The Company Relevance of GIS Metadata
with special emphasis on
the Design of a Metadata System for GIS
at the VSN group

Drs. Clemens J.M. van Brunschot

Dissertation submitted in part fulfillment of requirements for the degree of Master of Science in Geographical Information Systems at the Manchester Metropolitan University

Submitted: February 1997

Abstract

GIS metadata is data about data, processes and environments relevant to the use of a Geographic Information System. It can be seen as the 'glue' connecting various datasets comprising a GIS, and it supports search and maintenance of GIS elements. GIS metadata is relevant to many companies, especially in those many cases where certain factors may present obstacles to rational GIS use. The more important the GIS 'house' is for establishing the strategy and/or operation of a company, the more this company needs a formal, automated GIS metadata management system (MMS). Although this does not hold very strongly for the present use of GIS in the central office of the VSN group (a major company producing collective personal mobility in the Netherlands), an MMS in this instance can be demonstrated to pay back its investment in less than three years.

In the VSN group case presented, MMS is intended to enhance effectiveness and efficiency and to prevent unacceptable risks, relating to the Atlas GIS operation. On this basis the study develops a full MMS design for Atlas GIS at the VSN group. In this process limited but rational use is made of the emerging European metadata standard, which may act as 'glue' connecting various GIS houses in Europe.

The study points to numerous advantages of GeoKey, a standard software package for MMS implementation. Yet, in the VSN group case, preference is given to a tailor-made MMS on the basis of the design developed.

Table of Contents

DISCLAIMER

The results presented in this thesis are based on my own research as supported by the Department of Environmental and Geographical Sciences of The Manchester Metropolitan University. All assistance received from other individuals and organizations has been acknowledged and full reference is made to all published and unpublished sources used.

This thesis has not been submitted previously for a degree at any Institution.

Signed: C.J.M. van Brunschot

Eindhoven, Netherlands (January 31, 1997)

SPECIAL ACKNOWLEDGMENTS

On September 1995 this study took a formal start. Right from the beginning my two coaches: Dr. D.I. Heywood (The Manchester Metropolitan University) and Prof. Dr. H.J. Scholten (Vrije Universiteit, Amsterdam), did a great job in helping me find a fine balance between what was called for by science, the subject, work and - last but not least - family life. For that matter I owe a lot to my patient wife Lenie, and also to my employer who sponsored the project.

Special thanks to the interviewees (L. de Jong, A. van Loevezijn, and P. Pronk) and to those who responded to my questionnaire in the Atlas GIS discussion group on the Internet: Reid Campbell, Marcelo Thome Caminha, Loren Muehlius, Beat Burri, Jacques Dupont, Matthew Rummel, and Thomas Ott. They all had to wait long to see the results of their contributions.

I am grateful to the Geodan company, who offered a temporary licence for using the GeoKey software, and to Bart Kusse, Eric Cats and Sjef Janssen for reading drafts of thesis sections.

Mathilde, thank you for the appointments you arranged.

1. Introduction

The past decades have witnessed an enormous growth of volumes of data available to anybody who is aware of its presence, and available to its reach. This holds for laymen as well as for professionals. In both categories people may be hindered by not knowing what the available data really represents and can be used for. Both laymen and professionals may find themselves confronted to data in such a way that the overview is easily lost. When this is the case he/she needs data about data (metadata or meta-information). This metadata imposes meaning and structure to what appeared to be blurred. It is a great asset for people who find it difficult to locate the data they need.

This may be quite interesting as a general topic. Our concern in this thesis will relate to professional use of data. More specifically to company use of data in and for geographic information systems (or: GIS). It is interesting to note (without proof given) that library, book shop, data bank, and Internet queries for the term ‘metadata’ normally yield more GIS-related literature than general information technology (or: IT) literature[1]. But that is not the reason for this study.

The reason for this study is the growing need in a Dutch company for internal GIS metadata. The author is employed in this company: the *VSN group*, a large company that provides collective transport to people. Section 1.1 will describe the VSN group and their GIS. For this moment it suffices to note that this study stems from a real and practical need for metadata. The study will aim at the design of a meta-information system for VSN’s GIS. But this effort will not be undertaken without a thorough study of GIS metadata and of its relevance for a company. The study area will be defined in Section 1.2.

1.1. The VSN group and GIS

The VSN group is the major company to offer both regional and urban public transport in the Netherlands. Most transport is done by bus and light rail, although small scale public transport is developing. Activities include taxi and private collective transport, as well as technical vehicle maintenance for third parties. The VSN group[2] employs (in full-time equivalents) 17,000 people. It operates 8,100 vehicles, of which 5,300 public transport buses. On an average day 1.3 million people make use of VSN’s transport services

The VSN group is a single company whose shares are held by the Dutch Minister of Traffic, Public Works and Waters. It has grown from the former association ESO where until 1989 individual bus

[1] To be fair: this may be the case partly because a more common noun for meta-database in database theory is: ‘data dictionary’.
[2] This (and other) information about VSN’s figures and strategic issues has been taken from a range of internal documents.

companies where members. Nowadays these individual companies are subsidiaries whose shares are owned by the VSN group.

Following developments receive a lot of attention at the VSN group:

Box 1-1: Some developments relevant to the VSN group

External developments:

- stagnating public transport volume
- declining transport market share (because of rising private car use)
- budget cuts in public transport subsidisation
- chance of privatisation or split-up
- commercialisation
- regionalisation
- public tendering and competition
- risk of being excluded in public tendering

Internal developments:

- trend toward small scale public transport
- more private collective transport
- mergers of subsidiaries
- new impetus for company wide strategy building
- subsidiaries eager to be (more) independent

VSN's corporate Mission Statement:

"Achieving business continuity by providing people's transportation."

While the VSN group keeps growing (mainly by acquiring private transport companies), its size and dominating position have led to an urge in national politics to enhance competition in the market for public transport. Recently one of the VSN subsidiaries lost part of their business to a U.S.A. based transport company (Vancom) in one of the pilot public tenders. More public tendering is likely to follow.

These strong external developments have led to strong internal developments where business and marketing strategy have received more coordinated attention than in the past, challenging present business and organisation views and culture.

Most planning and operating is done by the subsidiaries within strategic and tactical guidelines from the Utrecht based central organisation. The author is a staff employee in the fields of Information, Analysis and Research, in the central department of Marketing and Transport. Among his activities are market research, GIS, and scenario studies.

The GIS concerned is a stand-alone system consisting of Atlas GIS 3.01 (for Windows 3.1) on a Pentium desktop computer. Other software (a.o. IDRISI, FoxPro, MS Office, Lotus 1-2-3, SPSS) is also available. The first (DOS) version of the system was implemented in the middle of 1992.[3]

VSN's GIS contains area, line and point features, of company and market/environment entities. It is mainly used for visualisation for purposes of decision making, communication and negotiation, and for geographic analyses for the purpose of market research. Further description of VSN's GIS will now take place in the form of of a SWOT[4] analysis on VSN's GIS.

Box 1-2: SWOT Analysis of VSN's GIS (assuming Atlas GIS)

	Strengths	**Weaknesses**
Threats	*(from success to failure?)* • no system problems • good track record of analysis/maps • difficult to transplant success to subsidiaries' environments • high expectations • spatial data available from subsidiaries • GIS compatible across VSN group	*(great risk)* • dependency upon one main operator • poor documentation of applications • conversions often necessary • spaghetti line entities (therefore no link with transportation network) • no integration with IS/IT policy • do data integration offered • Atlas GIS tables not relational (although Oracle is becoming the standard and is used for the general Transportation Database)
Opportunities	*(improvement as an option)* • further development as a DSS tool	*(improvement may be imperitive)* • GIS as a toolbox (applications could be written) • distances not measured along road networks • no formal link with business strategy

It can be concluded that:

- Atlas GIS has drawbacks that might create future problems if line topology and relational databases become imperitive
- the present system is vulnerable from a management point of view
- improvements may be imperitive
- a documentation system of data and procedures seems to be an important improvement
- (partial) portability of this system to a possibly different future GIS would be a convenient feature

The need for a meta-information system at the VSN group will yet be clarified in this study, but for the moment it suffices to note that the meta-information system would be welcomed as a means of providing GIS data documentation in order to reduce company risk and to enhance efficiency.

---

[3] In two mini-conferences in April and May 1994 - with cooperation of Prof. Dr. H. Scholten - company attention was drawn to central GIS applications. Interest started growing, and subsidiaries started to make plans for GIS introduction. The three largest of them (NZH, ZWN group and Midnet) have also bought Atlas GIS in 1995. Some others are considering buying. All Atlas GIS software is operating under MS Windows 3.1. There has been a fair amount of coordination with the central organisation.

[4] Strengths - Weaknesses - Opportunities - Threats

1.2. Aims, Objectives, Contents

With the above in mind it is possible now to define the study area. The problem is reflected by the thesis title: The Company Relevance of GIS Metadata with special emphasis on the Design of a Metadata System for GIS at the VSN group.

Two questions will be raised:

1. **What company relevance does GIS metadata have?**
2. **How can a GIS metadata system be designed for use at the VSN group?**

These questions define two major aims in the thesis:

A. Demonstrate the company relevance of GIS metadata in general

B. Design a GIS metadata system for the VSN group

Ad A.

The company relevance of GIS metadata in general will be demonstrated through six objectives:

I. *Identify GIS needs of a general company* (Chapter 2)
Method: literature survey.

II. *Provide an overview of what GIS metadata is all about* (Chapter 2)
Method: literature survey

III. *Provide an overview of conceptual models for metadata, metadata systems, and related systems* (Chapters 2 and 3)
Method: Literature survey

IV. *Explore important metadata projects* (Chapter 4)
Methods: literature survey, scanning the World Wide Web

V. *Explore the most relevant metadata standards* (Chapter 4)
Methods: literature survey, scanning the World Wide Web, symposion participation

VI. *Identify GIS metadata needs of a general company* (Chapters 4 and 5)
Methods: literature survey, structured personal interviews, questionnaire in Internet discussion list

Ad B.

A design of a GIS metadata system for the VSN group will be achieved through four objectives:

I. *Identify GIS metadata needs of VSN group* (Chapter 5)
Methods: 'soft systems analysis', cost/benefit analysis, risk analysis

II. *Specify VSN's GIS metadata system (processes, conceptual model, functional requirements, interface, technical requirements)* (Chapter 5)

Methods: ISDM, ER modelling

III. *Evaluate current software suitability for the VSN system* (Chapter 5)

Methods: pilot with one software program, evaluation

IV. *Make a plan for further action* (Chapter 6)

As explained at the outset of this chapter, metadata may have general and/or professional relevance. In this thesis relevance is restricted to company relevance, which is taken as an instance of professional relevance, the not-included instance being scientific relevance. Company relevance is not entirely different from general relevance. With company relevance, however, there is assumed to be an emphasis on rationality: carrying out activities in such a way that goals are likely to be met.

'Company' might be thought of as being a commercial enterprise, but this study takes a wider view, including non-commercial companies as well, since in both categories modern emphasis is on effectiveness and efficiency.

Information will be considered relevant to an individual company when it helps bringing reality closer to the company's mission, goals or objectives (or vice versa). The nature of such expressions of intentionality may differ greatly across (types of) companies. In fact, they may be entirely missing in an individual company. But companies with no clear mission, goals nor objectives are NOT within the scope of this study. Scholten et al. (1994) touch upon a number of commercial applications of GIS and GIS metadata. One could add that all these examples of business needs in turn derive their relevance from missions, goals and/or objectives of organisations.

With GIS shall be meant: 'A system for capturing, storing, checking, integrating, manipulating, analysing and displaying data which are spatially referenced to the Earth. This is normally considered to involve a spatially referenced computer *database* and appropriate applications software' (Handling Geographic Information, 1987, p.132). Although the second part of this study will consider metadata for a particular GIS (Atlas GIS at the VSN group) the next three chapters will consider general GIS. It may well be that much of what will be elaborated upon is useful for managing information systems that share some functionality with GIS.[5]

Trivedi and Smith (1991) studied the role of metadata in managing Very Large Spatial Databases (VLSDB's). They defined metadata as follows: 'Metadata is the data required to describe, locate and control data in a database. Metadata also includes data required to describe and locate process and environment entities in the database. It is the information that allows data identification and selection

---

[5] Like transport planning software.

based on properties of data such as content, sources and quality' (Trivedi and Smith, 1991, p.7). From these formulations it becomes clear that the authors see location and control as crucial to providing access to data. Stressing the need of metadata providing direct access to the data (in the sense that the metadata system facilitates direct selection of data for GIS application), we adopt the definition. While data describes the real world, metadata describes data (and process and environment). In this study the noun meta-information will sometimes be used to indicate that an interpretation of metadata is available.

Trivedi and Smith are focusing on instances of a central database inside a company (or organisation). Our scope is somewhat broader: including metadata on out-company data and on multiple in-company databases. In the next three chapters some attention will be devoted to the question if there are major differences between metadata on in-company GIS data and those on out-company GIS data.

While data needs managing, metadata likewise needs managing. Managing of metadata takes place through the use of a metadata system, a central element being a meta-database. But it is necessary to include the meta-database administrator, all hardware, all documentation, all rules, and all users in the concept of a metadata system. Trivedi and Smith (1991, p.6) mention four elements of meta-database functionality:

- 'common formats and procedures for data definition',
- 'common procedures for collecting, updating and maintaining metadata',
- 'common procedures for access control to metadata',
- 'authority and control over metadata'.

Again we might add the need for metadata to provide access to the data (in the sense that the metadata system facilitates direct selection of data for GIS application).

While a database is an abstraction of the real world, a metadatabase is an abstraction of the database (Ibid., p.6). In a database we have entity types (e.g. 'Region' and 'Bus_Stop') with entity occurrences (e.g. 'Utrecht' and '12445' respectively) which are represented by data record occurrences. In the metadatabase, however, the entity types themselves become occurrences of a meta-entity (e.g. 'Spatial_Feature'). The meta-entity occurrences are represented by metadata record occurrences (after Wertz, 1986, p.55). In Sections 2.3 and 3.1 it will be made clear that too much focus on traditional data dictionary solutions should be avoided in this study. In this study we shall (depending on the context and the authors quoted) pick different names to refer to the same thing: 'metadata management system', 'metadata system', 'metadatabase', 'metadata information system', 'meta-information system' or 'metadatabase management system'.

Design refers both to a creative process and to the result of it. In this study a metadata system will be designed. As Petch (1993a, p.3-5) notes design is driven by problem, user and context, judged by general criteria, and strongly influenced by fashion and personal taste. Furthermore, design of computer systems may follow one from hundreds of developmental methods (Hobson, 1991, p.205). The impression is made that it is quite feasible to devote a full thesis to the design process itself. Such is not the intention in this study. Care will be taken not to pull attention too far away from metadata and metadata systems. Therefore literature used for matters as ISDM, ER modelling, and interface design will be limited.

The actual building of the very metadata system will not be part of this study. Rather the study will end by inspecting the capabilities of one popular instance of existing spatial metadata software in the framework of our requirements, and by drawing conclusions as to the whole study and the further development of the intended system.

The study is centering around two rather different questions: one theoretical and one practical of nature. The theoretical part of the study will provide a framework for the development of the metadata system, so that its relevance may transcend the needs of the VSN group.[6]

---

[6] GIS deployed by the VSN group is Atlas GIS. Sections 1.1, 4.4 and 5.2 address a number of Atlas GIS specific issues.

2. GIS and Metadata in a Company

This chapter is meant to offer a theoretical framework for the use of GIS metadata in companies. Section 2.1 is an introduction to company use of GIS. Section 2.2 introduces the GIS house. In Section 2.3 basic functions of (GIS) metadata will be discussed. And in Section 2.4 attention is devoted to data warehouse systems (which receive a lot of company attention nowadays) in order to find out if the use of these systems may offer experience relevant to our study of metadata.

2.1. Company and GIS

It is hard to say something sensible about GIS metadata without considering GIS first. It seems ridiculous to assess a company's need for GIS metadata (systems) without first answering the question why GIS has relevance in the first place. This section will deal with the relevance of GIS for business companies and other organisations which may use GIS to help meet their missions, goals and/or objectives. After a summary of what GIS can be and is used for in a company, the well-known concept of the GIS House will briefly be decribed in order to build the bridge between GIS data and GIS metadata.

In 1994 Grothe et al. published the results of a study that was aimed at finding evidence in the Netherlands to consolidate or falsify critical opinions about the company relevance of GIS. They drew a random sample from several categories[7] of companies that were suspected to find relevance in GIS. The sample was stratified for business sector and size. Questionnaires were mailed in April 1993, adressed to the Boards[8] of 2572 companies. Although literature was enclosed in order to raise interest, response was low: 16% for companies with 100 or more employees[9], 11% for companies with 20-99 employees (Grothe, 1994, p.67-70). All in all the sample cannot be regarded as providing a representative picture, but it is well worth to find out what management in these companies felt.

