE 8 enables for a number of clients to communicate with several databases as long as they are connected in a network. The objectives, when designing ArcInfo 8, were to develop software able to work with several different data formats, both files and existing geodatabases. Among these formats are ArcInfo 8 coverages, shapefiles, CAD files, as well as two versions of ArcSDE. (ESRI 2000)

ArcInfo 8 manages metadata in the metadata editor of *ArcCatalog*. The metadata in the editor are structured into properties and documentation. Properties, e.g. bounding box coordinates, are derived automatically from the dataset when new metadata are created. Documentation, on the other hand, is descriptive information provided by the user, e.g. completeness and positional accuracy of the dataset. (ESRI Inc. 1999, p.18) The properties of the metadata are automatically updated with the current values every time it is displayed. (ESRI Inc. 1999, p.35) They are also automatically moved, copied and deleted along with the data source. Properties therefore become a part of the data source itself. The metadata managed by the editor are stored as XML data, either in a file alongside the data source or within a geodatabase. (ESRI Inc. 1999, p.117) Four metadata stylesheets are provided by default: ESRI, FGDC, FGDC FAQ and XML. (ESRI Inc. 1999, p.116) Each of them is stating a set of metadata elements to use. The FGDC stylesheet is based on the CSDGM standard, and categorizes metadata information into seven sections: identification, data quality, spatial data organization, spatial reference, entity and attribute, distribution, and metadata reference (see Section 6.6). The ESRI stylesheet presents many elements defined in the FGDC standard but extends it with more detailed properties of the data. (ESRI Inc. 1999, p.128)

9.2 Intentions and results of the study

Our original intentions were to evaluate the usefulness of the ArcInfo 8 metadata editor for e.g. acquiring quality information from a source dataset. Some producers do not label their data with all, for routing, important quality parameters, and we wanted to find out if the metadata editor could create any of these. Of course, we did not expect the editor to "invent" quality parameters such as positional accuracy and completeness. It would, however, with knowledge of transformations and other operations, be possible for the editor to create some kind of lineage information. From the lineage information other quality parameters could be implied. This would possibly provide a complete set of important quality parameters for the dataset. The set of quality parameters would serve as a foundation for developing a system for quality labeling of map data. Our second intention with this study was to see if the metadata editor could provide us with information about how quality parameters are altered when different operations are performed on the source dataset. This could be used as a simulation of how the quality of ISAB's map data is altered by the map generating process.

In reality, the editor uses only the quality parameters already present in the dataset, and is not able to create new ones, not even lineage information. Insufficiently labeled datasets therefore can not be improved by the editor, why ISAB needs source datasets to have sufficient quality information enclosed. Since we had no metadata editor compliant quality information enclosed with available source datasets, we were not able to evaluate the metadata management of the editor. It would have been possible for us to provide the dataset with fake quality information just to study the changes in it as operations were performed. We would, however, most likely not attain knowledge about the algorithms for metadata management, since these algorithms are developed by ESRI and considered confidential. To be able to benefit from the metadata management of ArcInfo 8, ISAB would have to perform all operations of the map generating process with ArcInfo 8, leaving the software developed at ISAB. This is obviously not an option, why we decided not to continue this study on a more detailed level.

PART III

Implementation of a Quality System for Map Management

10 Quality management systems

This chapter aims at giving an introduction to quality management systems. The chapter will provide mostly generic guidelines and ideas about how a quality system and its elements should function and what benefits they will provide.

10.1 What is a quality system?

A quality system is an important aid to control and govern the activities within an organization in order to introduce and maintain production of high quality products or services. A more formal definition of a quality system is according to STG (1997, p.9) "an organizational structure, responsibilities and authorities to control the business with respect to quality" (author's translation). A quality system thus contains all activities of a business that must be considered for assuring quality of products or services, e.g. documentation, development of company quality policies, and quality manuals. The system itself does not result in any assurance of quality, but provides an approach that enables achievements of consistent results and gradual improvements. Requirements and guidelines on how to introduce, implement and maintain a quality system can be found in the ISO 9000 series of standards.

Quality management consists mainly of three interacting activities; quality control, quality improvement and quality assurance. (Hoyle 1994, p.12) Quality control provides the means to fulfill requirements on quality. The control checks how well requirements are met and can be performed before, during and/or after final results are created. This enables prevention, discovery and correction of mistakes and shortages, which otherwise may lead to undesirable changes of quality. (Hoyle 1994, p.13; Andréason 1995, p.21) Quality improvement is concerned with an objective to gain higher quality. The improvements are made by correcting shortages identified in the quality controls. Quality assurance includes all planned and systematic activities, e.g. quality controls and improvements, necessary to provide confidence that a product or service will fulfill requirements on quality. (Hoyle 1994, p.18; Stenborg 2000, p.15)

Instructions for quality assurance often state e.g. that every step of the production process must be quality controlled separately before the next step is approached. If only the final product was to be controlled it would be almost impossible to locate possible errors introduced at an early stage of the process. Correcting these errors will hence require an enormous effort. Performing a gradual quality control will provide a continuous status of the quality throughout the work process and perhaps make controls of the final product/service by the time for delivery unnecessary.

10.2 Benefits from implementing a quality system

The purpose of a quality system is to enable an organization to achieve, sustain and improve quality. The system will help the organization to achieve quality goals (Hoyle

1994, p.21), both the goals identified by the organization itself as quality policies and the requirements implied by customer needs. The quality system puts the objective of the ISO 9000 series of standards into practice. Any problem, which may jeopardize customer satisfaction at any step of a production process, must be prevented. (Hoyle 1994, p.4)

Seen from the customers' point of view, there are two main reasons for organizations to implement a quality system. First, the customers can require that products and services meet the expectations and understandings agreed upon. The organization will, by means of the quality system, be able to assure the customers of this fulfillment. Secondly, the customers may want the business governed in a way assuring that potential suppliers meet with requirements of quality, today and in the future. With a quality system the supplier can assure the customers that systematic and functional efforts are made to thrive for high quality of products and services. The supplier himself also benefits from activities within the quality system concerned with quality assurance. Products and services will be produced in a cost efficient way, which will yield a strong financial position and enhance competitiveness in the long run. (Vägverket 1999a, p.3)

10.3 Documentation

One of the most important issues of implementing a quality system is, according to ISO 9000, documentation. (Hoyle 1994, p.225) The development and use of documentation is supposed to be a dynamic and profitable operation (CEN 1994, p.8), and should help customers as well as auditors to understand how a quality management system addresses the requirements of the ISO 9000 standards. Both the quality system and its documentation are supposed to reflect the way the organization operates, its processes and routines. This should be done efficiently and easily comprehensible in order to create an aid for all employees. The documentation should enlighten and explain the business, but not include more than is needed for the situation. (STG 1994, p.20) Therefore there is, just like with any other issue in the ISO 9000 series of standards, no standard way to design these documents, nor how many should be established. (Peach 1992, p.118; Hoyle 1994, p.139)

Appropriate documentation is essential for several activities within the quality system (CEN 1994, p.8):

- attaining required quality
- evaluating the quality system
- carrying out quality improvements, and
- maintaining the improvements.