Out of 407 companies 36% said to be using spatial data (defined as: data that can be linked to a particular location on earth and could be pictured by use of coordinates). As much as 88% of the utility companies used spatial data. There was a positive correlation between use of spatial data and company size. Respondents from Marketing and Logistics departments were the main users of spatial data

---

[7] The categories being: utility companies, trade and retail companies, transport and communication sector, financial services, business services, companies with 1000 or more employees.
[8] This is not so say that it was the Boards who completed the questionnaires. In fact, the author was one of the respondents although he is not even aspiring to be a member of the Board.
[9] The response among the largest of these componies approached 100%

(Grothe et al., 1994, p.73-75).

Out of 144 spatial data using companies 53% said to be using GIS. Again this percentage was highest among utility companies. GIS use ranked highest (70%) among spatial data using companies counting 100-999 employees. Main application areas appeared to be, in descending order of frequency:

- Customer spotting and mapping
- Optimisation of distribution networks and centers
- Analysis of market shares/positions
- Distribution planning
- Mailing
- Geographic registration of real estate
- Locational planning
- Geographic registration of utilities
- Media planning

Almost ¾ of the 77 GIS using companies mentioned marketing/logistics related applications (Grothe et al., 1994, p.83-88).

Following table shows the position of GIS within the 'strategic raster' conceived by McFarlan and McKenney:

Box 2-1: Strategic position of GIS in a sample of Dutch companies (Grothe et al., 1994, p.97,110)

Strategic position	Percentage of 75 companies
Operational support	28%
Vital (current strategic support of structured activities wth no evident strategic strategic role in the future)	1%
Intermediate (now an operational tool, likely to become a strategic planning tool)	49%
Strategic planning tool, expected to persist in the future	22%

From this and more material presented by Grothe et al. (1994) one may draw the (conservative) conclusion that in many Dutch companies GIS is certainly considered to be relevant, especially as an operational tool. It is deemed to become a strategic tool as well.

Is company relevance of GIS restricted to the Netherlands, or is this an internationally established fact? Grothe et al. (1994, p.21-25) present interesting findings from international market research:

- GIS worldwide is expected to account for 6 Billion investment dollars in 1996
- Worldwide GIS investment growth rates as high as 32% per annum have been observed

- European growth is stronger than US growth (where current GIS penetration is higher)
- Within Europe the United Kingdom and the Netherlands are regarded the most mature markets, where familiarity and use of GIS technology are well developed.

Dataquest conducted a survey among European GIS users during the second and third quarters of 1995 to better understand the issues of the GIS end use environment. The questionnaire was distributed among the readers of the monthly European magazine Mapping Awareness. Gartzen and Hale (1995) published initial results. There were 924 respondents, 43% of whom were in local or central government, 18% from utilities and telecommunications, 17% from service sectors, 11% from education and research.[10]

Some of the results bear importance to our subject (Ibid., p.2,5,6):

- The more specific the respondents are in terms of mapping standards used in their organisation, the more they consent to the statement ‘My organisation’s progress with GIS has been excellent’. Gartzen and Hale conclude that “standards breed success”.
- European GIS users appear much less satisfied with their GIS systems and vendors than users in North America who were previously polled by Dataquest.
- Satisfaction with data access and data sharing fall short of respondents’ expectations, but this holds for all of the 14 product features and vendor services.
- From these and other findings the authors conclude (among other conclusions) that users should not try to build a GIS without also building a data maintenance program and data standards.

It follows that there are significant differences between the Netherlands, Europe and the world regarding acceptance and other market aspects of GIS. But the material presented does *not* in itself demonstrate that GIS is any more (or any less) important to any of these regions. However, the mature state of the Dutch GIS market has been attributed to the fact that in the Netherlands space is quite scarce and therefore calls for spatial planning (Grothe et al., 1994, p. 106). Taking everything above into consideration we dare expect that GIS is indeed relevant to many companies in many regions of the world.[11]

2.2. The GIS House

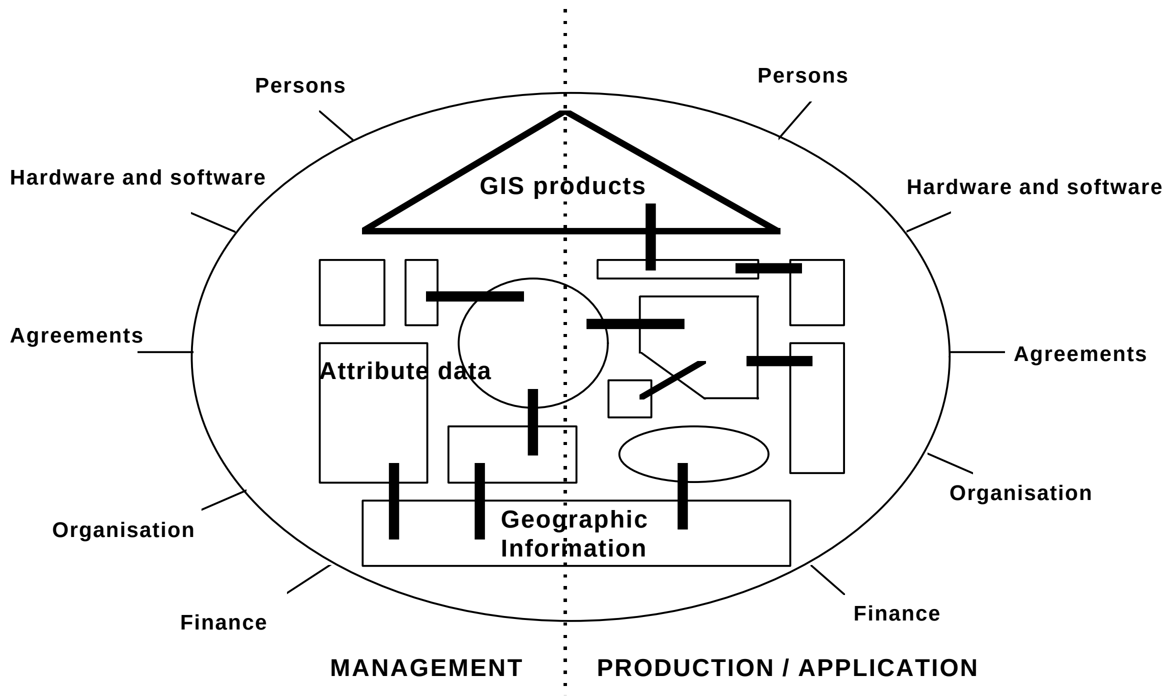
According to Scholten et al. (1994, p.203-204) the structure of GIS data can be seen as a house. See

---

[10] Note that this sample’s composition differs from the one in Grothe et al. (1994), who did not include the governmental sector.
[11] In GIS in Business (1995) many recent applications of GIS in business settings can be found.

Box 2-2. The roof is the set of display and other output possibilities: the end products of the GIS data processes. In contrast to the presentation roof we have the geographic database as the basis, the foundation of the GIS house: the geographical part of the data collections. This data represents geographical features, their topography and topology. Five aspects of the geographical data are important in the GIS house framework:

Box 2-2: The GIS House (after Scholten et al., 1994, p.204)



1. geographical data frequently is *not* specific to a particular organisation but to an area of interest that may be shared with other organisations,
2. this data is highly defined and standardised,
3. this data has been measured with a certain degree of accuracy,
4. this data is kept up to date
5. gathering and management of this data is expensive.

These aspects make it worthwile considering cooperation: within an organisation as well as between organisations sharing an area of interest.

From Scholten et al. (1994, p.204): ‘Between the output of the GIS and the geographic core dataset are the alphanumeric databases. Alphanumeric data and geographical data are integrated and processed into information, in the form of tables, graphs and maps for instance, by all kinds of processes. The

databases are the building bricks of the house. The data processes such as the modelling of water-flows or the modelling of noise, however, are the mortar you need if you want to end up with the information requested.
‘The GIS house is a concept for information supply. Around the GIS house there are a number of other factors at work. After all, for implementing the processes hardware and software are needed. Agreements have to be reached and standards devised for the hardware and software, but also for the data. An organisation is needed which makes available personnel, space, facilities and finance. And last, but by no means least, there are the people themselves, who can set to work with an information system but who are also responsible in their management of it.’

The GIS house stresses the links between the various datasets. It focuses on the system aspects of GIS and the main structure aspects of GIS data. From this it is only a small step to GIS metadata. In fact, it was already referred to by Scholten et al. In Kuggeleijn and Padding (1995, p.7.39) we see metadata pictured as a vertical, segmented bar alongside the GIS house, its segments having multiple connections with geographical information, attribute data, and GIS products.

The GIS house concept is useful to any organisation using or contemplating use of GIS. But it is particularly useful[12] to organisations that are concerned about GIS fitting into the organisation’s policy (mission, goals, objectives). These organisations are comprised in what we have called ‘companies’.

2.3. The Function of GIS Metadata

The metadata definition in Section 1.2 mentioned the main functions of metadata: identification and selection of data based on properties of data such as content, sources and quality. Trivedi and Smith (1991, p.5) have more to say for that matter: ‘We discuss the role of metadata in data management, query processing and helping the user. (...) Metadata is the data about entities in a database. It is systematic and typically deductive information about the contents, use and organization of the data in the underlying database. It acts a a repository for data from the logical and physical levels of the database. Metadata plays an important role while managing information as a resource that is shared by many users at all levels in a enterprise. It can be retrieved, manipulated and displayed in various ways. (...) Typically metadata users would include[13]:

- the system,
- all end-users from the very naive to the quite experienced and

[12] Because of its close relationship with organisational and IT aspects.
[13] Enumeration added (CvB).

- the database administrator.’

It is important in our study to stress that metadata users show variability in their needs for metadata. The GIS system itself and the database administrator will need metadata that shows every detail of data about their data. The interface does, then, not need to be absolutely user-friendly. On the other extreme is the non-experienced GIS user with his/her need to know what can be done with the GIS, and how, and where the necessary data can be picked up. The user-friendliness of the interface becomes crucial then. The experienced GIS user may be somewhere near the middle of the scale. Some metadata systems may focus on the detailed, technical aspects of data while others may equip end users with a helicopter view. When both categories of data properties need to be queried, multiple metadata management systems may be called for. Alternatively, it is a single system that may answer as diverse queries as (Trivedi and Smith, 1991, p.7):

‘The basic metadata queries about data entities include:

1. What is the data?
2. Where is the data?
3. What are the data characteristics?
4. Where did this data come from?
5. What is the quality (accuracy, lineage, resolution) of this data?

‘The basic metadata queries about process entities include:

1. What can one do with this data?
2. What happened when?
3. What types of spatial and geometric operators are available?
4. How can data be imported (acquired)?
5. How can entities like various data layers and spatial objects, be created, deleted and updated?
6. How can the results of analysis be presented?

‘The basic metadata queries about environment entities include:

1. What does the primary and secondary storage profile look like?
2. What are the user profiles or who is who and who is doing what?
3. What are the characteristics of various physical devices?’

In our treatment of metadata and metadata management system we shall avoid focusing too much on systems that are primarily devoted to providing detailed technical information. This may very well limit our description of e.g. data dictionary and data directory systems, which ‘(...) have been the conventional mechanisms to store and manage metadata. Much work has been done in the area of *extending*[14] the ideas of what information can and should be handled by the data dictionary, how this can be done and how the information can be used’ (Ibid., p.5). The relationship between data dictionaries and metadata systems in general will be clarified in Section 3.1.

2.4. The Data Warehouse

At first glance this section may seem to be an odd intrusion. It will become clear, however, that there are good reasons to expect that the data warehouse experience may add to our understanding of GIS metadata systems.

According to Inmon (1992, p.29) a data warehouse is a subject oriented, integrated, nonvolatile, time variant collection of data in support of management's decisions. Subject orientation means that the system is organised around management significant subject areas or entities rather than around atomic data as they appear in operational applications. Integration is said to occur when data passes from the application oriented operational environment to the data warehouse, whereby consistency is enforced. Nonvolatility means that data maintenance is done outside of the data warehouse. The time variance of data warehouse data shows up in a relatively long time horizon, in the existence of time series, and in the inherent relationship between data and time. Among key issues in data warehouse design are granularity (the level(s) of aggregation and summarisation of atomic data), partitioning (the organisation of data into relevant physical units), and the way that extracts will be taken from the operational environment (Ibid., p.29-56).

Why can data warehouses teach us anything about spatial metadata information systems? Basically because of following reasons:

- like GIS, a data warehouse is usually intended to be a Decision Support System
- like GIS, a data warehouse typically receives data from a wide range of sources
- like most GIS users, data warehouse users may not be interested in the lowest level of technical data detail, but they need to know the relevance, quality and location of data
- like in GIS, lineage of data is of paramount importance because of the time variance of the data warehouse
- all this is to say that metadata is as much an issue in the data warehouse as it is with GIS and that we can learn from a short treatment of the metadata aspect of data warehouses.

---

[14] Italics added (CvB).

‘It is with metadata that the most effective use of the data warehouse is made. Metadata allows the end user/DSS analyst to navigate through the possibilities. Metadata sits above the warehouse and keeps track of what is where in the warehouse. Typically the things the metadata store tracks are:

- the structure of data as known to the programmer
- the structure of data as known to the DSS analyst
- the source data feeding the data warehouse
- the transformation of data as it passes into the data warehouse
- the data model
- the relation between the data model and the data warehouse and
- the history of extracts’ (Ibid., p.100).

Metadata also includes common procedures for accessing data in the data warehouse (Ibid., p.127).

The role of metadata in the data warehouse environment is very different from its role in operational environments. In the operational environment the IT professional is dominant. The IT professional is computer literate and knows his/her way around systems on the basis of background and training. He/she uses metadata on a casual basis, much like the other system documentation. However, the data warehouse serves the DSS analyst community, in which there usually is not a high degree of computer literacy but other kinds of specialisation. The DSS analyst needs as much help as possible in order to use the data warehouse environment effectively, and metadata serves this end quite well. In addition, metadata is the first thing the DSS analyst looks at in planning how to do informational/analytical processing. ‘Because of the difference in the communities served and because of the role that metadata plays in the day-to-day job function, metadata is much more important in the data warehouse environment than it ever was in the operational environment’ (Ibid., p.133-135).

‘But there are other reasons why data warehouse metadata is important. A second reason (...) is that of managing the mapping between the operational environment and the data warehouse environment. (...) Data undergoes a significant transformation as it passes from the operational environment to the data warehouse environment. Conversion, filtering, summarization, structural changes, etc., all occur. There is a need to keep careful track of the transformation, and the metadata in the data warehouse is the ideal place to keep track (...). The importance of keeping a careful record of the transformation is highlighted by the events that occur when a manager needs to trace data from the data warehouse back to its operational source (...). In this case the record of the transformations describes exactly how to get from the data warehouse to the operational source of data’ (Ibid., p.135-136).

‘There is yet another important reason for the careful management of metadata in the data warehouse environment. (...) data in a data warehouse exists for a lengthy timespan - from five to ten years. Over a five-to-ten-year time span it is absolutely normal for a data warehouse to change its structure (...). Keeping track of the changing structure of data over time then is a natural task for the metadata in the data warehouse. Contrast the notion that there will be many structures of data over time in the data warehouse environment with the metadata found in the operational environment. In the operational environment it is assumed that at any one moment in time, there is one and only one correct definition of the structure of data’ (Ibid., p.136-137).

Metadata takes on yet another dimension in the face of storing and managing external and/or unstructured data. This data is registered, accessed and controlled in the data warehouse environment. Typical contents of data warehouse metadata for external/unstructured data is:

- document ID
- data of entry into the warehouse
- description of the document
- source of the document
- date of source of the document
- classification of the document
- index words
- purge date
- physical location reference
- length of the document
- related references (Ibid., p.178-179).

In many cases it will not be possible or economical to store all external data in the data warehouse. Instead, an entry can be made in the metadata of the warehouse describing where the actual body of external data can be found, and the external data is stored elsewhere. (Ibid., p.180)

‘It is through the metadata that a manager determines much information about the external data. In many instances, the manager will look at the metadata without ever looking at the source document. Scanning metadata eliminates much work for the manager in ruling out documents that are not relevant or are out of date. Properly built and maintained metadata, then, is absolutely essential to the operation of the data warehouse insofar as external data is concerned’ (Ibid., p.179).