Organizations should document these activities to such an extent that everyone understands the purpose of them. This makes it possible to establish how the activities are carried out and to measure their specific performance. Documentation of the activities will also increase the possibility to determine effects of changes in the work processes, made to improve the quality.

According to Peach (1992) the documentation of a quality system should, preferably, be arranged hierarchically consisting of four layers. Each layer develops a steadily increasing level of detail about the organization's operations and methods. The layers are;

quality manual, company operating procedures, work instructions, and records. (Peach 1992, p.118) Figure 10.1 shows how the layers are hierarchically structured and what each layer should represent.

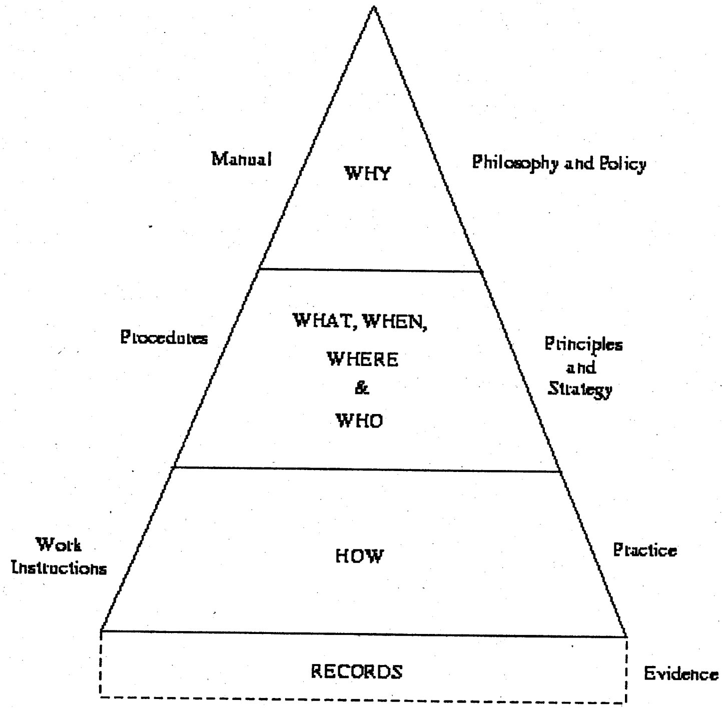

Figure 10.1. The hierarchical structure of the documentation in a quality system. Redrawn from Peach (1992 p.119).

The quality manual is a toplevel document, which contains stated quality policies, the objectives for implementing the quality system and a general description of the system. (Hoyle 1994, p.124) It is intended to answer a *why* question. The quality manual is in the documentation hierarchy followed by procedures, which describe the activities that must be carried out within the quality system. These documents explain the *what, when, where* and *who* of the system and record the purpose of the procedure, responsibilities of carrying out the activities and simple descriptions of the activities. (Peach 1992, p.119) The work instructions are more specific than the documented procedures. The level of the documented work instructions is usually machine-, task, or product-specific. Someone who knows and performs the task should write the documentation. This will make it much more likely that the documentation is current, properly applied, and that any required changes are made. (Peach 1992, p.119) The base in the documentation pyramid in Figure 10.1 comprises records. The records document knowledge about and experiences from the quality system. They constitute the evidence that all activities are carried out according to the intentions of quality system, and thus ensure customers, auditors and employees that the organization will fulfill the requirements for quality.

10.4 Introducing a quality labeling system on ISAB

The ISO 9000 requirements and guidelines for introducing a quality system are generic. It is therefore a quite complicated task to apply those requirements and guidelines to the business of a specific organization such as ISAB. Our project only includes the development of a quality labeling system for ISAB's map data. It is, however, still appropriate to use the guidelines of ISO 9000.

According to Rydstedt, the first step of introducing a quality system for map management is to establish a quality plan, in which requirements on quality are to be specified. The next step is to make sure that acquired source data meet the specified quality and that the quality is retained during updating and maintaining processes. An important issue is to state the routines for quality assurance, i.e. how to verify that the data meet the quality that has been specified. It is a matter of forming routines for creation and maintenance of the data in a way that the quality issues are satisfied.

What we want to do, to assure a certain quality on ISAB's map data, is to develop a system for labeling the data with quality information. Therefore only some aspects of the quality system ideas will be implemented. The most important issue, as we see it, is documentation. ISAB's map generating process must be documented in order to see how the data are processed and in what steps the quality of the data is altered, to the better or to the worse. These steps can then be evaluated to provide information about the quality change of the map data.

11 Development of a quality system

This chapter presents the outcome of this project. We here build a foundation for implementing a system for quality labeling of map data. We also have produced a proposal for how this quality labeling can be done and an appropriate structure for how to present the quality information.

The aim of this Master of Science Thesis is, as described in Chapter 1, to develop a quality system for map management in the field of location-based services. As found in previous chapters, this is a rather extensive process. The most beneficial part of the quality system would be to quality label the map data to enable evaluation of how much the map data contribute to the quality of the services. In order to label the map data with quality information it is necessary to identify the steps in the map generating process. Then the steps that may result in a change of quality can be discerned. This will provide us with the foundation of how and where in the process it will be suitable to label data with quality information.