Let us close this section and chapter by pointing to a few aspects of data warehouse metadata that bear particular relevance for spatial metadata and associated management systems:

- The nature and use of metadata differ greatly between operational and DSS environments.
- Metadata may in many cases be accessed instead of the data itself.
- Metadata is essential in finding data in the warehouse and in tracing it back to its sources.
- Metadata (referencing data outside the system) may offer a serious alternative to necessarily including all relevant data inside the system.
- Metadata are necessary in the case of data structure changes within data time series.

3. Conceptual Models for Metadata Management

In this chapter, after introducing some basic concepts, several conceptual models for metadata and metadata systems will be presented. We shall make a journey along four basic models that have been used for the sake of metadata and of the systems to manage metadata. We shall see (in Chapter 4) that practical metadata systems borrow from these basic models. But before all that some concepts call for clarification.

First the idea of concept itself. Man perceives reality largely in terms of prior conceptualisations or categories. Perception depends on the concepts that are handed over through culture. And a significant part of the professional's overall environment is made up of the concepts held by other persons in his/her sociocultural milieu (Sjoberg and Nett, 1968, p.35-36). A concept may be viewed as a building block for a shared understanding of reality.

This is important to our study, since we shall often be dealing with *conceptual models*: representations of reality in terms of concepts. A conceptual model is among other things useful to create a basis and a framework for the development of information systems. Therefore an overview of conceptual models for GIS metada (systems) will not only greatly add to our understanding of GIS metadata, but also provide points of departure for the VSN GIS metadata system in-statu-nascendi. 'The conceptual model is a description of the actual data[15] and the natural relationships found in that data. It is generally accepted that a good conceptual model should be meaningful to both the database manager and the non-technical business person or end user. Therefore this type of model must be devoid of any reference to the technical details of a database management system' (Wertz, 1986, p.33).

Metadata has been defined in the Chapter 1 as the data required to *describe, locate and control* data in a database. Metadata includes data required to describe and locate process and environment entities in the database management system. It is the information that allows data identification and selection based on properties of data such as content, sources and quality.

This metadata must therefore be efficiently and effectively accessible for a range of procedures[16]. This calls for a database management system, since metadata is also data. This leads us to the concept of a metadata management system (or: MMS[17]): a DBMS devoted to metadata. Of course there is a close relationship between data and metadata. Metadata is data about data. The form of this relationship

---

[15] In these sentences one may read 'metadata' where 'data' is mentioned.
[16] These procedures have been categorised in the previous chapter.
[17] In this study we shall (depending on the context and the authors quoted) pick different names to refer to the same thing: 'metadata

depends on the models for both data and metadata. The same holds for the relationship between DBMS and MMS.

Three statements may clarify this:

- In some data models part of metadata is an integral part of the data itself. In this case we have a *self-describing database* (Trivedi and Smith, 1991, p.5). An important instance is found in the Object Oriented Database Management Systems (OODBMSs), which are deemed superior to Relational Database Management Systems (RDBMSs) by Trivedi and Smith (1991, p.10). Most metadata, however, is stored in separate deductive (meta)data layers.[18]
- Most DBMSs contain a data dictionary, which is a sort of MMS used by the DBMS itself for data definition and for maintaining data integrity (Reeve, 1993, p.37-39) (Trivedi and Smith, 1991, p.4). This pertains to GIS to the extent that GIS *is* a DBMS (it may in fact be argued that GIS is more). Data dictionaries may be accessible for metadata queries by the user, in which case they show much general MMS functionality.
- We have mentioned metadata being data about data. Strictly speaking this somewhat contradicts with our very metadata definition, since metadata includes data about *processes in* and *the environment of GIS and/or DBMS*. On the other hand the contradiction is rather limited in the sense that these processes and environmental factors contribute much to what data one has available and what can be done with it (and might therefore be seen as attributes of data).

In the field of metadata we are often speaking about (meta)information on a possibly large number of separate databases spread over a wide geographical area. For these databases (and the data contained) to be part of an MMS we come to the matter of standardisation. This holds a fortiori in case MMS users are potential users of the distributed data itself. Data calls for data standards, so that one can use the data in combination with other data within an information system. Metadata calls for both data standards and metadata standards: data standards being an attribute of entities in the MMS; metadata standards allowing data description within the conceptual metadata model. In Chapter 4 we shall meet metadata standards.

3.1. The Data Dictionary

In section 1.2 we referred to Wertz (1986) for a description of the relationship between entities and meta-entities. In fact Wertz' work is about data dictionaries. This having said, it might suggest that we assume data dictionaries to be the same as metadatabases. This is certainly not the case. Originally data

---

management system', 'metadata system', 'metadatabase', 'metadata information system', 'metadatabase management system'.

[18] The dichotomy mentioned (referring to the physical structure of the metadatabase) does not seem to be quite the same as the one of active versus passive data dictionaries (referring to the integration of procedures) mentioned by Wertz (1986, p.58).

dictionaries were created for the removal of field/variable definitions from the various modules of a 3rd generation computer program to bring it together in a central 'data dictionary' so that the various modules would not cause definition or temporal conflicts, and software maintenance would become more easy (after Newman, 1991, p.74).

Definitions of data dictionaries can however be quite broad. Wertz (1986, p.56): '(...) a data dictionary is also[19] a computerized database of metadata, or data about data, together with the software and procedures used to create and maintain the dictionary database.' One one hand this definition is broad enough to describe a general metadatabase concept. On the other hand it does not mention information retrieval functionality. The latter, however, seemed also to be the case with Trivedi and Smith's definition of metadata in section 1.2. Much the same as Trivedi and Smith, Wertz does seem to include the retrieval function in the concept of a data dictionary.

But *is* a data dictionary the same as a metadatabase? We shall argue that traditional data dictionaries constitute a subset of medatabase systems, and one that shows limitations[20] when it comes to GIS metadata.

Wertz suggests that the use of data dictionaries might solve several operating problems as well as data-related (or information-related) problems. They are summarised in Box 3-1.

Box 3-1: Problems tackled by data dictionaries (Wertz, 1986, p.6-14):

Operating problems	**Information-related problems**
• redundant data	• inconsistent data
• redundant processing	• inconsistent representations
• inconsistency	• inconsistent timing (mix-up of dates)
• difficulty of change (maintenance)	• lack of understanding
• lack of adequate documentation	• lack of good data
	• lack of organisation

This list includes elements that would indicate that data dictionaries are much more than just a subset of metadatabase systems. Wertz' work however focuses strongly on data dictionary use by technically involved people like system developers and programmers. Focus is much more on professional housekeeping of system elements than on retrieval of (meta)data by more or less naive end users. Certainly this is the case with the data dictionaries[21] that come with standard database management systems. Trivedi and Smith (1991, p.4) mention several extensions of data dictionaries that have been proposed to overcome the limitations of traditional products.

---

[19] Wertz (1986) offers multiple definitions of a data dictionary, which taken together still can be seen as 'broad'.
[20] However, the limitations are much more severe in the case of 'data directories' which are described by Wertz (1986, p.57) as: 'a machine-readable definition of a computerized database'.

In Fourth Generation Environments the extensions of data dictionaries '(...) do not simply record information about data. They provide a means of managing it. (...). The[22] data directory component of the system is usually a table within the system which stores information about the other tables in the system (...) and about the columns within each table. Typically, for each column, the data dictionary will store[23]:

- 'its name,
- its role (equivalent to the type in more conventional languages e.g. integer, character, string),
- prompts that are to be issued to the user when a value should be entered,
- default values for variables that should be displayed to the user when input is required (...),
- headings that are to be used when a value is displayed if the user does not specify otherwise,
- checking/validation that should be applied to a value when it is supplied by the user before it is accepted as input into the database itself,
- error and help messages related to the column.

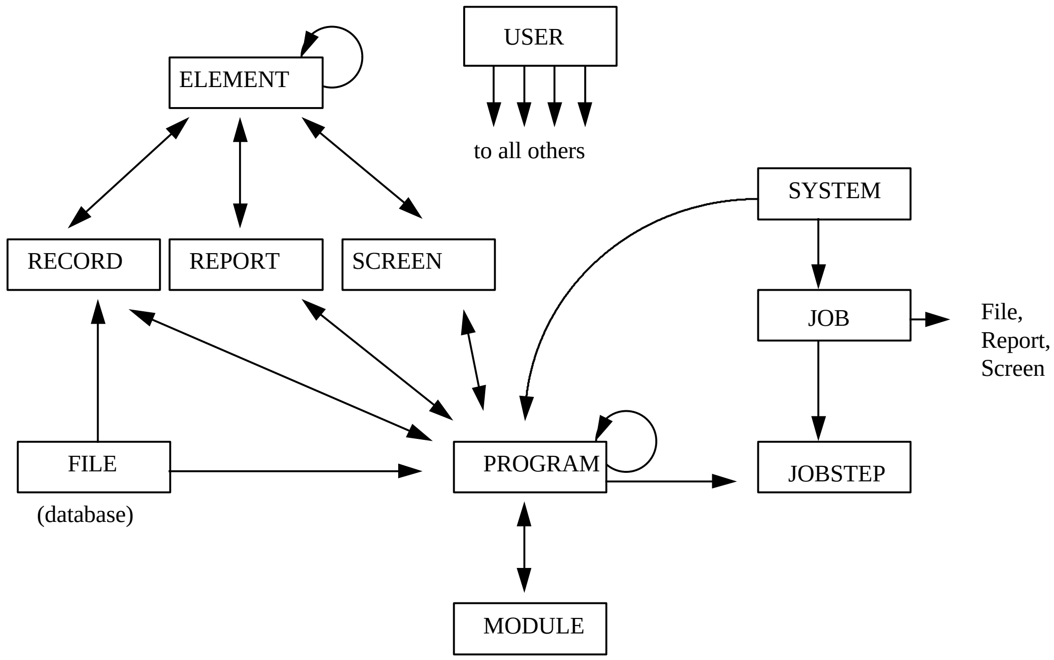
'In the future it is clear that the current data management systems will be enhanced by including some artificial intelligence facilities to make it easier for users who are not experienced database designers to construct usable systems' (Newman, 1991, p.76-77).

Now back to traditional data dictionaries. Wertz (1986, p.114-120) depicts an example of a conceptual model of a simple data dictionary aimed at accomplishing the housekeeping of systems, jobs, programs, and files or databases within a 3GL[24] using system development department. The conceptual model is drawn in the form of an ER (entity-relationship) diagram, where arrows are relationships between entities, and arrowheads represent the 'many' side of these relationships.

---

[21] Often mere data *directories*.
[22] Our quote jumps to an earlier text part (including the enumeration).
[23] Next to this enumeration Newman (1991, p.77) adds a few (rather technical) characterististics of more sophisticated systems.
[24] Namely COBOL.

Box 3-2: Conceptual Model of a Simple Dictionary (after Wertz, 1986, p.117)
(further simplified by omitting 'segment' and 'schema', where were considered analogous to 'record' and 'table/database', respectively)



Box 3-3: Short description of some of the meta-entities in Box 3-2 (after[25] Wertz, 1986, p.117-120).

META-ENTITY	DESCRIPTION	ATTRIBUTES
USER	An individual, responsible[26] for some of the dictionary content	Name Description
ELEMENT	Smallest unit of data (field). An ELEMENT may be a group of ELEMENTs	Name Definition Physical description Validation criteria
RECORD	Tuple (row) in a (table) FILE	Name Description
REPORT	File/Hardcopy listing of data (output)	Name Description
SCREEN	Screen listing of data (input/output)	Name Description
FILE	Collection of RECORDs	Name Description Physical characteristics
PROGRAM	Computer program (result of compilation). A PROGRAM may be a group of PROGRAMs	Name Definition/Description Size Revision data
MODULE	Standard piece of source code	Name Description Size Revision data
JOBSTEP	Unit of a PROGRAM's functionality	Name Definition/Description
JOB	Series of JOBSTEPs, used to create a FILE, a REPORT, a SCREEN	Name Description
SYSTEM	Family of JOBs, PROGRAMs, FILEs etc.	Name Description Revision data

[25] With a low degree of interpretation (CvB).
[26] Note that there are no USERs who only take (meta)data from the dictionary.

The above refers to dictionaries of general data. This study, however, is about GIS metadata, that is: spatial metadata. What is special about the spatial? 'The answer to the question (...) is: referencing. The two-dimensionality of maps, with the variety of co-ordinate systems that strain to cope with the near-sphericality of the land surface, and the three-dimensionality that attempts to cope with its topology, clearly make special demands upon systematic meta-description. But spatial data lend themselves to visualisation and so they can also add a dimension to our cognition and hence to our means for searching on metadata. Referencing requires attention to scale and to resolution. It also requires attention to spatial accuracy of measurement: to precision and to completeness of coverage. Therefore instrumentation matters, as does the inherent structure of the recorded data: raster or vector, with or without further topological structure. Areal identification schemes differ. In addition to various co-ordinate systems, there are postal/street addresses, postcodes, and a variety of areal (often administrative) boundaries defined by various agencies. These and the subject matter of the spatially-referenced data are subject to change over time. This is only part of the problem of temporal currency, but frequency of update may matter greatly'(Burnhill, 1991, p.30).

We conclude that:

- traditional data dictionaries fulfill at the best only part of the need for GIS metadata,
- 'state-of-the-art' data dictionaries may fulfill all GIS metadata needs. But in these cases the concept of 'data dictionary' is seldomly used. This can be seen clearly in Newman (1991), who offers a historical tour along data dictionary developments that lead to systems that deserve less and less the label 'data dictionary', as with Information Resource Dictionary Systems (IRDS) '(...) which are constructed around the concept of 'level pairs', where each pair comprises a 'type'and an 'instance' of that type. Types and instances (or: classes and objects in the more trendy 'object oriented' programming idiom) can occur at several levels. Something can be a type at one moment, then an instance of a higher level type a moment later' (Newman, 1991, p.78).[27]

Promising for the design of our metadata system seem to be:

- the use of an ER diagram to depict relationships between the building blocks of a metadata system
- the list of operating and information-oriented problems, as a cave-at when defining functional requirements
- the 'level pair' concept of (meta)data in the IRDS approach, which implicitly appeared already in section 1.2 (under the heading Metadata System)

[27] Cf. Wertz (1986, p.55), referred to in section 1.2 (heading: Metadata System).

3.2. A Meta-database for a VLSDB

Trivedi and Smith (1991) offer a conceptual framework for 'integrated[28] metadata management in Very Large Spatial Databases (VLSDBs)'. In their introduction they explain that VLSDBs have all the characteristics of generic spatial databases, but with spatial data in 'very large quantities' which means that the metadata can no longer be handled by a 'conventional system' because of volume, complexity and heterogeneity.They add that a GIS is a leading example of a spatial database (Ibid., p.1).

In Trivedi and Smith's conceptual framework metadata consists of three types of metainformation entities describing:

1. data in the database
2. processes in the DBMS
3. environment characteristics

ad 1. Metadata entities (Ibid., p.22-26)
The metadata[29] entities (that describe objects or entities which are units or aggregates of data) are stored in 3 modules: the data dictionary, data directory and data quality modules. The data dictionary module stores the metainformation about the data content and the conceptual schema. The data directory module stores the metainformation about the location of the data and how it can be accessed. The data quality module stores metainformation related to data quality like positional accuracy, attribute accuracy, consistency, completeness, and lineage. It seems somewhat confusing that information on such matters as preprocessing (the data stream up until data entry) and processing history (all changes of coverages or objects) are seen as being part of separate modules[30] (Ibid., p.26).

The metainformation of data content is quite dependent upon the underlying data model, as shown in Box 3-4.

Box 3-4: Data model dependent metadata entities (Trivedi and Smith, 1991, p.25)

Record based	Relational	Object-oriented	Deductive
element	relations	objects	facts
record	attributes	attributes	terms
group	domains	relationships	predicates
file			rules

[28] With *integrated* metadata management Trivedi and Smith (1991) seem to mean metadata being closely linked to the data model itself (self-describing databases). However, this integration is stronger in the case of OODBMS which Trivedi and Smith deem superior to RDBMS (Ibid., p.10) from a data management viewpoint. Also from a metadata management viewpoint the OODBMS offers advantages, since the very semantics of the object-oriented data model already contain part of the metadata. The practical example of an MMS that Trevedi and Smith present in their last chapter is nonetheless based on an Arc/Info application.
[29] Note that 'metadata' has a wider and a narrower meaning with Trivedi and Smith (1991).
[30] In stead of the data quality module.

The other metadata entities are independent of the underlying data model: databases, applications, reports, schema graphs and sub-graphs, and constraints.

Ad 2. Metaprocess entities (Ibid., p.21-26).
The metaprocess entities (that describe the processes available in the DBMS) fall into five basic categories:

- modelling primitives (generalisation, aggregation, grouping, classification)[31]
- operators (spatial, geometric, aggregate)
- transactions
- procedures and functions (system defined, user defined, object methods, constraint evaluators)
- programs

It would seem that preprocessing and processing histories of data are *not* metaprocess but metadata entities. Elsewhere (Ibid., p.8) Trivedi and Smith do incluse these histories in metaprocess entities. The metaprocess entities refer to the system and its functionality. The metaprocess entities are elements in a system documentation module. They are largely independent of the underlying data model.

Ad 3. Metaenvironment entities (Ibid., 24-26)
The metaenvironment entities are divided into:

- physical devices (input devices, output devices, terminals, storage[32] and memory)
- users (individuals, groups, organisations)
- software maintenance
- others

The metaenvironment entities are largely independent of the underlying data model.

The overview above pertained to metadata as such. Now attention is turned to the management of metadata (in the more general sense, including metaprocess and metaenvironment entities). We are still following the conceptual model by Trivedi and Smith (1991, p. 26-30).

All this metadata needs to be defined, stored and manipulated within the data management system, thus allowing for a kind of ‘self-describing’ database. This calls for organisation and structure at the ‘metaobject level’. At this level the authors mention four basic entities: fields, forms, catalogs and modules. The others are tabular and transient forms, fileboxes and orphan catalogs. ‘The primary purpose of these entities is to describe data about the entities in the underlying database and to assist in

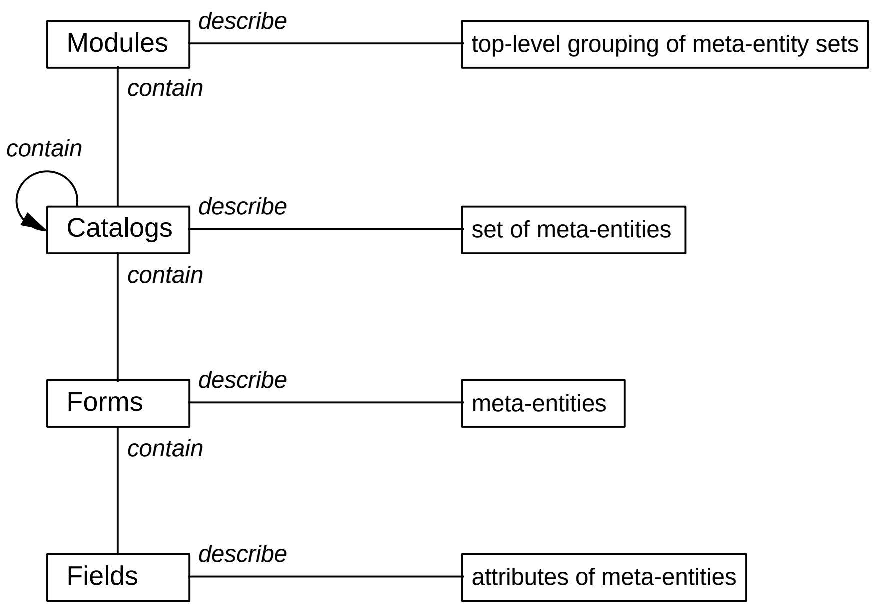
---

[31] This holds in the case of an OODBMS (Trivedi and Smith, 1991, p. 19-20).
[32] Including secondary storage profile, disk i/o statistics, information on available data compression procedures, and paging

manipulating, querying and making inferences on the metadata which is stored in them' (Ibid., p.26).

A field[33] is an object that describes an attribute of a meta-entity; it is an instance of a set of fields (fields class). A form describes a meta-entity itself; it contains a.o. fields, and is an instance of a set of forms (forms class). A catalog describes a set of meta-entities; it contains a.o. forms and catalogs (!). A module decribes a top level grouping of meta-entity sets; it contains catalogs. An illustration and a series of examples may clarify these concepts.

Box 3-5: The relation between fields, forms, catalogs, and modules
(Trivedi and Smith, 1991, p.28, with 'entity/concept' replaced by 'meta-entity')



mechanism characteristics like page distribution over multiple disks
[33] The rest of this section is based upon Trivedi and Smith (1991), but after a certain amount of interpretation, simplification and exemplification in order to reduce some of the ambiguities that were perceived by the author (CvB). By the same token the word entities is sometimes replaced by the word meta-entities. Who is interested in what Trivedi and Smith have to say about features and properties are advised to consult the original text. These items seemed not to be essential for the present explication.

Box 3-6: Examples of field, form, catalog, module (not by Trivedi and Smith)

Field: Coverage data model

- name: Cov_Model
- value type: Char string (6)
- value: Vector

Fields classes: {vector, raster}

Form: Coverage

- name: Coverage
- field 1 (Cov_Name): Municipal
- field 2 (Cov_Descr): Municipalities on 1-1-1995
- field 3 (Cov_Model): Vector
- etc.

Forms classes: {all coverages}

Catalog 1: Set of administrative coverages

- name: Admin_Cov
- entry 1: Municipal
- entry 2: Province
- entry 3: Country
- etc.

Catalog 2: A set of coverage sets

- name: Cov_Set
- entry 1: Admin_Cov
- entry 2: Stat_Cov
- entry 3: Ad_Hoc_Cov
- etc.

Module: GIS for Marketing Department

- name: GIS_Mrktng
- entry 1: Cov_Set
- entry 2: DB_Set
- entry 3: Report_Set
- entry 4: Query_Set
- entry 5: Device_Set
- etc.

About operations on metadata Trivedi and Smith (1991, p.30) write: ‘The following basic classes of operations are required for definition and manipulation of forms, fields and catalogs. If we assume a primitive for each of the operations then we have the following types of primitives[34]:

[34] From Trivedi and Smith (1991, p.30): ‘Create primitives are required for creating form class, form instance, and catalog. Delete primitives are required to delete form classes, form instances, fields and catalogs. Edit primitives for forms include adding a field, editing form name and form properties. Edit primitives for fields include editing field names, field properties, and field value. Edit primitives for catalogs include adding and deleting an entry, editing catalog name, editing catalog description.
‘We can identify three kinds of search: Feature based, Content based and Property based. Feature based search includes getting the field by name; the form by name, form class, keywords; and the catalog by name, catalog type and keywords. Content based search includes getting the field by value, getting forms by fields, getting catalog by entry-name and description. Property based search gets fields by field properties, forms by forms properties and catalogs by catalog properties. Lookup primitives allow the properties and

1. Create
2. Delete
3. Edit
4. Search and Lookup
5. Transfer
6. Miscellaneous

This concludes our overview of Trivedi and Smith's (1991) concepts of a metadata management system. The order they impose to metadata and metadata management may indeed help in defining a new metadata system, although we perceived some ambiguities regarding the relation between MMS and the object-oriented data model, regarding the relation between properties, features, attributes, and values, and regarding the nature of (pre-)processing history.

3.3. Lineage

Lanter (1990a, 1990b) wrote two NCGIA[35] reports concerning GIS[36] lineage statements. Following the proposals of a National Committee for Digital Cartographic Data Standards Lanter (1990a, p.2) defines lineage as: 'information describing source materials and transformations used to derive final digital cartographic data files'. It is proposed that GIS output is accompanied by a lineage report that shows source materials and processing history to facilitate interpretation and proper judgement of output.

A lineage report is produced by a lineage information system that details:
1. characteristics of cartographic resources
2. input/output relationships
3. transformations

Ad 1[37]. Characteristics of Cartographic Resources (Lanter 1990a, p.3/1990b, p.5,8,9)
Source layers are original inputs to a spatial database. Source documentation includes lineage information concerning:

- name
- feature types
- date
- responsible agency

features of fields, forms and catalogs to be queried. Transfer primitives allow transfer of information between metaentities and include two kinds of transfer commands - move and copy. Miscellaneous primitives include (1) Initialialization primitives for initializing forms and catalogs and (2) Cataloging primitives for filing forms and catalogs into other catalogs or modules.'
[35] National Center for Geographic Information and Analysis (USA).
[36] Lanter restricts his proposals to layer-based GIS.
[37] In the following explication we draw both from Lanters lineage problem analysis (1990a) and from his proposed solution (1990b).

- scale
- projection
- accuracy

This information is stored in a Source Description Frame.

Ad 2/3. Input/Output Relationships; Transformations

Product layers form the output of the GIS while intermediate layers bridge the gap between source layers and final products. This bridging of a gap is referred to as 'temporal topology': a network of temporal sequences, where layers are connected by data flows characterised by transformations. This concept was borrowed from Tomlin's cartographic model. These data flows represent the input/output relationships between the three types of layers. Input/output relationships are governed by commands that cause processes (transformations). Both product layers and intermediate layers have source layers and call for documentation on their 'parent' layers and the transformations used to derive them. The latter are documented in Command Frames. Source layers themselves call for documentation on their 'child' layers. Product layers call for documentation of their use in Product Frames which contain following attributes:

- Name
- Use
- Users
- Responsibility
- Date

Following table may clarify the relationship between layer types, lineage metadata and documentation frames:

Box 3-7: The relationship between layer types, metadata and documentation frames (after Lanter, 1990a/b)

Layer Type	Metadata	Documentation Frames
Source Layer	Cartographic Resources Child Layers	Source Documentation Frame
Intermediate Layer	Parent Layers Child Layers Transformations	 Command Frame
Product Layer	Parent Layers Transformations Product Attributes	 Command Frame Product Frame

According to Lanter (1990a, p.6-12) existing automatic systems[38] for tracking lineage in aspatial or spatial systems fall considerably short in providing means for a correct interpretation of GIS output.

Manual systems could do the job, but ‘one problem with the manual approach to lineage reporting is that it requires a conscientious effort on the part of all users to fill out forms at each step of a spatial application. A means for automatically creating lineage documentation and associating it with a derived GIS data product is clearly preferable.’ (Lanter 1990a, p.9).

The functional design that Lanter (1990b) proposes, assumes the availability of knowledge-based techniques for semantic networks. ‘Knowledge representations are data structures in which knowledge about a particular domain is stored. Such structures are important because they are useful for representing information about recurrent patterns extracted from our experience (...). A semantic network is a general framework in which knowledge is organized as a set of nodes connected by labeled links. Since relationships are explicitly represented as links in a graph, a semantic network traversal algorithm could make relevant associations by simply following links (...).’

The functional design resulted in a program called GeoLineus. It automatically stores lineage metadata information (partly after interrogating the user about sources) and facilitates queries in plain English to extract information on sources, transformations, and data products. The effects of updates can be traced (and regenerated) all the way between sources and products (Lanter 1990b, p.12-13).

Together with Surbey of the Southern California Edison GIAS Lab Lanter later used GeoLineus in an auditing study aimed at the nature and quality of data processing at the GIAS Lab, a production GIS facility equipped with ARC/INFO. The ‘Create from Log’ option of GeoLineus was used to extract data lineage metadata from the ARC/INFO Log files. While the GIAS analysts weren’t able to recall the relationships between 806 GIS layers and the original cartographic sources, GeoLineus was able to find the traces to 54 cartographic sources (along with most of the transformations applied). Because documentation of these sources was largely non-existent, the study resulted in the determination NOT to reuse the existing GIS database in order to avoid legal liabilities! (Lanter and Surbey, 1994)

The developments by Lanter seem important to our undertaking since data lineage is definitely an important metadata aspect that should be taken into consideration for the functional design of VSN’s GIS metadata system. ‘Analysis of a derived map’s lineage provides understanding of the geographic reality it represents. It also makes obvious when a GIS application is based on flawed logic. This is important. If the theory encoded in a GIS application is not clearly understood, then a correct interpretation of the resulting map is not possible. Recording this information along with documenting sources has critical relevance to decision makers using information derived by GIS applications’ (Lanter 1990a, p.5).

---

[38] Including history files, audit and log files, version control systems, map librarians, and polygon overlay attributes.

Briggs et al. (1993, p.92) summarize the benefits of GIS lineage:

- error detection
- management accountability
- external accountability
- quality reporting.

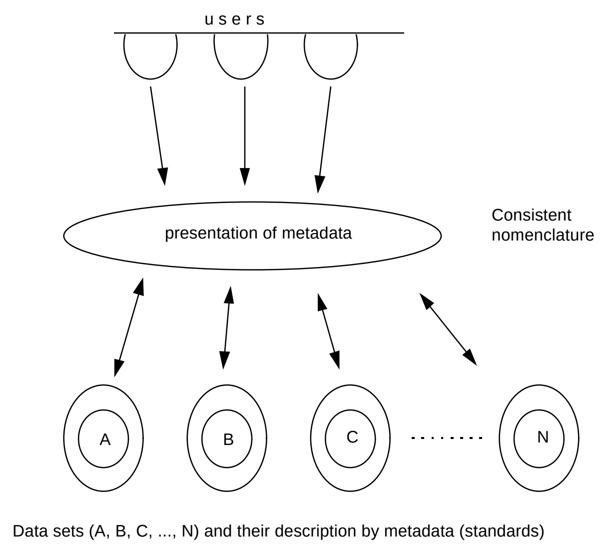
What exactly is the relationship between lineage and metadata? Briggs et al. (1993, p.102): 'The difference between a metadatabase and a lineage is that a metadatabase records where data can be found (producer/directory) and other information such as date of production and cost. It is a database of data sources. A lineage records the operations and methods used to produce information. A part of the lineage source information is included but not to the same level of detail as the metadatabase. This could be changed.' However, we find it difficult to see why metadata should focus primarily on data *sources* and why *processes* (as in lineage) would not be an important aspect of metadata. We find substance for our view with Trivedi and Smith (1991) in the previous section.

3.4. The Clearinghouse

In a general sense a clearinghouse is a place where goods or vehicles have to pass in order to make their transfer official. As a concept this idea has been applied to data. The place becomes then a virtual place or a computer network, where data can be obtained in a legal manner. This process is highly dependent on the availability of adequate metadata.

The concept of the clearinghouse is highly popular in the case of spatial data. In the next chapter we shall meet implementations of it. This section focuses on the more general aspects of the clearinghouse as a concept. The section will be rather short since the concept in itself is so simple. The real problems come with actual implementations, when it comes to maintaining standards for data and metadata. The diagram in Box 3-8 focuses on the metadata aspects of the clearinghouse.

Box 3-8: Metadata in the Clearinghouse (after RAVI, 1995a)



Users have access to a presentation of metadata that also serves as an interface to obtain the datasets of interest. The functionality of a well-developed clearinghouse may be to facilitate:

- storage of (or provide infrastructural links to) data sets
- storage of descriptions of data sets (including pathways)
- searching for metadata
- browsing metadata
- browsing data sets
- selecting data sets
- copying/conversion of data sets
- financial transactions derived from (meta)data use

The data sets may or may not conform to one or more standards. The same goes for the metadata. Essential is that the metadata plays an adequate role in describing relevant aspects of the data sets. This is not different from what we saw with other conceptual models in previous sections. Typical, however, is that the datasets may originate from distributed and totally different organisations who may have different ways of defining and storing data and metadata. Provisions may therefore be necessary to handle a variety of (meta)data standards. This may very well call for descriptions of metadata (say: meta-metadata). A practical solution is to make only part of metadata fields obligatory. Like the data, metadata may be geographically distributed.

Infrastructure is even more an issue in and around the clearinghouse than with other information systems. This has to do with the variety and the geographical spread of supplying and demanding parties. The Internet (especially the World-Wide-Web) is suggested as a convenient and well-distributed (but sometimes slow) infrastructural basis for a Wide Area Network, facilitating bottom-up growth of the clearinghouse. Because of the infrastructural emphasis the concept of a geo-/spatial data/information infrastructure is often used in relation to the clearinghouse.

The material in this section was adapted from: Berends 1995, RAVI 1995a, and van Cann 1995a.

4. Spatial Metadata Projects

In the previous chapter we saw four basic models for metadata and its management. In this chapter we shall review several applications of these models, although the traditional data dictionary model will be absent.

The sections will be organised geographically: first the United States (where the NSDI program and its Clearinghouse have proven to be exemplary in the field), then European developments (where coordination between national infrastructures is quite important), finally the Netherlands since this is where the author lives (and where the VSN group in particular is established) and finds a primary frame of reference.

The relationships between the projects described here and the basic models from the previous chapter are as follows:

- Sections 4.1 and 4.2, about the USA and European metadata standards, deal with essential infrastructure to describe lineage (Section 3.3), and implement a Clearinghouse (Section 3.4) in case the metadata should be usable outside a limited organisational setting. Section 4.3.4 is an example of a Clearinghouse (Section 3.3).
- Sections 4.3.1, 4.3.2, 4.3.3 and 4.4 follow some of the concepts in the model for metadata management described in Section 3.2.