11.1 Identifying the steps of the process

We have started documenting work instructions for the map generating process at ISAB. These work instructions constitute one part of the documentation that is recommended by ISO 9000 to be used when implementing quality systems for a business. The instructions should be exhaustive enough to enable identification of the process steps and thereby also the steps affecting the quality of the map. The collection of instructions must be a dynamic document that should be updated whenever changes occur. Every time a new source dataset is included in the process, the new steps must be added to the work instructions.

The work instructions are written with the objective that any person with appropriate background knowledge should be able to generate a map following the instructions. They will provide a detailed description of the different methods to generate a map to use as a foundation for ISAB's services. We have chosen to structure the instructions in a way that each method for map generation is described in sections called *preprocessing, map generation* and *postprocessing*. In *preprocessing* is presented what has to be considered in order to prepare the source map data for conversion into ISAB's map data format, and the source extra data for addition to the routable map. In *map generation* the conversion of the source map data into a routable map is performed. Finally, in *postprocessing*, the extra data are added to the map. These sections actually are the main steps in the map generating process and the ones to evaluate and label with quality, in order to get an understanding of the final quality of the map. The steps can be illustrated schematically as in Figure 11.1 and are further described in the text following the figure.

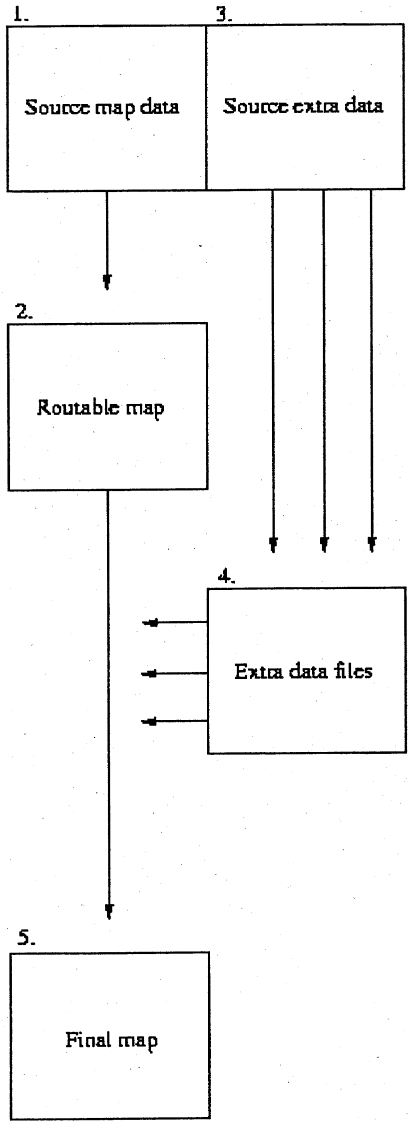

Figure 11.1. Overview of the map generation process at ISAB.

1. The map generating process starts with acquiring appropriate map data from different producers of digital road data. This *source map data* should contain a road network with attributes for routing, such as speed limits, street names and driving directions.

2. The map data are then converted into ISAB's map data format, which requires a minimum of storage capacity and is optimized for navigation applications. Different conversion methods are used depending on the format of the source map data, all resulting in a *routable map*. Since the routable map contains only a basic set of attributes, and therefore does not hold much information, more attributes must be added.

3. Additional data (extra data) are acquired from different sources. These *source extra data* contain e.g. companies, road capacities and regions and will enable ISAB to offer more and better services.

4. The extra data must be converted into a format appropriate for addition to the routable map. This format will in the continuation be called *extra data files*. For every new attribute added to the map, the map gets more comprehensive by containing more information.

5. The *final map* holds all attributes needed for ISAB's different services.

11.2 Quality changes in the process steps

In the previous section the steps in the map generating process were identified. They will in this section be evaluated with the objective to find out how they contribute to the quality of the final map. The importance of quality labeling of source map data and source extra data is implied by the discussion in Chapter 7. It was there realized that several aspects of quality are important for routing applications and must be presented with the source data. This source data quality will be affected by the conversions and additions in the map generating process. Therefore, thorough quality documentation of the process is important.

For discussing quality labeling, we have divided the map generating process into two sections. The section *Processes on map data* includes conversion of map data into ISAB's map data format. The section *Processes on extra data* includes conversion of extra data into extra data files and addition of them to the routable map. Each section starts with a short summary of the discussion in Chapter 7, showing which quality aspects are important for respective source data.

11.2.1 Processes on map data

Source map data

As found in Chapter 7 several of the quality elements are important to present with the source map data:

- Completeness, i.e. knowledge of how many of the objects in the real world that actually are represented in the map, is essential in order to estimate the quality of the services. For the objects in the map it is essential that they have attributes with values, e.g. roads must have speed limits.
- Consistency is important for map data, especially topology and format consistency, since the road network must have correct topological relations, and all data need to be

stored correctly, according to specified structure, to enable conversion into ISAB's format.
- Absolute positional accuracy of the road network must be high enough to enable use in GPS applications.
- Resolution of the road network is important in order to show at which scale routing is possible.
- Temporal accuracy is important since its time reports give an understanding of the usefulness of the map data.
- Thematic accuracy is important for quantitative and qualitative attributes in order to show if they have correct values.

Conversion into ISAB's map data format
Conversion of source map data into ISAB's map data format means that only the format of the map data is changed. The conversion methods are developed to manage the different formats of the source data. Thus, if the source map data have a high quality the conversion methods function well. Assuming high quality source map data the conversion process does not directly affect many of the quality elements.

The completeness of the data is changed when incorrect records are not accepted by the conversion process and therefore not included in the routable map. This is, however, not only a disadvantage. The completeness is deteriorated, but since the faulty records are omitted the other quality elements are indirectly improved. The total quality of the map is not improved in this step, but it is probably not deteriorated to the same extent as the completeness.

Depending on the reference system of the source map data, transformations may need to be performed when converting the data into ISAB's map data format. This will to some extent affect the absolute positional accuracy. The size of the change of the positional accuracy can usually be derived from the particular transformation.

11.2.2 Processes on extra data

Source extra data
As found in Chapter 7 several of the quality elements are important to present with the source extra data:
- Completeness, i.e. knowledge of how many of the objects in the real world that actually are represented in the extra data, is important. It is also essential that these objects have attributes with values, e.g. companies must have an address.
- Consistency (not topological) of the extra data is important for enabling addition to the map data.
- Absolute positional accuracy for extra data with coordinate based locations must match the accuracy of the map data in order to be connected to the correct road segment.
- Resolution for extra data is important in order to show with which detail attributes are presented.
- Temporal accuracy is important since its time reports give an understanding of the usefulness of the extra data.