4.1. The United States

Spatial data infrastructural developments in the United States are considered to be internationally relevant, mainly because of their exemplary function. Some elaboration on their history has, therefore, been included in Appendix A.1.

At its June 8, 1994, meeting the Federal Geographic Data Committee (FGDC) approved the “Content Standard for Digital Geospatial Metadata”. This standard specifies the information content of metadata for a set of digital geospatial data. The FGDC invites and encourages organisations to use the standard to document their geospatial data. The main reason to document their data is to maintain an organisation’s investment in its geospatial data. The standard specifies information that helps prospective users to determine what data exists, the fitness of this data for their applications, and the conditions for accessing the data. Metadata also aids the transfer of data to other users’ systems. The standard establishes the names (and groups) of metadata elements, their definitions, and information about the values that are to be provided for the metadata elements. Information about terms that are mandatory, mandatory under certain conditions, and optional (that is, at the discretion of the data provider) also is provided by the standard (adapted from FGDC, 1995d).

The Content Standard for Digital Geospatial Metadata[39] includes detailed definitions in the following groups of data elements (Ibid.):

1. ‘Identification Information - basic information about the data set. Examples include the title, the geographic area covered, currentness, and rules for acquiring or using the data.
2. ‘Data Quality Information - an assessment of the quality of the data set. Examples include the positional and attribute accuracy, completeness, consistency, the sources of information, and methods used to produce the data. Recommendations on information to be reported are in the Spatial Data Transfer Standard (...).
3. ‘Spatial Data Organization Information - the mechanism used to represent spatial information in the data set. Examples include the method used to represent spatial positions directly (such as raster or vector) and indirectly (such as street addresses or country codes) and the number of spatial objects in the data set.
4. ‘Spatial Reference Information - description of the reference frame for, and means of encoding, coordinates in the data set. Examples include the name of and parameters for map projections or grid coordinate systems, horizontal and vertical datums, and the coordinate system resolution.
5. ‘Entity and Attribute Information - information about the content of the data set, including the entity types and their attributes and the domains from which attribute values may be assigned. Examples include the names and definitions of features, attributes, and attribute values.
6. ‘Distribution Information - information about obtaining the data set. Examples include a contract for the distributor, available formats, information about how to obtain data sets online or on physical media (such as cartridge tape or CD-ROM), and fees for the data.
7. ‘Metadata Reference Information - information on the currentness of the metadata information and the responsible party. Examples include currentness and information about the organization that provided the metadata’.

The standard has sections that specify contact information for organisations, temporal information for time periods covered by the data set, and citation information. The standard does not specify how metadata is to be organised in a computer system or in data transfer. This decision was made because of the variety of means of organising data in a computer or in transfer, the differences between data providers to describe their data holdings, and the rapid evolution of means to provide information through the Internet (adapted from Ibid.). Peter Schweitzer of the United States Geographical Society (see Stein, 1995) has developed one of the computer programs used to check the syntax of metadata records using the FGDC Content Standard. Metadata records must use a particular specification for metadata encoding in order for that particular validation service to be helpful. The service is available through the Internet.

The FGDC metadata standard is considered to be rather complex. Yet data providers are encouraged to use the entire standard, including the optional elements (Schweitzer, 1995, Q6). To show the complexity of the standard one single syntax rule will be reproduced here (FGDC 1994, section 5):

Attribute =

Attribute_Label +
Attribute_Definition +
Attribute_Definition_Source +
1{Attribute_Domain_Values}n +
0{Attribute_Units_of_Measure}1 +
(Attribute_Measurement_Resolution) +
(1{Beginning_Date_of_Attribute_Values + 0{Ending_Date_of_Attribute_Values}1}n) +
(Attribute_Value_Accuracy_Information) +
(Attribute_Measurement_Frequency)

Complex[40] as it may be (or: seem), the FGDC metadata standard is considered a very useful one. It turned out e.g. to be perfectly adequate in describing all geospatial data held by the Staring Centrum, the Institute for Research of Rural Areas, in the Netherlands (Bisoen and Heerikhuisen, 1995b, p.8).

We adopt following conclusions from within the FGDC (Tosta, 1995, p.7.12): ‘The NSDI is an ambitious and challenging effort that requires far more interaction and coordination among individuals and organizations than has been demonstrated in the past. The FGDC can only begin to chart a course, it cannot accomplish the task. New ways of looking at the world, sharing information, identifying incentives, and building partnerships must be identified and encouraged. Discussions in the United States and other countries about national spatial data infrastructures are the likely precursor to the concept of a global spatial data infrastructure. These discussions must be encouraged’.

---

[39] The standard can be found on the Internet (FGDC 1994).
[40] Some explanation is appropriate: A = C + D means that meta-entity A is a combination of C and D; (B) means that B is optional; n{X}m means: between n and m (inclusive) occurrences of X.

4.2. Europe

The European Community is striving toward a European Geographic Information Infrastructure (EGII): a stable, European-wide set of agreed rules, standards and procedures for creating, collecting, exchanging and using geographic information (GI). The EGII would also ensure that European-wide base datasets are readily available and that metadata services exist so that data can be easily located by potential users (EGII Policy Document, 1995). For efficient operation of the EGII a set of standards is required that will allow data interchange, system interoperability, and metadata expression. European (CEN) and international (ISO) initiatives are well on their way to promote standards in the field of geographic (meta)information. In 1995 draft GI standards were published by the CEN/TC 287 technical committee[41].

At the moment that this is written (October 1996) all standards are available in draft version from the Internet. All standards are aimed to be be finalised in 1997, after which formal procedures will still have to be followed in order to create an official European standard[42] (RAVI, 1995b, p.18). In August 1996 (GIS users etc., 1996), looking forward to the Second "Geodata for All in Europe" EUROGI workshop, attention was drawn to the open letter of EUROGI chairman M.J.D. Brand (Brand, 1996) in which serious questions were raised concerning the tension between USA/European initiatives (focusing on data standard-isation) and ISO initiatives (focusing on GIS standardisation and Open GIS development).

Appendix A.1 contains additional details on the line of thinking behind the European standards. The set of standards prepared by CEN/TC 287 is divided into four categories, further divided into documents (RAVI, 1995b, p.18-21; van Oogen, 1995a):

Overview

An introductory overview of all components of the standard, including:

- *Reference Model* (an overview of the family of standards and its application areas).
- *Overview* (what is geographic information (GI), what elements need standardising?).
- *Definitions* and *Dictionary*.

---

[41] The paths for CEN/TC 287 were initially prepared by European organisations for official and commercial cartography, CERCO and MEGRIN. Heavily involved in all EGII activities is EUROGI: the European Umbrella Organisation for Geographical Information (EUROGI, 1994/95).

[42] CEN/TC 287 should not be confused with CEN/TC 278, the technical committee that - within the EC DRIVE program - developed the GDF draft standard for geographic information related to road transport and traffic telematics (EDRM 1994). The GDF standard will still have to be adapted to the CEN/TC 287 standard.

Data Description

How to describe geographic information using conceptual schemas? Including:

- *Techniques* (adoption of the formal language EXPRESS[43]).
- *Application Schemas* (rules for development and use of data models).
- *Geometrical Subschemas* (definitions and representation of primitives for modelling geometrical and topological data models).
- *Quality* (GI quality parameters and their integration into the application schemas by means of subschemas).
- *Meta-information* (structure of general descriptions of data sets, including availability, classification models used, and generic data quality aspects).
- *Data Transfer* (transfer schemas and implementation mechanisms for GI transfer).

Referencing

- *Metrical Position* (how to describe locational reference systems by coordinates?).
- *Indirect Positioning* (how to describe locational reference systems in case positioning is done indirectly?).
- *Time* (how to describe the temporal dimensions of geographic data?).

Processing

Operations on geographic data and its metadata:

- *Query and Update* (development or adoption of formal languages including appropriate spatial operators; identification systems for querying and updating geographic (meta)data).

Although 'meta-information' was used above in a rather narrow sense (in relation to the description of data *files*), it can be seen that the CEN/TC 287 data standard will also entail an elaborate metadata standard. The standard is created to develop datasets in a standardised manner and in the same process provide systematic description of these datasets, that is: to create metadata[44].

What use is the European (meta)data standard to an average company using geographic information somewhere in Europe?

- Companies involved in international projects will benefit from the compatibility of national data sets. This holds for companies cooperating with other GI using companies in other countries, as well as

---

[43] EXPRESS is an ISO standard.

[44] In fact, the CEN/TC 287 is also created to coordinate the development of derived data standards (RAVI, 1995b, p.10), in which case its meta-character is once more emphasised.

for companies with GI-aided international marketing or logistic operations.

- Many companies will be able to make good use of geographic core datasets developed in the GI2000 project. This project offers great added value in the current situation where commercial map-making has as a rule concentrated on specialised fields where profit possibilities exist, i.e. road maps or thematic maps of urban areas. As might be expected, such map-making enterprises have seldomly taken the responsibility to secure nation-wide data coverage, including areas where little profit potential exists. Part of the base data sets is intended to be placed in the public domain (GI2000, 1994/95).
- Cost of geographic data collection - accounting for about 60% of the cost of GIS analysis (Scholten et al., 1994, p.206) - may significantly be reduced because of partial elimination of redundancy and incompatibility.

In Chapter 5 an attempt will be made to make partial use of CEN/TC 287[45] (and/or the US FGDC Content Standard) when defining the conceptual model for the GIS metadata system at the VSN group.

4.3. The Netherlands

In Section 2.1 we saw that the Netherlands represent a relatively mature GIS market. Its scarce resources of physical space have necessitated efficient physical planning and have created a great need for (often detailed) spatial data. The need for spatial metadata probably has led to a need for spatial metadata. It is a fact that there are several Dutch metadata projects that have drawn a great deal of both national and international attention from the professional community. Of these we treat four projects[46].

4.3.1. The RIVM Experience

The RIVM is the National Institute for Public Health and Environment, located in the town of Bilthoven. The RIVM is an important advisor to the Ministery of Public Health and Environment. It has seven sectors with an average of 5 laboratories each. One of these laboratories is the important Laboratory for Soil and Groundwater Research (LBG). The LBG has 20 employees, 15 of whom are involved in analytical GIS operations on a regular basis (van Cann, 1995b, p.22).

At the LBG ARC/INFO was introduced in 1989. The use of the system has shown considerable growth since then. In fact, ARC/INFO GIS plays an essential role in the LBG spatial information provision. GIS instruments were largely project oriented until 1993, when problems became apparent related to

---

[45] E.g. with reference to the Draft standard for metadata (CEN 1996a).
[46] To date the other ones are NexpRI (the Dutch Centre for Expertise on Spatial Information), the Province of Gelderland, and CBS (the Statistical Bureau of the Netherlands).

control, availability, quality, integration, and capacity of spatial data provision. A number of developments were started that were intended to lead to a totally new spatial data infrastructure at LBG, an important component of which is the newly developed infrastructure for spatial data management: GeoBase (Thewessen et al., 1995b, p.3).

The data management infrastructure facilitates control and availability of all existing centrally registered data at the LBG. Metadata plays an important role in this concept: it is the key to the data. Next to systems, data and metadata[47], the data management infrastructure consists of standards for direct access to data for the sake of model applications. The same goes for procedures for maintenance, access, and security (Ibid., p.11).

GeoBase has been developed largely in line with the data warehouse concept[48] (Ibid., p.16). It facilitates tracing, evaluation (through metadata), browsing, viewing, selection, and export (to different formats) of data. GeoBase acts as a ‘window’ to all important LBG data, wherever they reside.

The metadata serves as a kind of glue, providing integration of the other data types (Ibid., p.15). Centrally controlled it supplies information on the GoeBase data. Each GeoBase file has its own associated metadata, including information on a.o. quality, scale, definitions and interpretations of items, etc. Metadata also has important functions in authorisation of data access, history of file derivation, and showing how up-to-date files are.

The metadata is part of the central LBG database. Because the metadata structure is different from the spatial data structure the logical part where the metadata resides is also called the meta-info database. This meta-info database can be divided into:

- Standard tables, containing a.o. definitions of keywords, objects, users and projections. These tables function as lists of supported application areas and projects.
- File descriptive tables, containing the metadata of files, which is divided into a number of groups of tables[49] (Ibid., p.24).

The GeoBase proficient user interface offers following functionality (Ibid., p.33): searching, browsing, exporting a file, querying file metadata, importing a file, searching for a map. According to source files are divided into measurement data, basic data, remote sensing data, derived data, project data, and output data. According to topology files are divided into physical location data, physical attribute data,

---

[47] LBG’s information structure is closely related to other infrastructure components at RIVM, e.g. the meta-information system AIDA.
[48] See section 2.4.

various spatial data, and non-spatial data (Ibid., p.19-22).

The metadata is divided into 12 screens, describing relevant aspects of files[50] (Ibid., p.35/36):

1. General 1: general metadata, like title, summary, data type (see previous paragraph), contact persons, supplier etc.
2. General 2: general metadata, like quality, temporal extent, restriction to user types, metadata status, publications wherefore data used.
3. Columns: a description of all columns in the data file, i.e. name, description, data type, length in bytes, decimals, key y/n.
4. Spatial aspects: spatial extent and coverage.
5. Map: list of maps associated with the file.
6. Geographic aspects: map projections used for digitising file data.
7. Derivation: the file's relations to source files, transformations, macro's and scripts.
8. Project: the project in which the file was created.
9. Relations: the file's relations to other GeoBase files. E.g. a spatial attribute file always has a relation to a spatial location file.
10. Extra info: a list of (unformatted text) files possibly providing additional information on the file. These files can be browsed.
11. Versions: list of different versions of the file.
12. Remarks: users can read and write remarks about the file.

For data exchange with the GeoBase databases the transfer format GeoTalk was created, supporting a large number of file formats, like (for tables) dBase, Ingres and INFO, (for vector data) ARC, Spans etc., and (for grid data) ARC, Idrisi etc. (Ibid., p.47)

There are important similarities as well as dissimilarities between GeoBase and the system that the VSN group needs. The greatest similarity is that both systems contain company internal data and metadata, unlike systems we met in the previous sections. The greatest differences are that GIS at the VSN group is (basically) single user/standalone, and that there is considerably less volume, diversity and complexity of data and procedures. On the other hand GeoBase metadata are indeed data about data, while the VSN system will probably also contain meta-process and meta-environment data, much like what we saw with Trivedi and Smith (1991) in Section 3.2.

When developing the functionality of VSN's metadata system we can go back to the GeoBase report (Thewessen et al., 1995b) and try to make use of the fundamental work that was done there, e.g. in creating the ER diagrams.

---

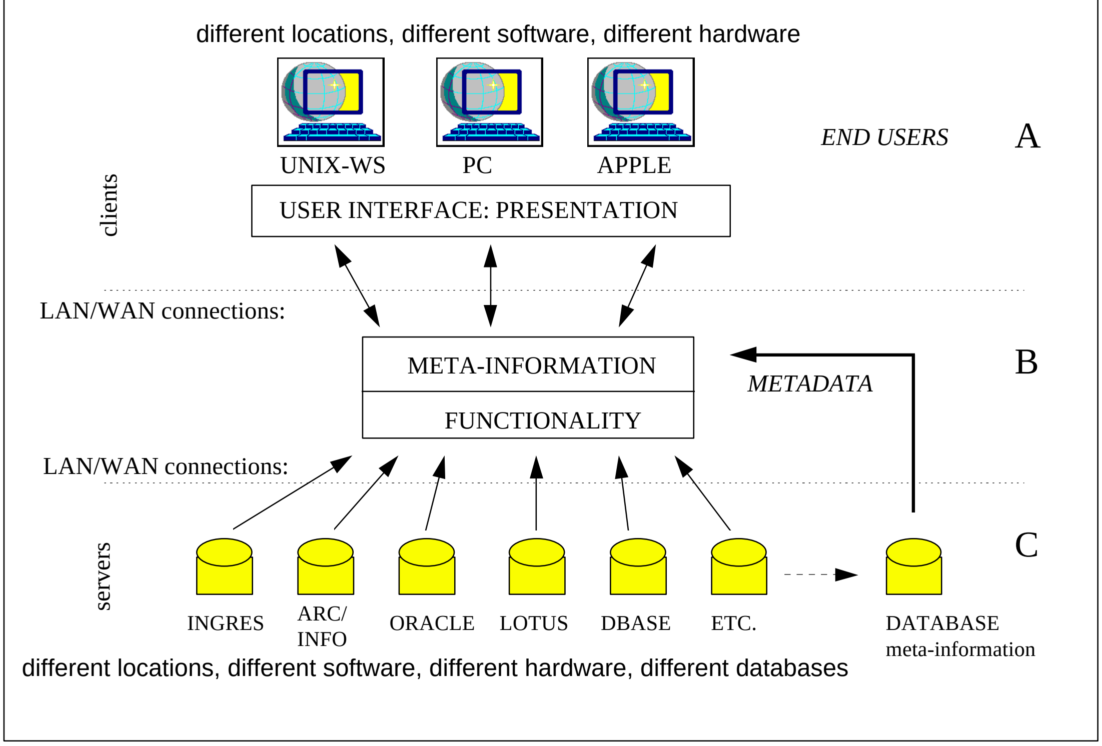
[49] Since not all of the metadata applies to all of the files.
[50] Not-applicable metadata fields are 'dimmed' in the metadata screen.

4.3.2. Metadata in the Geo-Key System

The Dutch Ministry for Transport, Public Works and Water has witnessed (and is still witnessing) a steady growth in the use of GIS, especially within the executive body of Rijkswaterstaat. Along with this development there is a shift from single-user GIS applications to organisation-wide projects, and also growing volumes of spatial data. All this data needs to be stored and managed. There has been multiple storage of identical data. It has even been found that identical data was bought more than once from the same source. These findings have prepared the way toward an organisation-wide infrastructure for distributed data management (Kuggeleijn and Padding, 1995, p.7.37-7.38). This system will extend to all over 100 Ministry locations in the country (van Cann, 1995a, p.8).

A choice has been made for a so-called Open GIS Environment (OGE), that is independent of the variation in hard and software, since the conclusion has been drawn that a single platform is not feasible. Within an application there usually is a separation between three layers or tiers: data storage (databases), data processing (functionality), and presentation (user-interface). In the three-tiered OGE concept there is a central role for data processing, for which reason it is supported by a metadata system, since metadata in this concept is the key to available data. This metadata is independent (but, of course, not ignorant) of location, software, hardware, and databases. Following diagram clarifies the idea.

Box 4-1: The Three-Tiered Concept (after Kuggeleijn and Padding, 1995, p.7.45)



Advantages of the OGE are to be found in efficiencies of application development and data administration.The metadata system allows (authorised) viewing of the actual data over a network connection. The metadatabase is updated in real time as data is changed. (Kuggeleijn and Padding, 1995).

Metadata at Rijkswaterstaat is divided into 4 main groups of variables (Ibid., p.7.40):

- general
- access
- content
- specific blocks, dependent on dataset type.

Rijkswaterstaat dataset types include (Ibid.):

- geographical
- remote sensing
- attribute
- map composition
- software program

Openness to the outside world will in time become reality, probably through the Internet. Note that the system includes Wide Area Network (WAN) connections to similar organisations. This outside openness is step 2. In step 1 (finished by the end of 1996) the system will be implemented all over the Ministry. There will also be a step 3: connecting the metadata system to the future National Clearinghouse that will be built[51] (van Cann 1995a, Kuggeleijn 1995).

The metadata system is an implementation of the software package GeoKey, that was developed by Geodan, a leading Amsterdam based GIS house. GeoKey runs under DOS/Windows and Unix. Each dataset is (in the Rijkswaterstaat implementation) described by 40-50 fields, only 12 of which are mandatory (in order to lower thresholds). GeoKey is allegedly flexible and user-friendly. Geodan claims that the data selection module[52] (GeoKey Select) can be used within 30 minutes of training[53]. GeoKey Select has a graphical user-interface showing a pictorial impression of the (spatial) data[54] and

---

[51] See section 4.3.4.
[52] Other GeoKey modules are: Edit (for customising/filling/maintenance of the metadatabase), View (for viewing slected spatial data), Browse (for browsing selected non-spatial data), Transport (for transfer between different software types), and Plot (for plotting/printing).
[53] The author can confirm this statement from his own observation while working with the demo diskette.
[54] GeoKey View or a scripted link to a GIS allows more detailed viewing of the actual data.

facilitating data selections on dataset type, keyword, values of variables, elements from a hierarchical set of themes, area type element, and geographical location (Geodan 1995a/b, van Cann 1995a).

The main value of the GeoKey implementation at Rijkswaterstaat to our study is probably that it has given us a clue as to the possible usefulness of GeoKey as a tool for the future metadata system of the VSN group. GeoKey may be a candidate for a pilot at the VSN group. It will, in that case, be relevant to see if GeoKey would be able to contain and access all relevant VSN GIS metadata (including GIS *process* and *environment* entities) and if it would show sufficient integration with Atlas GIS data structures[55].

4.3.3. Metasys Prototype at the Staring Centrum

In Wageningen (the Netherlands) we have the Staring Centrum (SC-DLO), an institute for research on rural areas, as one of the twelve institutes of the official Agricultural Study Service (DLO). It creates and manages several national geographical files relating to soil, soil use, landscape, and ecology. The files are used inside and outside the Staring Centrum for both national and regional studies (Bregt et al., 1995, p.7.81; Bisoen and Heerikhuisen, 1995e, p.2).

The Staring Centrum has assessed that their data is far from fully open for use. This holds for both data and metadata. For that reason the SC-DLO is preparing for the development of a metadata information system. In 1995 two students have caried out a study to investigate the opportunities and risks of a metadata information system. Their valued work has resulted in a proposal and a prototype (Metasys) for the creation and management of such a system. In this section (and Appendix 2) we draw from their report in five volumes (Bisoen and Heerikhuisen, 1995a/b/c/d/e).

The search for relevant metadata entities and attributes started from the metadata project at RIVM (that we pictured in Section 4.3.1.). The metadata content of RIVM's GeoBase was compared to the U.S. Content Standard for Digital Geospatial Metadata (CSDGM, see Section 4.1.). The same exercise was carried out for the existing (simple) GISHELP metadata information system at SC-DLO. It was shown that RIVM's GeoBase was incompatible with SC-DLO's hard-/software platforms. Following these analyses all relevant metadata was defined according to the CSDGM[56] (Bisoen and Heerikhuisen, 1995b). See Appendix A.1 for more detailed information on the work done on Metasys.

The latest news we heard from the Staring Centrum (February 1996) was that there have been major

---

[55] GeoKey is sold for 50, 100 or an infinite number of users (Geodan 1995a), while the VSN group would probably count much less users.
[56] In this process minor additions relative to the CSDGM were made.

changes of plans:

- the construction of the SC-DLO metadata information system is to follow the broader national clearinghouse pilot projects where SC-DLO is involved in (see Section 4.3.4).
- the draft European CEN standard (see Section 4.2) is to be used rather than the U.S. CSDGM

The lists of functional requirements, entities and attributes, and screens (only partially contained in this section) may prove useful for the developments of VSN's metadata system. Furthermore we have seen that it makes sense (for the SC-DLO) to try and learn from other projects (as we have been doing ourselves in this chapter).

4.3.4. National Clearinghouse for Geo-Information

In the Netherlands there is, as a follow-up to earlier projects (RAVI 1992), a project aimed at arriving at a National Geo-information Infrastructure (NGII): the combination of policies, datasets, agreements, standards, technology (hardware, software, and electronic communication) and knowledge that serves a geo-information user to accomplish his/her own task (RAVI, 1995c, p.7). The coordination of geo-information since 1990 is a task for the Secretary of State for Housing, Physical Planning and Environment. In this task he is supported by RAVI, the Advisory Council for Information on Real Estate, which is concerned with policies, geo-datasets, geo-information technology, (meta)data standards (especially the European standard) and knowledge infrastructure (Ibid., p.3,7).

To facilitate access to geo-datasets metadata in a geo-information clearinghouse is deemed crucial. RAVI plays an active role in creating a Clearinghouse for National Geo-information, but essentially as a follower of geo-information users' initiatives. In August 1995 a project plan for a national clearinghouse was submitted to the Minister of Internal Affairs in the framework of the National Action Programme for Electronic Highways (NAP-aanvraag 1995).

Following aims of the national geo-information clearinghouse are envisaged (RAVI, 1995c, p.12):

- better and faster availability of geo-information for policies of official bodies and companies,
- prevention of redundant work, by ascertaining what datasets are already available,
- improvement of service provision to the citizen, by integrated supply of information,
- forming a channel for purchase and delivery of geo-datasets as half-products or information services,
- quality improvements, as a result of mutation reports from a wide field of users.

The Dutch geo-information clearinghouse will be easily accessible to a large field of users. Therefore the World Wide Web has been proposed as a platform[57]. The preliminary choice for the Internet has led to the insight that metadata should be offered in two quite different ways (Berends, 1995, p.49):

- well-structured/high-quality (by the European metadata standard), for most professional use,
- less-structured/quality-not-guaranteed, to keep it interesting for many other types of users.

According to RAVI (1996, p.37) priority is in providing professional users with free metadata. Later on, paid metadata and other users may be added.

The clearinghouse project is characterized by a high degree of cooperation from the field of professional land information users, and by the use of pilots. Eleven pilots have been proposed by now (February 1996). Not all of them have been defined yet. All pilots where metadata is defined follow the European draft CEN/TC 287 metadata standard. All projects end somewhere in 1996, so that the National Geo-information Clearinghouse can become effective in 1997. We now mention three of these pilots, namely the projects where the Staring Centrum (from Section 4.3.3) is involved in. The pilot descriptions were taken from: Deelnemende Pilots in het Nationale Clearinghouse Geo-informatie, 1996.

A. The Idefix pilot project is aimed a gaining experience with new insights, concepts, methods and techniques in relation to the realization of a national clearinghouse. There is an emphasis on three basic clearinghouse functions (for environmental geodata):

1. searching and browsing metadata,
2. transferring, browsing and selecting data, using a geodata warehouse on a WWW server,
3. same as 2., but as a company-internal system (without World Wide Web).

Participating parties are: RIVM, Survey Department, Staring Centrum, Province of Gelderland, and RAVI. There is support from commercial companies and other organisations: CAP-Volmac, Geodan, IDE, and NexpRI (Thewessen et al., 1995a, RAVI, 1995a).

B. The Staring Centrum conducts their own pilot project “Tuning of geo-information demand and supply”, aimed at enlarging knowledge of and insight into the decision making process in the geo-information market. Focus is on the roles and forms of clearinghouse metadata exchange in data purchase processes.

C. The Staring Centrum also participates in a clearinghouse pilot project shared with the other institutions of the Cooperating Geographical Data supplying Institutions (SAG): Rijks Geologische Dienst, RIVM Laboratorium voor Bodem- en Grondwateronderzoek, and TNO Grondwater en Geo-

[57] The URL of the Idefix pilot project is http://idefix.geodan.nl/

energie. The pilot aims at encompassing the SAG WWW-catalogue of digital geodata in the national geo-information clearinghouse.

Following conclusions can be drawn from this section and may prove useful in defining the VSN metadata system:

- The European draft CEN/TC 287 metadata standard already plays in important role in arriving at geographical metadata in the Netherlands.
- There is a riding train heading for a national geo-information clearinghouse, where the VSN group somewhere in time may connect to for gathering and/or providing geographical (meta-)data.

4.4. Atlas GIS Metadata

Since the second part of this thesis will concern metadata for a specific GIS (Atlas GIS), it makes good sense to have a look at Atlas GIS metadata projects. We have conducted an exploratory study into the wishes and experiences of other Atlas GIS users concerning metadata. The results will be discussed after a short introduction of this specific GIS.

Atlas GIS is a powerful desktop (vector) mapping package originally produced by Strategic Mapping Inc. (SMI) in Santa Clara (USA). It is available for DOS and for DOS/Windows 3.x, for single-user and network implementations. The system is used all over the world[58]. Functionality is aimed primarily at presentation and geographical analysis. Procedures are accessible through a menu and subject to additions through a scripting language, using (alternatively) Visual Basic or C. Among Atlas GIS (for Windows) file types are (Atlas GIS Reference Manual, 1994):

- Geo Files (quadruple) with geographical features (regions, lines, points) organised in layers and columns, one key column being the _ID field.
- Attribute Tables (dBASE-compatible), linked to the _ID field in a Geo File.
- Column Settings Files, containing Attribute Table structure and linking information.
- Point Tables (separate files for points, like a point layer in a Geo File).
- Project Files: definitions of maps, including file links. One project file may include the references for several of each of file types mentioned above.

In 1996 products of Strategic Mapping, including Atlas GIS, were sold to Claritas inc., a subsidiary of the Dutch VNU publisher. Claritas retained the data products and sold the GIS products to ESRI, the

---

[58] Atlas GIS has to deal with heavy competition especially from Mapinfo. Both of these products using vector data models, Atlas GIS's strengths are in its vaste capabilities for thematic mapping, while Mapinfo is better suited as a mapping interface to various

producer of ArcInfo and ArcView. This all caused uncertainty for the Atlas GIS users, who received signals that ESRI might be interested in Atlas GIS primarily as a means of transferring an extra customer base to use of its desktop ArcView. However, ESRI keeps repeating that Atlas GIS deserves a separate right of being, though recently (Dangermond, 1996) it was made known that there will be an interface to use Atlas GIS data and functionality within ArcView.

In the course of this study we asked Atlas GIS users about their use of both Atlas GIS and metadata. First interviews were personal, later we posted a questionnaire in the AGIS-L discussion list on the Internet. Appendix 3 contains a detailed report of this research and its resulting data. The research was somewhat limited since the VSN group's wishes will anyway be leading in defining our own metadata system in chapter 5.

Of these Atlas GIS users, four were in the USA, two in the Netherlands, one in Canada, one in Germany, and one in Switzerland. Following general conclusions may be drawn from the study:

- Unlike the VSN group, each of these nine Atlas GIS sites shows more than one user.
- On most of these sites there is little or no metadata made explicit.
- Most users in these sites are aware of their metadata needs.
- Available metadata in few of these sites is organised in a formal system. Paper or a simple file prove useful.
- One user mentions a project where Atlas GIS metadata were subjected to the USA metadata standard, discussed in Section 4.1.
- Two users mention a need for an application that automatically scans Atlas GIS files to collect metadata.

From Appendix A.2 we derive the following list of metadata entities/attributes suggested by the users involved:

---

database types.

Box 4-2: Metadata suggested by nine Atlas GIS users

Category	Metadata suggested
GENERAL / DATA	Availability of data (internal, external) Authorisation Quality of data
GEOGRAPHICAL DATA (Geographical Files)	Map layers Date of latest change Version code Description Number of elements Type (region, line, point) Geographic extent Scale Source Suitability for specific applications Remarks
ATTRIBUTE DATA (Attribute Tables)	Who created data? Who maintains data? Source Collection method Fields: name, type, description Size Location
MAPS (Project Files)	Level of detail Status Description
PROCESSES	Purpose of data How does it work?

We are confident that this information will prove useful in specifying the VSN metadata system.

5. The VSN group Company Case

Chapters 2-4 were concerned with present GIS metadata literature and projects. This effort was primarily undertaken as a preparation for the development of a Metadata Management System (MMS) in a specific company. In this chapter an MMS will be designed for the VSN group, which will be restricted to the Utrecht based corporate headquarters (and specifically the department of Marketing and Transport). After a recapitulation of what is directly relevant from earlier chapters - the theoretical part of this study - we shall dig deeper into: relevance of GIS metadata for a company in general (5.1), the problem situation in the VSN group case (5.2), aims and objectives (5.3), the business case (5.4), specification of the desired system (5.5), and a pilot with existing software (5.6).

5.1. Company Relevance of GIS Metadata

What then is the company relevance of GIS metadata in general? While trying to answer this question, it seems to be reasonable to assume that we are talking about a company that is already a GIS user. In Section 2.1 we demonstrated that GIS use in fact pays for many companies. It is good to remember that with ‘company relevance’ we assumed (in Section 1.2) an emphasis on rationality: carrying out activities in such a way that goals are likely to be met. ‘Company’ was not restricted to commercial operations, but to organisations with a mission, goals and/or objectives, striving to effectivenesss and efficiency.

Let us take a small step back to what was in Section 2.2 said about the GIS House concept. We saw that the GIS house serves as a concept showing the interrelatedness or structure of GIS data. Or, somewhat broader: all aspects of GIS application. Metadata serves as the glue[59] connecting elements of the GIS house. For our treatise on the company relevance of GIS metadata we like to expand this view.

Reviewing the previous chapters one may be struck by the many gaps that need to be bridged before rational use of GIS inside a company is possible. These gaps or discrepancies have to do with modern professional life being complex. Gaps may exist within and between categories of company entities. Examples of company entities (or fields) that call for bridging gaps, are:

- Environment
- Company
- Mission, goals, objectives

[59] Concept borrowed from RIVM. See section 4.3.1.