- Thematic accuracy is important for quantitative and qualitative attributes in order to show if they have correct values.

Conversion into extra data files

Conversion of extra data into extra data files may be performed by several different processes depending on the original form of the extra data. Analog extra data need to be converted into digital form, which can be done by digitalization or scanning. Digital extra data must sometimes be manually edited e.g. to match the road network in the routable map. Digitalization and manual editing require human intervention, which will affect the quality of the data. Digital extra data may also need to be edited by automatic processes, such as geocoding and transformations. These processes do not, unlike the manual editing processes, require human intervention. Thus there can be no errors due to human mistakes, but possibly due to the automatic processes. Finally, the format of the extra data is changed when converting them into extra data files.

All work with human intervention involves a risk of affecting some aspects of the quality of the extra data. In manual editing the extra data is adjusted to the routable map. The absolute positional accuracy of the extra data is then deteriorated, while the relative positional accuracy in relation to the map is improved. This is advantageous when adding the extra data to the map, but disadvantageous when using the map in applications based on coordinates. Also digitalization affects the positional accuracy of the extra data, to which extent depends on the skill of the analyst. The completeness is affected by mistakes in the digitalization process; data completeness if objects in the analogue material are omitted, attribute completeness if an attribute is not captured, and value completeness if the value for an attribute is not registered. The rules for domain consistency may be violated if an incorrect value is assigned to an attribute. The thematic accuracy is affected if the attributes are given incorrect values.

The automatic editing processes affect the quality elements differently depending on how they are used. Transformations are used on extra data with coordinate based locations, and affects positional accuracy only. Geocoding on the other hand, is used on extra data that have address based locations, with the aim to match the addresses in the routable map. It will therefore not affect positional accuracy, but instead completeness, and indirectly consistency and thematic accuracy. The completeness of the extra data will deteriorate due to misspelling of addresses or an incorrect address format. This deterioration is not only disadvantageous for the final map, since the data omitted from the extra data in fact are the ones that are faulty. Some aspects of quality of the extra data, such as consistency and thematic accuracy, are thus increasing with this completeness deterioration. The format consistency improves since addresses that do not agree with the specified format are omitted. The thematic accuracy may both increase and deteriorate. The reasons for increase are similar to the ones for consistency, objects with misspelled addresses do not match the addresses in the map and are consequently omitted from the extra data. The increase may also be due to corrections of misspellings, if the geocoding process is set to an address match lower than 100%. On the other hand, this might also result in a lower thematic accuracy of the extra data if two similar addresses are confused. The geocoding process changes the address based extra data into being coordinate based.

This results in a positional accuracy of the extra data, which will be the same as the positional accuracy of the map since the coordinates are derived from it.

The different source extra data are converted into extra data files. Some of the extra data have been processed by the methods described in this section, and a selection of them adjusted to the routable map. Other extra data are converted directly from source files to extra data files. The conversion into extra data files does not affect the quality since the extra data is only restructured, not altered.

Addition of extra data to the map
Extra data are added to the routable map to increase the number of attributes and enable a more extended use of the map. The quality of the added attributes can never be higher than the quality of the extra data files. It may, however, be lower depending on the match with the routable map. The successfulness of the match depends on the extra data file, the routable map and the addition processes.

Extra data files that have been created based on the routable map will most likely be added to the map without major changes of quality. The reason for this is that the selection of matching records was made already in the conversion processes. If incorrect records are still present in the extra data file, they will not be added to the map. Hence, the completeness will be deteriorated and other elements indirectly improved, as when source map data are converted into ISAB's map data format.

Extra data files that have been created without being adjusted to the routable map will not be added to the map without changes of the quality of the extra data. The selection of matching records will be performed in this step instead of the previous. This results in the same change of quality as when the selection was made in the previous step, e.g. for geocoding.

11.3 Quality labeling of ISAB's map data

The quality of ISAB's services depends on the quality of the final map, which in turn depends on the quality of the source data and the map generating process. The source data quality is thus the foundation of the map quality, and is set by the producers. It is for ISAB not possible to affect the quality of the source datasets. The final map, however, can be improved by addition of more source datasets, which results in improved performance of the services. Every compiled part of the final map will still have its own quality, only the quality of the final map has improved.

We would like to have two measures of the quality for the map data. The first measure is more detailed, following each dataset throughout the map generating process, registering the quality change in every step. The other measure provides the quality of the final map.

The first quality measure is for internal use only and should document how the important quality parameters of the source dataset are affected in every step. This can reveal insufficient methods, i.e. where quality shortages occur, and be the base when evaluating which method is most preferable to use. The measure will also provide information about how the quality of a potential dataset will change throughout the process. The measure should be documented for each source dataset in some sort of "quality report". It should include the quality effects of every process used for conversion and addition of the dataset. The software of the map generating process should in every step be able to register changes in the quality, e.g. that 850 of 867 features were approved by the process (e.g. added to the map or geocoded), and also which features were not approved and possibly the reasons for the non-approval. The software should generate as many of these "quality comments" as possible. Additional changes of the quality must be reported by the person in charge of the specific step. Finally, the original source quality and the changes in this quality throughout the map generating process are, for every source dataset, compiled to a quality of the specific dataset. A compilation of this quality of all source datasets will constitute the total quality of the final map, the second quality measure.

The second quality measure should state the quality of the final map. The map is here considered as one unit, not a compilation of several datasets, but is quality labeled per category of data. If the map e.g. has companies acquired from three different source datasets, the company quality of the map should comprise the compiled quality of all company datasets. The measure can be used for an assessment of the map quality, and can provide clues to why routing is sometimes incorrect. If e.g. the topology is bad, the routes will obviously not be correct. The measure can also be used to show customers that the map has a known and documented quality.

The structure of the second quality measure should be set in some sort of "quality specification". To set the structure, the database specification should be considered, since this defines what objects should be in the map, their attributes and domains. The values of the second measure must come from the first measure, since this is where the quality is compiled.

Appendix A shows an example of how the second quality measure may be presented for a simple map with only a few attributes.

11.4 Methods for quality evaluation

The system for quality labeling of ISAB's map data is intended to serve as a foundation for further quality thinking within map management. This could e.g. include methods for deciding which datasets should be used and/or bought for improving the map data, by a quality comparison.

It is difficult to assess which datasets are most appropriate to use for improving the map. The major difficulty is that the quality parameters represent very different aspects of quality. This makes it impossible to make a direct comparison between two parameters. One example might be comparing two potential new datasets, one with high positional accuracy and the other with high completeness. If the datasets are to be used for a specific application, this application can be evaluated concerning which quality parameters are the most important, resulting in determination of which dataset is the fittest. This decision is based on the functional quality of the dataset.

ISAB provides several services and thus do not use the map data in only one application. Therefore, a more descriptive quality measure of the map data is needed, where all relevant quality parameters must be considered. To enable this "descriptive comparison" a method for evaluating whether, and to which degree, a new dataset will improve the quality of the final map, must be established.

One possible solution for quality evaluation is to establish some sort of "quality function". This function would consider all aspects of quality to enable assessment of a total quality of a dataset. Each of the parameters could then be assigned a constant, the value of which depends on the importance of the parameter. The function would yield a "quality value" of the dataset to be used in comparison with other potential datasets. It would be preferable also to enable assessment of the most probable outcome of addition to the map for each potential new dataset. This would require a "simulation" of the map generating process integrated in the "quality function".

11.5 Effects

The system for quality labeling will have a rather great impact on ISAB's map generating process. Not only the quality of already existing materials must be specified, but also the quality goals expected when implementing the system. The quality thinking must imbue the processes and employees must be aware of the importance of quality. Source datasets can not be acquired just for having the correct content concerning objects and attributes, but also different quality aspects of the content must be considered. The quality must match the goals and requirements specified by ISAB.

The software generating the maps must be somewhat reconstructed in order to register quality changes in the processes. Additionally, new software must be developed to manage the quality system and to compile the overall quality measures for the final map.

As for the work instructions, they must be completed to also hold instructions for registering quality changes in every step. Therefore, the new steps and also the quality requirements must be added to the work instructions every time a new source dataset is included in the process.

12 Conclusions

This chapter presents the conclusions we have made from working with this Master of Science Thesis. The chapter summarizes the "subconclusions" made in previous chapters.

For the development of the system for quality labeling of ISAB's map data, three standardized quality models have been studied; ISO 19113, CEN ENV 12656 and the quality model of the Swedish Technical framework. Despite the differences between the models the resemblance is compelling. Additionally, all models provide the possibility of user defined quality parameters, and can therefore be adjusted to fit any application. Therefore, it makes no difference which quality model is complied with when discussing which quality parameters to use for quality labeling of routable map data. We have therefore considered parameters of all three models, since they show different aspects of the parameters and their usefulness.

The quality parameters found important for routing applications can be divided into quality overview elements and quality elements. The important information contained in the quality overview elements consists of the age and producer of the map data, as well as the purpose and method for production. The important quality elements are:

- Completeness, which is essential in order to estimate the quality of the map data considering the proportion of entities represented in the map data. For the objects in the map it is essential that they have attributes with values.
- Consistency, which is important for map data, especially topology and format consistency, since the road network must have correct topological relations, and all data need to be stored correctly.
- Positional accuracy (absolute and relative), which must be high enough to enable use of the map data in GPS applications. Extra data to be added to the road network must have positional accuracy matching the one of the map.
- Resolution, which is important in order to show at which scale routing is possible, and with which detail attributes are presented.
- Temporal accuracy, which is important since its time reports give an understanding of the usefulness of the map data.
- Thematic accuracy, which is important in order to show if quantitative and qualitative attributes have correct values.

Data producers label their data to various extents. Our study showed that products developed in the last few years are more likely to be labeled according to an accepted quality model than older products. However, some producers keep this information confidential why we have not been able to make any general conclusions about the use of quality labeling.

We studied the metadata editor of ArcInfo 8 with the aim to find a way to improve insufficiently labeled datasets. The editor, however, only used the quality parameters already present in the dataset, and was not able to create new ones. Insufficient datasets therefore can not be improved by the editor, why source datasets must have sufficient quality information enclosed.

The most important issue, when introducing a system for quality labeling is, according to ISO 9000, documentation. ISAB's map generating process must be documented in order to see how the data are processed and in what steps the quality of the data is altered, to the better or to the worse. These steps are evaluated to provide information about the quality change of the map data.

The quality labeling system is to be based on the producers' quality labeling of the source data. The system will to this original quality add the quality changes in the steps of the map generating process. This will yield two quality measures. The first measure is more detailed, and follows each source dataset throughout the map generating process. The second measure provides the final quality of each data type in the final map.

13 Discussion

Systems for quality labeling of geographic information is a rather new application area. Therefore, we have run into some difficulties when trying to find relevant information. There is a lot of information about quality on certain subjects, such as quality assurance of medical care, but almost nothing to be found on geographic information. Generic guidelines for implementing a quality system can be found in the literature about the ISO 9000 series, but since these guidelines are established to match any business, nothing is especially applicable to geographic information.

We believe that quality of geographic information will be given more attention in the future since the use of digital geographic data is constantly increasing. It would be an obvious waste of company resources to purchase or capture geographic data, not knowing the quality of the data. The development of new geographic information standards, e.g. the ISO 19100 series, will also increase the awareness of quality among producers. Since there previously have been only a few guidelines on the subject, it has been difficult to proceed with e.g. quality labeling although producers might have been interested in this. Proceedings would have resulted in an uncertainty whether the labeling was performed accurately, and a risk of investing in something that might not be useful in the future. Due to the increasing use of digital data and the new geographic information standards, we feel that if we were to write this Master of Science Thesis in a couple of years, we would find more information on the subject.