- GIS software/application
- GIS output
- DBMS software/application
- Users
- Departments
- Geographical data
- Attribute data
- Hardware
- Locations
- Time

Some quite real examples of such gaps may clarify this point:

1. Within a company there may not be a clear understanding as to the availability of geographical data on a neighbouring country.
2. Two related companies may have identical data needs without knowing.
3. The company's Board may be unaware of GIS' relevance for decision support.
4. One GIS user might not know what another GIS user has been doing.
5. A GIS user might not know the quality status of attribute data, since these were prepared in a different department.
6. A GIS user may not recall how he used to produce map X from available data.
7. Data may be digitised by one department while another department within the same company has bought it last year.
8. Since data structures changed between two dates, it may seem impossible to construct a time series of geographical data for a spatial diffusion analysis.
9. A GIS user may be unaware of data being available in a different location of the same company.
10. A company may have three locations all using different operating systems, preventing data interchange.

It will be demonstrated that GIS meta-information (or, for that matter, an MMS) could alleviate or eradicate many of the problems arising from such gaps.

Following (triangular) table demonstrates some of the applications and related advantages of meta-information (systems) in bridging gaps within and between six categories of company entities:

Box 5-1: Example applications/advantages of GIS meta-information (systems) in bridging gaps, per combination of company entities (GMI stands for GIS meta-information)

	Environment	**Company**	**Department**	**GIS user**	**GIS application/output**	**Data**
Environment	GMI can show what spatial data remains relevant after company diversification	GMI as a basis for external accountability of a company producing spatial studies	GMI facilitates external audit of departmental GIS procedures	GMI shows a GIS user what data is available about the company's environment	GMI may show which scripts are available for spatial analysis on a geographical market segment	GMI may show differences in data structures across geographical market segments
Company		GMI aids in investigating opportunities for combined data purchase across companies	GMI facilitates in-company GIS audit	GMI may show a GIS user what information needs the company has for decision support	GMI may learn a company's Board what relevance its GIS has for decision support	GMI shows what data is available within the company at large
Department			GMI helps prevent data purchase duplication across departments	GMI may show which maps user X produced for department Y	GMI may show if GIS application X was used on request by department Y	GMI may turn data into information, usable by a specific department
GIS user				GMI facilitates data exchange between GIS users	GMI shows which map was produced by which GIS user (for efficient re-production)	GMI identifies the GIS user responsible for specific data maintenance
GIS application/ output					GMI facilitates new GIS applications based on existing ones	GMI shows data suitability for specific GIS use
Data						GMI shows inconsistencies between datasets

This table will - we trust - bring the message home that:

GIS meta-information can be quite valuable to a company,
for bridging some of the gaps otherwise preventing rational GIS use.

For the sake of completeness, four related questions call for an answer:

1. Does this conclusion hold for all GIS using companies?
2. Is a meta-information *system* needed to make the advantages manifest?
3. What can and what cannot be expected of a GIS meta-information system?
4. How important are standards for spatial metadata?

In answering these questions use shall once again be made of insights gathered in the course of this study. We shall favour clarity above verbosity.

Ad 1. GIS meta-information relevant to *all* GIS using companies?

Every GIS using company needs GIS meta-information. Without it, GIS use is impossible. However, the necessary degree of meta-information explicitation depends upon the company's complexity and the risks it is willing to accept. With one GIS user, all meta-information may fit perfectly well into the user's mind. But this user may choose a different employer or be involved in a fatal accident. With two users, risks are lower, but coordination may already call for explicitation of some of the meta-information. With many users, risks are very low, but meta-information must absolutely be explicitated. Complexity may be augmented by diversity of GIS use across users, rather than by their sheer number.

Ad 2. Is a meta-information *system* required?

One might argue that where (meta)information exists, a (meta)information system automatically exists. For how could it be possible to deal with data and information otherwise that in a systematic way? Although the argument holds, it may act somewhat confusing. As with information, information systems do differ greatly in their degree of formality. Some companies' meta-information needs may be satisfied with a rather personal or informal system (like hardcopy output from a word processor), while other companies absolutely need a formal system. And systems may still be on a scale from 0 to 100% automation. Roughly spoken, the formal and highly automated meta-information system suits a company best if that company relies on GIS as an essential element of its essential operation or primary process. The needed degree of meta-information system automation and formality is probably lower in a company where GIS is used for decision support 'only'.

Ad 3. What can(not) be expected of GIS meta-information system?

It can be noted that not all of the 10 example 'gaps' were covered by applications/advantages mentioned in Box 5-1. Especially one may have missed issues related to changing data structures or to diversity in locations and platforms. The reason for this is that even in formal GIS meta-information systems ranges are discernable: from a most humble and dedicated meta-information system on the one extreme, ranging to the Open GIS Environment and Data Warehouse on the other extreme. So it is clear that a formal meta-information system in itself is not a sufficient condition for bridging some of the mentioned gaps faced by modern companies. Complexity removed *by* a meta-information system may very well be proportionate to complexity *of* a meta-information system.

Ad 4. How important are spatial metadata *standards*?

A company may design metadata in splendid isolation, or, on the other extreme, completely rely on official metadata standards for description of spatial data. To explain the extremes again, we picture two quite different companies. Company A has - perhaps for strategic reasons - developed both their GIS software and all of their data themselves, and never supplies data to the outside world. Company B is - maybe, again, for strategic reasons - using standard GIS software and purchased data only, and it is a supplier of (enriched) spatial data. Because purchased data as well as standard software will at some point in time probably be subjected to metadata standards (irrespective of the view the company itself takes in this matter), company B would act much more rational in conforming to official metadata standards than would company A. In fact, company B might in future have to pay a price for not readily adopting the metadata standard to begin with. (This is, however, not to say that company B would have to limit their metadata to what was covered by the metadata standard. Some extension might be necessary to suit company-specific needs.)

While metadata (and meta-information) acts as glue between the elements of the GIS house, metadata standards act as glue between different GIS houses.

5.2. The Problem Situation

In Section 1.1 we found that the VSN group is a large company with subsidiaries active in collective transport of persons. The VSN group shows a strong trend away from large scale public transport, and into small scale transit operations. Turbulence is not only internally, but also (primordially) in market and environment.

GIS is used as a visualization tool (for decision support, communication and negotiations) and as an analytical tool (for market and transportation research). Atlas GIS: desktop (powerful PC), Windows 3.1, stand-alone, with entities being points, lines and regions. Line entities are ‘spaghetti’ lines: topology is missing, although operating transit networks is core business for the VSN group. GIS at the VSN corporate headquarters is essentially used by one operator. The main user can be seen as a (market) information specialist and a researcher (Grothe et al., 1994, p.45). In the past one additional person received a minimum of training on an earlier version of Atlas GIS; he lacks experience.

Weaknesses of VSN GIS that might be alleviated by metadata (system) include:

- dependency upon one operator,
- poor documentation,
- many conversions and complex data processes needed.

In line with Section 5.1 expectation is that a metadata system might bridge some of the gaps otherwise preventing rational GIS use. Rational GIS use includes, then, aspects like:

- effectiveness (meeting demands from higher organisational levels)
- efficiency (doing it all within limited time frames)
- preventing unacceptable risks.

Actual developments that may prove relevant for GIS and GIS metadata in the VSN group are:

- GIS more and more used in VSN subsidiaries,
- relational Oracle databases gaining wider acceptance within the company,
- use of Oracle CASE gaining wider acceptance as an ISDM tool,
- trend to integration of data (standards, data warehouse),
- from the external developments should be mentioned the uncertainties for Atlas GIS users that may be associated with the acquisition of Atlas GIS by ESRI, the manufacturer of ArcInfo and ArcView. Though ESRI does a lot to convince Atlas GIS users, a number of users in the AGIS discussion group on the Internet expect the software as such to be discontinued somewhere in the (near?) future.

Among possible future developments that may prove relevant for VSN GIS (metadata) are:

- stronger call for line topology in GIS data structures,
- more than one operator,
- closer relationship between GIS on the one hand, and IS/IT policy and formal Decision Support on the other hand,
- changing organisational structure within VSN group,
- European geospatial metadata standard finding wide acceptance.

For an analysis of the problem situation we shall borrow from the original Soft Systems Analysis (SSA), developed by Peter Checkland and his colleagues at Lancaster University, as explicitated in

Reeve and Cornelius (1993, Section 5.5.3). Use of SSA will be followed by more formal ISDM methods, because in the specification phase the VSN-internal ideas of what is needed will turn out to be rather well developed, and because those formal methods permit exact communication with the persons who are to develop the system according to a specification. In this section we make use of earlier work submitted as Tutor Assessed Assignment for Module 10 (GIS in Organisations) of the UNIGIS postgraduate diploma course in GIS (van Brunschot, 1994).

SSA is performed here in 6 consecutive steps. Not all methods recommended from SSA will be used, however. This is due to the fact that the author is - essentially - the sole user. Furthermore, the urge for a metadata system comes from the author himself. Except for the author there is only general awareness of the (potential) problem, i.e. the need for risk prevention. (The operator’s environment is interested mainly in GIS output.) Therefore there is little case for discussions, interviews and iterations. This chapter was discussed with the Head of the Department and with the second (backup) operator. They made some comments that were embedded in this text. These comments were mostly related to details, except for the remark that the labour cost rate in Section 5.4 initially was set too low.

Analysis starts from the *present* situation. Focus is on GIS as a decision support tool, since problem consequences are felt relatively sharply there.

Problem Definition (Step 1)

The problem and its framework have been outlined at the outset of this chapter. These points do not have to be repeated here. The essence of the problem is in the consequences of poor documentation. These consequences have to do with effectiveness, risks and efficiency. They have impact on corporate headquarters’ interests in a.o. decision support. Although decision support is ultimately a VSN Board problem, it is mainly within the Marketing and Transport Department that the need for a map or geographical analysis is recognised.

Problem: Poor documentation of the GIS of the VSN group from the viewpoint of corporate interests.
Problem boundary: Marketing and Transport Department.

Problem Structuring by means of a Rich Picture (Step 2)

Major Actors and their dominant characteristics:

- VSN Board (outside boundary): needs Decision Support rapidly.

- Head of Dpt.: needs a map often within two hours (for the Board's DS problems).
- Operator (the author): knows the system and the data, but may be ill (or worse) or might forget about data and/or processes.
- Second (backup) operator: knows only 5% of the system and its data.

Major objects: GIS (hardware and software), Data, Documentation System, Output

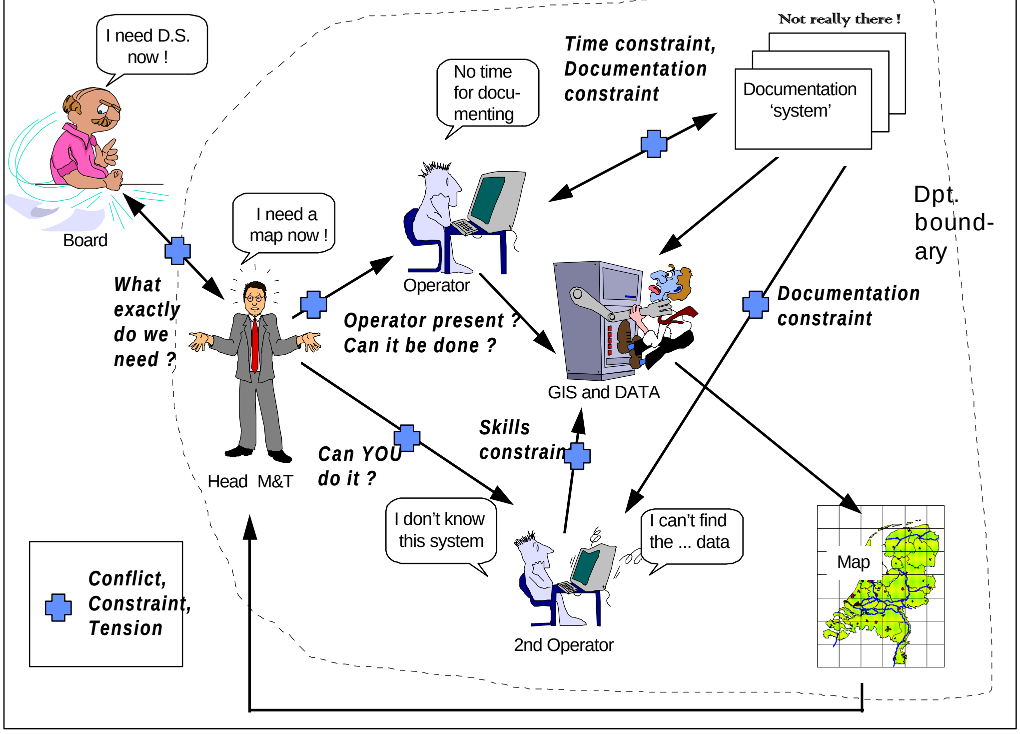
Processes: Documenting (data), Substituting (operator), Requesting (map), Using (GIS), Searching (data).

Box 5-2 contains the *Rich Picture* that has been produced to express the essence of the problem.

Remarks:

- Head of M&T translates DSS needs into map needs.
- The 2nd Operator does not know the GIS and needs training, no matter what documentation is available.
- Tension stems from constraints in:
  - time (map making within 2 hours, work load)
  - documentation (availability, know-how)
  - skills/experience (second operator)
  - personal presence (of operators)
- Effectiveness and efficiency of GIS operation may be low because of invisibility of data and processes and because of suboptimal use and duplication of data.
- The documentation system is the problem (or: solution) that the analysis focuses upon.
- The rationale for the system will be explored further on Section 5.4 (after stating the objectives in Section 5.3).

Box 5-2: Rich Picture of the Problem



System Identification (Step 3)

***C**lients*:	GIS operators
***A**ctors*:	GIS operators
***T**ransformation*:	Storing and browsing GIS metadata, Selecting data
***W**eltanschauung*:	GIS data/process documentation is in VSN's corporate interest
***O**wnership*:	Head of M&T Dpt., Operators
***E**nvironment*:	Constraints in personal presence, time (for documentation and map making), skills/experience (of second operator) call for transparency of GIS and its data and processes.

Root Definition: "A system designed for easy and rapid storing and retrieving GIS metadata for use by *all* GIS operators, to facilitate map (and geographical analysis) production within two hours most of the time (by *any* of the operators) on request of the Head of M&T". No alternative root definitions are needed, since there is no tension related to the system's root definition.

<u>Conceptual Model: what does the system do in ideality?</u> (Step 4)

Type of system: documentation system (for storing and retrieving metadata on GIS data and processes, for the sake of finding appropriate data and processes).

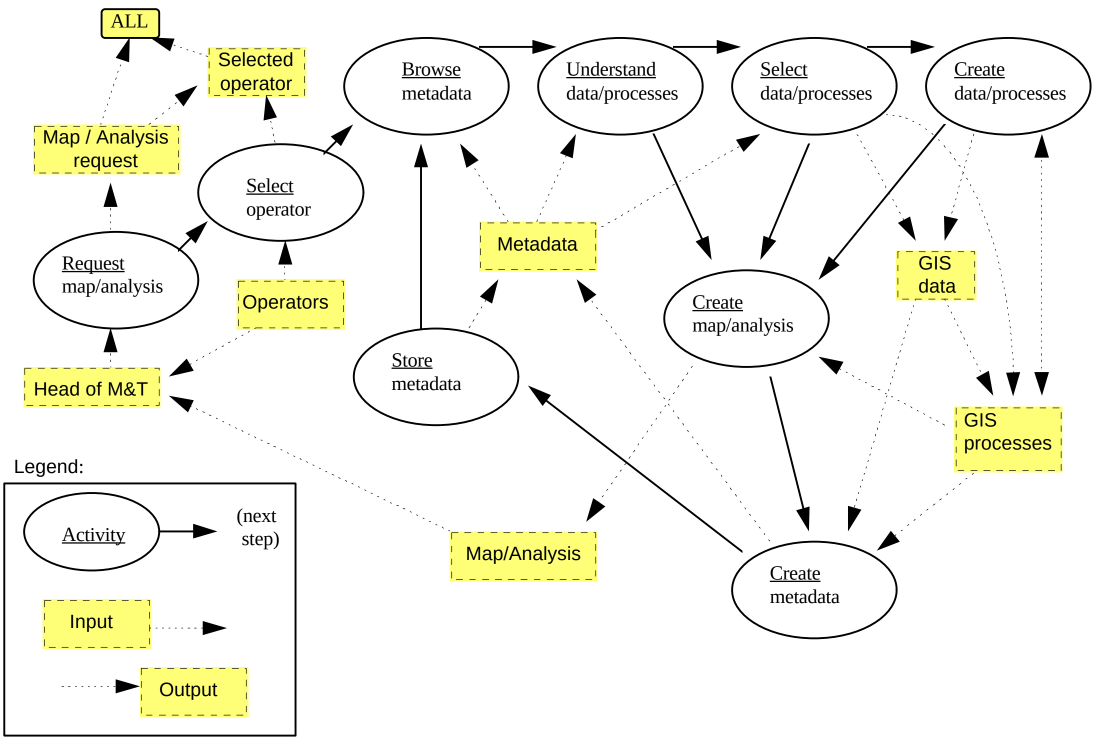
Box 5-3 names the *Transformation Activities* and *In*- and *Outputs*.