Working with this Master of Science Thesis, we have experienced that the area of standardization is very confusing. The reasons are that standardization organizations develop standards within the same areas, why it is difficult to know which one to use. It is also difficult to understand how the organizations and the standards are related to each other. One example is the relation between ISO, CEN and the Swedish Technical Framework. ISO has published some 12 000 standards, and is presently working on 30 standards that are strictly concerned with different aspects of geographic information. Furthermore, several of the 12 000 other standards may also to some extent be applicable, why there are many ISO standards concerned with geographic information. Corresponding standards have been published by CEN, adjusted to European needs, and by Stanli as part of the Swedish Technical Framework, adjusted to Swedish needs. It is thus difficult to decide which standard is most appropriate for a certain application. Especially since there is not much literature to find on the subject.

When examining quality parameters to find out how important they are for labeling ISAB's map data, we have experienced that the parameters are not very well described in the standards. We found it difficult to assess exactly how they should be used and how they could be applied to ISAB's map data. As a complement to the standards, we needed to study more literature on the subject, in order to understand what the parameters mean and how they can be used. It is, however, not obvious that the meaning of the parameters is identical in the standards and in the complementary literature, since the parameters can be used in many different applications.

It has, due to the issue of confidentiality, been difficult for us to acquire information about quality labeling among data producers. The arguments NavTech and other companies use are quite relevant. However, it seems a bit strange that information about the main marketing argument can not be acquired without signing agreements on confidentiality. We feel that it is understandable if a company does not want to share information about quality assessment methods, since these are typically developed by the company itself. The assessment methods can then be considered as one of the company's products, and therefore confidential. However, sharing information about what quality parameters are used for labeling and the values of these, we do not find company specific and confidential. The set of parameters used have most likely already been specified by a standardization organization and not invented by the company. The company has only established the combination of parameters to use. We therefore do not feel that sharing this combination would result in any disadvantages for the company. The values of the parameters are somewhat revealing and might be kept confidential. However, since anyone who wants to know the quality of the company's products, e.g. potential customers, can acquire this information, we feel that there is really no need for keeping it confidential.

14 References

Published

Aalders H. (1997). *Quality metrics for GIS*, in Kraak M.J. et al. *Advances in GIS research II*, pp.277-286, Taylor & Francis Ltd, UK.

Andréason K. ed. (1995). *Grundkurs i kvalitetssäkring*, Quality Management Training Institute (QMI), Halmstad, Sweden.

Campbell J.B. (1996). *Introduction to Remote Sensing* 2nd edition, Taylor & Francis Ltd, UK.

Cederholm T. (2000). *Huvudsaken är modellernas innehåll – inte språken*, Stanli Nyheter 3 2000, pp.4-5.

Celander L. (1999). *STEP: En skonsam men effektiv introduktion*, Azelia AB, Göteborg, Sweden.

CEN (1994). *EN ISO 9000-1 Quality management and quality assurance standards – Part 1: Guidelines for selection and use.*

CEN (1995). *Geographic Data Files* 1st draft.

CEN (1998a). *ENV 12656 Geographic information – Data description – Quality.*

de Blij H.J. et al. (1996). *Physical Geography of the global environment* 2nd edition, John Wiley & Sons Inc, USA.

Dent B.D. (1999). *Cartography – Thematic Map Design* 5th edition, WCB McGraw-Hill, USA.

Duckham M. *Implementing an object-oriented approach to data quality*, Department of Geography and Topographic Science, University of Glasgow.

Eklundh L. ed. (1999). *Geografisk informationsbehandling – metoder och tillämpningar*, Bygforskningrådet, Stockholm, Sweden.

ESRI Inc. (1999). *Using ArcCatalog* - Draft, Environmental Systems Research Institute (ESRI) Inc., USA.

Essén K. (1999). Internationellt ramverk – nya tag på möte i Kyoto, Stanli Nyheter 4 1999, pp.14-15.

Essén K. (2000). *Internationella regler för datautbyte*, Stanli Nyheter 3 2000, pp.6,8.

Federal Geographic Data Committee (1998). *FGDC-STD-001-1998 Content standard for digital geospatial metadata*, Federal Geographic Data Committee, Washington, D.C.

Hofmann-Wellenhof B. et al. (1997). *GPS Theory and practice* 4th edition, Springer-Verlag, Wien, Austria.

Hoyle D. (1994). *ISO 9000 quality systems handbook* 2nd edition, Butterworth-Heinermann.

International Cartographic Association (1973). *Multilingual Dictionary of Technical Terms in Cartography*, Franz Steiner Verlag Gmbh, Germany.

ISO (2000a). *ISO/DIS 19101 Geographic Information – Reference model.*

ISO (2000b). *ISO/DIS 19113 Geographic Information – Quality principles.*

Jivall L. and Lidberg M. *SWEREF 99 – an Updated EUREF Realisation for Sweden*, Lantmäteriverket, Gävle, Sweden.

João E.M. (1998). *Causes and Consequences of Map Generalisation*, Taylor & Francis Ltd, UK.

Lantmäteriverket. *Introduktion till koordinatsystem och karttransformationer.*

Lantmäteriverket (1996a). *HMK-Geodesi, GPS*, Lantmäteriverket, Gävle, Sweden.

Lantmäteriverket (1996b). *HMK-Geodesi, Stommätning*, Lantmäteriverket, Gävle Sweden.

Lantmäteriverket (1996c). *HMK-Kartografi*, Lantmäteriverket, Gävle, Sweden.

Lantmäteriverket (1997). *HMK-Databaser*, Lantmäteriverket, Gävle Sweden.

Malys et al. (1997). *Refinements to The World Geodetic System 1984*, NIMA.

Molenaar M. (1998). *An Introduction to the Theory of Spatial Object Modelling for GIS*, Taylor & Francis Ltd, UK.

Nämnden för skoglig fjärranalys (1993). *Flygbildsteknik och fjärranalys*, Skogsstyrelsen, Jönköping, Sweden.

Palm C. (1998). *Kartteknik – En kartmakares anteckningar.*

Peach R.W. (1992). *The ISO 9000 Handbook*, CEEM Information Services, USA.

Pilesjö P. and Cederin K. (1998). *Introduktion till Geografiska informationssystem.*

Reit G. (1994). *SWEREF93 - ett nytt svenskt referenssystem,* LMV-rapport 1994:28, Lantmäteriverket, Gävle, Sweden.

Robinson A.H. et al. (1995). *Elements of Cartography* 6th edition, John Wiley & Sons Inc, Canada.

Stenborg E. (2000). *STG Hb 172 Geografisk Information – Kvalitetsmärkning*, SIS Förlag AB, Stockholm, Sweden.

SIS (1999). *SS 63 70 04 Väg- och järnvägsnät – Begrepps- och tillämpningsmodell*, SIS Förlag AB, Stockholm, Sweden.

STG (1994). *Framtid med kvalitet 9000- en introduktion till verksamhetsstyrning med ISO 9000 som modell*, SIS Kvalitetsforum, Stockholm, Sweden.

STG (1997). *ISO 9000 för mindre företag*, SIS, Stockholm, Sweden.

STG (1998a). *STG Hb 171 Tekniskt Ramverk för geografisk information*, SIS Förlag AB, Stockholm, Sweden.

STG (1998b). *STG Hb 171 Tekniskt Ramverk för geografisk information – Kvalitetsmodell*, SIS Förlag AB, Stockholm, Sweden.

STG (1998c). *STG Hb 171 Tekniskt Ramverk för geografisk information – Ramverket för samverkande informationssystem*, SIS Förlag AB, Stockholm, Sweden.

STG (1998d). *STG Hb 171 Tekniskt Ramverk för geografisk information – Översikt och grundläggande begrepp*, SIS Förlag AB, Stockholm, Sweden.

Ussisoo I. (1977). *Kartprojektioner*, LMV-rapport 1977:6 Lantmäteriverket, Gävle, Sweden.

Veregin H. (1999). *Data quality parameters*, in Longley et al. *Geographical Information Systems – Principles and Technical Issues* 2nd edition, chapter 12, John Wiley & Sons Inc, USA.

Wikström H. (1998). PM: *Läge och trender inom internationell standardisering*, Lantmäteriverket, Gävle, Sweden.

Worboys M. (1995). *GIS: A Computing Perspective*, Taylor & Francis Ltd, UK.

Unpublished

CEN (1998b). *CR 287002:1998 Geographic Information – Fundamentals – Overview*, found at http://www.agi.org.uk/pages/structure/committees/standardsold/36WDCENover.htm (November 16 2000).

ESRI Inc. (1998). ESRI White Paper: *ESRI Shapefile Technical Description*, found at http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf (October 27 2000).

ISO (1998). *Selection and use of ISO 9000*, found at ISO 9000 and ISO 14000, http://www.iso.ch/9000e/selusee.pdf (November 6 2000).

ISO/TC 176/SC 2 (1997). ISO/TC 176/SC 2/N *Quality management principles and Guidelines on their appplication*, found at ISO/TC 176/SC 2 Homepage http://www.bsi.org.uk/iso-tc176-sc2 (January 29 2000).

ISO/TC 176/SC 2 (2000). ISO/TC 176/SC 2/N524 *Guidance on ISO 9001:2000 clause 1.2 'Application'* (draft), found at ISO/TC 176/SC 2 Homepage, http://www.bsi.org.uk/iso-tc176-sc2/ (December 15 2000).

MapInfo Corporation (1999). *The MapInfo Interchange File (MIF) Format Specification*, MapInfo Corporation, Troy, NY, found at http://www.mapinfo.com/community/free/library/interchange_file.pdf (November 1 2000).

NIMA (2000). *NIMA TR8350.2 DoD WGS84 – its definition and relationships with local geodetic systems* 3rd edition, amendment 1, found at NIMA Homepage http://164.214.2.59/ Publications – Technical Report 8350.2 (February 15 2001).

OMG (2000). *OMG Unified Modeling Language Specification* Version 1.3, found at Unified Modeling Language, http://www.omg.org/technology/documents/formal/unified_modeling_language.htm (January 15 2001).

Open GIS Consortium. *Corporate brochure*, found at OGC Corporate Brochure, http://www.opengis.org/info/brochure/index.htm (November 17 2000).

U.S. *Executive Order 12906* (1994). Found at NSDI http://www.fgdc.gov/nsdi/nsdi.html (November 9 2000).

U.S. White House Office of the Press Secretary (2000). Press Release for May 1 2000, found at http://gps.laafb.af.mil/SA-PressReleases.htm (September 27 2000).

Vägverket (1999a). *NVDB – Kvalitetssäkring av dataleverantörernas verksamhet för dataförsörjning* version 1.0, found at the NVDB Homepage http://www.vv.se/nvdb (September 12 2000).

Vägverket (1999b). *NVDB – Specifikation av innehåll* version 3.0, found at the NVDB Homepage http://www.vv.se/nvdb (October 19 2000).

Vägverket (1999c). *NVDB – Specifikation av innehåll* version 3.0, bilaga 3 – *Kvalitetskrav på data*, found at the NVDB Homepage http://www.vv.se/nvdb (October 19 2000).

Magazines

Stanli (1997). Stanli Nyheter 3-1997.

Stanli (2000a). Stanli Nyheter 4-2000.

Internet

AutoDesc Inc. (2000). http://www.autodesk.com (October 31 2000).

Celigent (2000). http://www.celigent.com (January 15 2001).

CEN (2000a). http://www.cenorm.be/aboutcen/whatis.htm (October 24 2000).

CEN (2000b). Principles and methods for standardization http://www.cenorm.be/standardization/principles.htm (November 6 2000).

Dana P.H. (2000). Map projection overview http://www.colorado.edu/geography/gcraft/notes/mapproj/mapproj.html (November 1 2000).

Ertico (2000). Standardization http://www.ertico.com/what_its/standard.htm (December 6 2000).

ESRI (2000). Arc Info 8 & ArcSDE 8 http://www.esri-sweden.com/index.htm (October 11 2000).

Geographic Data Files (2000). http://www.ertico.com/links/gdf/gdf.htm (November 2 2000).

Geographic information standards (2000). http://www.gistandards.org.uk (November 6 2000).

Introduction to ISO (2000). http://www.iso.ch/infoe/intro.htm (September 11 2000).

ISO/TC 184 (2000). http://www.iso.ch/meme/TC184.html (November 7 2000).

ISO/TC 211 (2000). http://www.statkart.no/isotc211/ (October 24 2000).

ISO/TC176/SC2 (2000). http://www.bsi.org.uk/iso-tc176-sc2 (November 8 2000).

Lantmäteriet (2000a). http://www.lantmateriet.com Om Lantmäteriet – Geodesi och GPS – Kartprojektioner (September 19 2000).

Lantmäteriet (2000b). http://www.lantmateriet.com Produkter – Geografisk information – Geografiska Sverigedata – Dokumentation (December 5 2000).

NVDB (2000). http://www.vv.se/nvdb/index.asp (March 1 2001).

SIS (2000). http://www.sis.se (September 11 2000).

Standardization and the UML (2001). http://www2.awl.com/cseng/titles/0-201-89542-0/techniques/standards.htm (January 15 2001).

Stanli (2000b). http://www.stg.se/projekt/stanli/ (September 11 2000).

SWEPOS (2001). http://swepos.lmv.lm.se/index_gnss.htm (March 12 2001).
USNO NAVSTAR Global Positioning System (2000). http://tycho.usno.navy.mil/gpsinfo.html (September 27 2000).

Personal communication

Blomberg I. SIS Swedish Standards Institute (January 10 2001).
Kottman C. Vice President and Chief Scientist, Open GIS Consortium (January 10 2001).
Morel R. Quality Assurance Specialist, Navigation Technologies B.V. (January 30 2001).
Rydstedt B. Research & Development responsible, Lantmäteriverket (December 1 2000).

Appendix A – Example of the second quality measure

This is an example of how the second quality measure, as described in Chapter 11, might be structured. The example is given for a map covering entire Sweden with roads divided into two road classes, major roads and minor roads. The road network in the map has three attributes, speed limits, road names and road capacities. Furthermore, companies and municipalities have been added to the map as extra data.

An example of how to interpret this quality measure is given at the end of this appendix.

1 Road network

- 1.A Lineage: The road network has originally been produced by the Vägverket. The network has been transformed from RT 90 to WGS 84 using equations recommended by Lantmäteriverket and converted into ISAB's map data format.
- 1.B Topological consistency: 99% of all topological relations in the road network are correct.
- 1.C Temporal accuracy: The road network was last updated in October 1999. All roads in the map are valid.

1.1 Major Roads

- 1.1.A Data completeness: 95 % of the major roads are represented in the map.
- 1.1.B Attribute completeness: All relevant attributes have been captured.
- 1.1.C Positional accuracy: 95 % of the major roads have an absolute positional accuracy better then 10 meters.

1.1.1 Speed limits

- 1.1.1.A Value completeness: 98 % of the major roads in the map have speed limit.
- 1.1.1.B Domain consistency: 100% of the major roads with speed limits have values within the specified domain.
- 1.1.1.C Thematic accuracy: 95 % of the major roads with speed limits have correct values.

1.1.2 Road names

- 1.1.2.A Value completeness: 90 % of the major roads in the map have road names.
- 1.1.2.B Domain consistency: 100% of the major roads with road names have values within the specified domain.
- 1.1.2.C Thematic accuracy: 92 % of the major roads with road names have correct values.

1.1.3 Road capacities

1.1.3.A Value completeness: 90 % of the major roads in the map have road capacities.

1.1.3.B Domain consistency: 100% of the major roads with road capacities have values within the specified domain.

1.1.3.C Thematic accuracy: 97 % of the major roads with road capacities have correct values.

1.2 Minor Roads

1.2.A Data completeness: 90 % of the minor roads are represented in the map.

1.2.B Attribute completeness: All relevant attributes have been captured.

1.2.C Positional accuracy: 95 % of the minor roads have an absolute positional accuracy better then 20 meters.

1.2.1 Speed limits

1.2.1.A Value completeness: 95 % of the minor roads in the map have speed limit.

1.2.1.B Domain consistency: 100% of the minor roads with speed limits have values within the specified domain.

1.2.1.C Thematic accuracy: 85 % of the minor roads with speed limits have correct values.

1.2.2 Road names

1.2.2.A Value completeness: 60 % of the minor roads in the map have road names.

1.2.2.B Domain consistency: 100% of the minor roads with road names have values within the specified domain.

1.2.2.C Thematic accuracy: 90 % of the minor roads with road names have correct values.

1.2.3 Road capacities

1.2.3.A Value completeness: 20 % of the minor roads in the map have road capacities.

1.2.3.B Domain consistency: 100% of the minor roads with road capacities have values within the specified domain.

1.2.3.C Thematic accuracy: 95 % of the minor roads with road capacities have correct values.

2 Companies

2.A Lineage: The companies have originally been produced by the Yellow Pages and the CompanyRegister AB. The companies have been geocoded and then added to the map.

2.B Temporal accuracy: The company file from the Yellow Pages was last updated August 2000. The file from CompanyRegister AB was last updated in February 2001.

2.C Data completeness: 70 % of the companies are represented in the map.

2.D Attribute completeness: All relevant attributes have been captured.

2.E Positional accuracy: 95 % of the companies have an absolute positional accuracy better then 20 meters.

2.1 Name

2.1.A Value completeness: 100 % of the companies in the map have a name.

2.1.B Domain consistency: 100% of the companies with names have values within the specified domain.

2.1.C Thematic accuracy: 95 % of the companies with names have correct values.

2.2 Phone number

2.2.A Value completeness: 80 % of the companies in the map have a phone number.

2.2.B Domain consistency: 100% of the companies with names have values within the specified domain.

2.2.C Thematic accuracy: 95 % of the companies with names have correct values.

3 Municipalities

3.A Data completeness: 100 % of the municipalities are represented in the map.

3.B Attribute completeness: All relevant attributes have been captured.

3.C Positional accuracy: 95 % of the municipalities have a border with an absolute positional accuracy better then 200 meters.

3.1 Name

3.1.A Value completeness: 100 % of the municipalities in the map have a name.

3.1.B Domain consistency: 100% of the municipality with names have values within the specified domain.

3.1.C Thematic accuracy: 95 % of the municipality with names have correct values.

Interpretation of the quality measure

All levels are hierarchically linked, where every sublevel is a refinement of its superlevel. When interpreting this measure it is thus important to read not only the quality on one level. One example is the thematic accuracy for road capacity of major roads. 95 % of all major roads are represented in the map, 90 % of them have road capacities, and 97 % of these have the correct value. This yields a thematic accuracy of 0,95 * 0,90 * 0,97 = 0,83. Thus, 83 % of all major roads have been assigned a correct road capacity.