Box 5-3 Transformation Activities, In- and Outputs

Verbs **describing *Transformation Activities***	**Nouns** **describing *In-* and *Outputs***
Request Browse Understand Select Create Store	Head of M&T Department Map/Analysis Request Operators Metadata GIS Data GIS Processes Map/Geographical Analysis

Box 5-4 shows linkages between in- and outputs and activities. It constitutes a simplified conceptual model expressing the essence of the intended solution.

Box 5-4: Conceptual model of the essentially intended system



Comparison of Conceptual Model with Current Realities; Feasible and Desirable Changes

(Step 5)

1. The documentation system needs yet to be built (present hardcopy documentation is incomplete and often unclear).
2. The second operator still needs GIS training.
3. Ease and speed of metadata browsing is critical, since rapid finding and understanding the data is critical.
4. Ease and speed of metadata storage is also critical, because of constraints in time (work load) and skills (second operator).
5. Do not forget to document the documentation system.

Conclusion (Step 6)

The analysis provides a basis for further study aimed at defining a conceptual model and requirements of the intended system. Further study needs to go into existing needs (in detail), available (software?) tools, and metadata structures. In that process following future developments and consequences need also to be taken into account:

- portability to different GIS system?
- additional GIS feature types (e.g. network data structures)?
- emerging need for meta-environment entities?
- portability to subsidiaries' GIS?
- attributes stored in relational databases?
- GIS more formal role and framework?
- more (and different types of) users?
- GIS network version?
- (meta-)data through company-wide LAN?
- data integration (standards, data warehouse)?
- follow metadata standards?

5.3. Aims and Objectives for the VSN GIS Metadata Project

Wertz (1986, p.156) points out the importance of stating objectives in developing a metadata system (or: data dictionary, as he calls it): '(...) installing and maintaining a data dictionary is a lot of work. No one should undertake such a project without having a clear idea of the goals to be achieved. A well-understood and reasonable set of goals will also be invaluable in justifying and selling *and planning* the project.

'Since a data dictionary is such an open-ended piece of software, a clear statement of objectives will also keep the scope of the project under control. It will, of course, be necessary to continually reevaluate these objectives as new opportunities present themselves. Even if no one asks for a set of objectives, it is a good idea to take the time to list them, so that *you* know what you are doing'.

The 'goals' or 'aims' of the VSN GIS metadata project[60] have been derived from the problem analysis and (in the case of nr. 3) ultimately from management needs, as expressed in the past. The aims are:

1. Effectiveness: Increase the operators' ability to produce maps and geographic analysis in accordance with management information needs.
2. Efficiency: Reduce map/analysis production time (to two hours in most cases).
3. Reduced Risk: Less chance of error or inoperability.

Following are the 'objectives' that were derived from these aims[61]:

- Support file and data maintenance
- Support data creation
- Support GIS processes
- Keep track of GIS processes
- Keep track of files and databases
- Document sources and use of data

[60] That is: in this (second) part of the thesis specifically.
[61] Each of these objectives is related to more than one of the aims, mostly to all of them. Therefore we decided *not* to present a table showing the relationship between aims and objectives.

- Support hardware use
- Support multiple operators
- Support future IS development methodology
- Support data distribution over subsidiaries
- Support data/process migration to different GIS
- Document data lineage
- Document maps and analyses produced
- Standardise naming of layers, fields, files
- Standardise coding

As indicated by Wertz: it may be necessary to make alterations to these objectives in the course of the development process.

5.4. Building the Business Case

In this section it will be demonstrated that building a GIS metadata management system (MMS) seems to constitute an economically viable project. We shall integrate essential elements of cost/benefit analysis and risk analysis. Risk will be generalised to probability of costs (and benefits) occurring. Probabilities will be used to differentiate costs (and benefits) according to the situation *without* and *with* an metadata management system. Another rationale for this line of analysis is the fact that risk is best seen as the product of a potential damage and the probability of that damage occurring.

Two types of risk can be distinguished in relation to the intended system,
- risk of damage that may be prevented or reduced by MMS,
- risk of damage that may occur as a result of the MMS project itself.

The first risk type is in line with traditional cost/benefit analysis, the second type is the normal subject of risk analysis. However, in our case the risk of the project in minimal, because of following project characteristics:
- single implementor,
- basically single user,
- definite (and accepted) system purpose,
- low impact on others in the organisation,
- system not very complex.

Alter (1992, p.608) calls this a ‘minimum risk system’.

It is worthwile to mention some of Wertz’ (1986, p.150-151) evaluations of cost/benefit analysis in the MMS framework:
- ‘cost reductions’ are often fabrications,
- project cost is often underestimated,

- the analysis constitutes more or less an educated guess rather than a scientific calculation.

Even in the face of many uncertainties it can still be quite rational to conduct a cost/benefit analysis in order to convince others (and oneself) of the economic viability of the project. But we have to admit that some of the costs and benefits cannot be expressed in probabilities and amounts, or only in a highly speculative manner.

We shall now go through the elements of the adapted cost/benefit analysis, assuming that the MMS will meet the objectives formulated in the previous section. The results will be summarized in Box 6.4.

We shall assume that the system will be operational during a (depreciation) period of four years with 42 working weeks each. Labour cost is assumed to be NLG 125,- per hour. The intrest rate is set at 8%

In the 1996 budget an amount of NLG 20.000 was reserved for the MMS. We shall assume that 90% of this budget will be used. Furthermore we assume 20% probability of NLG 10.000,- unforeseen investment, that might otherwise have been used to generate 8% (accrued) intrest during 4 years. We assume that there is a 90% chance that 20 full working days are necessary to implement the system (thesis work not included).

We have identified five costs to be avoided or reduced by MMS:

- 5 percentage points probability reduction of NLG 25.000[62] cost because of system inoperability at crucial moments,
- 20 percentage points probability reduction of disk space enlargement (NLG 1.500),
- 5 percentage points probability reduction of NLG 2.000 redundant data buying (to 0%, in fact),
- reduction of damage due to errors (not quantified),
- reduction of cost of migration from Atlas GIS to a different system (not quantified).

We have identified four MMS benefits:

- 90% probability of 25% acceleration of (0,5 days/week) maps/analysis production,
- 90% probability of 50% acceleration of (0,5 days/week) data maintenance/derivation,
- better decision making and negotiations (not quantified),
- more satisfaction in GIS operation (not quantified).

---

[62] Highly speculative. Among damage prevented is also an MMS contribution to limit the consequences of hardware failure.

Box 5-5 summarizes the calculation and its result.

Box 5-5: Cost/Benefit Analysis for the VSN GIS metadata management system (MMS) (in NLG)

Cost/ Benefit Type	Description	Benefit (Cost)	In ... years	Probability without/with MMS		Result (4 years)	Subtotal
MMS investment	Investment (1996 budget)	(20.000)	4	0,00	0,90	(18.000)	
	Unforeseen	(10.000)	4	0,00	0,20	(2.000)	
							(20.000)
	Intrest	8% accr.	4	0,00	1,00	(7.210)	
							(7.210)
MMS operating	Labour costs (implementation, during 20 days, after thesis completion)	(20.000)	4	0,00	0,90	(18.000)	
							(18.000)
Cost avoidance by MMS	System inoperability at crucial moment	(25.000)	1	0,10	0,05	5.000	
	Disk space enlargement	(1.500)	4	0,50	0,30	300	
	Redundant data buying	(2.000)	1	0,05	0,00	400	
	Maps otherwise not produced	???					
	Errors	???					
	Easier migration to other GIS	???					
							5.700
Benefits from MMS	Map/analysis production 25% faster (basis: 0,5 day/week)	5.250	1	0,00	0,90	18.900	
	Data maintenance/derivation 50% faster (basis: 0,5 day/week)	10.500	1	0,00	0,90	37.800	
	Better decision making	???					
	More satisfaction in GIS operation	???					
							56.700
Assumptions:	Depreciation in: 4 years						---------
	42 working weeks with 5 working days each					**TOTAL**	**17.190**
	Hour cost: NLG 125,-						=======
	Intrest rate: 8%						

Total investment boils down to NLG 45.210. In one year an average gross benefit is made of (56.700+5.700) / 4 = NLG 15.600. The payback period can be estimated as 45.210 / 15.600 = 2,9 years, that is: 2 years and 11 months. Net benefit after 4 full years can be estimated at NLG 17.190,-

We conclude that the project seems to be viable in an economical sense, although some of the calculations were very speculative.

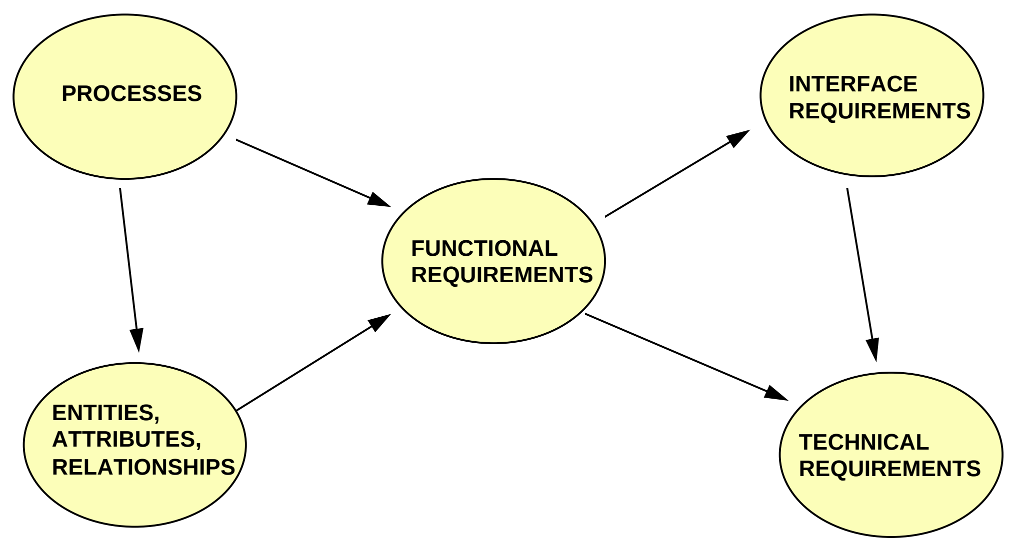
5.5. System Specification

In the previous section it was concluded that it seems worth wile to build a metadata management system (MMS) for GIS at the VSN group. Now it is time to take a firm step in designing the MMS.

This section will present a full functional specification of the intended system. In this process emphasis is not on the particular Information System Development Methodology (ISDM), but rather on the results of it[63]. In fact, the methodology employed borrows from several 'hard' and 'soft' types of analysis.

The structure of these specifications can be pictured as follows:

Box 5-6: Structure of the System Specification



The complete contents of the system specification can be found in Appendix A.3, which follows the five elements pictured in Box 5-6. This section will show only the essentials and crucial considerations that led to the system specification.

Appendix A.3 contains a full description of all processes and information flows that are relevant to the metadata management system (MMS). This analysis shows more detail than the provisional conceptual model pictured in Box 5-4.

[63] We spare the reader all the iterations performed in arriving at the results.

The MMS follows the relational data model. It consists of tables (relations between rows or ‘tuples’ and columns or ‘attributes’) interconnected through keys. The relational model seems to be ideal for conveying the structure of an intended system from a non-programmer to a programmer. It allows great rigour and clarity in data definition, while the disadvantages seem to be minimal in the case of the VSN MMS: since the metadatabase will be relatively small and simple the limitations of the relational data model will not pose any threats to our endeavour. Entity-Attribute-Relationship modelling (EAR- or ER-modelling) is used: this technique fits well the current situation where strong apriori ideas exist regarding logical structure.

The conceptual (or logical, or ER) model (Box A-8) presents metadata entities and their relationships. Our metadata entities can be divided in data, product, process and environment entities. It is important to keep in mind that actual data, products, processes and environments will *not* be part of the metadata management system. Rather, they are represented by (mapped into) the (meta)entities and their attributes in the MMS, which will be created in a Relational Database Management System (RDBMS).

A central position is taken by the entity ‘Product’. A Product may consist of several elemental Products. Each Product must have one link to one and only one of seven information/data entities (a file - including a map - , an element of a file - e.g. a coverage or a field - , or another Product). In fact, products and data come close to one another. The link between data and its source, data collection method, lineage, originator, operator, and output medium is achieved through the central Product entity[64].

The conceptual model shows some degree of integration with typical Atlas GIS data structures. This is a prerequisite for a metadata model that is not built as a full generic metadata model (that might be fitted to any GIS data structures). The building of a really generic metadata model is outside the scope of this thesis. The links to the paculiarities of Atlas GIS file and data structures are necessary to facilitate system maintenance. Design of a generic data model that would be both general and allow for all particular Atlas GIS file and data structures would call for quite professional IT capabilities, not available as a resource for this thesis. For future conversion of the VSN MMS to a different GIS this limitation of the present design will not constitute a problem since the conceptual model and the resulting system will facilitate the eventual shift to a different GIS system[65], because present data structures will always be documented in detail if the system is used properly.

---

[64] One of the reasons for this was the wish to keep the design straightforward, and therefore not too expensive to build.
[65] Which might become reality in a few years, e.g. because topological data structures or closer links to the Oracle DBMS might become imperitive or as a long-term result of the recent (May 1996) take-over of the Atlas GIS product rights by ESRI.

As motivated before (e.g. Section 5.1) it is our wish to adopt as much as possible from the CEN/TC 287 metadata standard (CEN 1996a). Considering the case of GIS at the VSN group (practically one user, not very complex applications, no data delivery to the outside world) the need for the VSN group to follow a metadata standard is minimal. An attempt to adopt the standard is made anyway, mainly with an eye on possible future developments where GIS packages and obtained datasets may more and more follow European standards for geodata. In this attempt, however, the (mainly organisation-*internal*) aims and objectives of the intended MMS were not allowed to get out of sight. The MMS is intended to make GIS operation less risky but also more *efficient*. This means that:

- situations must be avoided where completion of metadata would create a conflict with needs for efficient working;
- therefore only those metadata standard elements are embodied into the MMS which seem to help in GIS use;
- metadata elements are not restricted to metadata standard elements;
- an attempt is made to fit the Express-G structures for metadata definition into the Atlas GIS data structures, but no conflicts with Atlas GIS data structures are permitted;
- deviations from the CEN/TC 287 metadata standard are mentioned.

The entities and their attributes (Box A-9) do indeed show deviations from the CEN/TC 287 metadata standard. Appendix A.3 (Box A-10) discusses all deviations along with the rationales. Typically, there are five types of deviation from the European metadata standard:

- non-used mandatory elements
- mandatory elements used in a different manner
- mandatory elements used as optional
- revised naming
- revised data format

The described processes, entities, attributes, and relations dictate the functional requirements, interface requirements, and technical requirements. Again, all details are comprised in Appendix A.3.

This concludes the requirements for the VSN GIS metadata management system. We feel that it is more than just ‘requirements’: that appendix A.3 contains most of the system’s ‘design’. Yet, a lot remains to be discussed with the person(s) who is/are to build the system.

It is interesting to relate the system specification to analyses and projects mentioned earlier in this study. Taking the system's aims and objectives (see the end of Section 5.3) into consideration following observations can be made:

- Regarding the conceptual metadata models in Chapter 3, the MMS shows characteristics of lineage (Section 3.3) and a meta-database as described by Trivedi and Smith (Section 3.2).
- Regarding metadata projects in Chapter 4, the MMS shares elements with the Dutch metadata systems described in Sections 4.3.1 - 4.3.3.
- The MMS has no links to clearinghouse systems (Sections 3.4, 4.1, 4.2, and 4.3.4).
- The MMS does however have a relation with the European metadata standard (Section 4.2).

5.6. Pilot

To design a metadata management system (MMS) on paper is one thing, to bring it into actual existence is another one. This thesis does not include actually building the system. However, there are two good reasons to create a pilot project using existing tools:

1. It may turn out that *existing* software permits building the system. This is crucial to define the plan for further action, which is an element of the thesis.
2. A pilot project will probably give clues as to the technical applicability of the MMS design presented: some aspects of the current design may, after having a second thought about it, appear difficult to build.

One could say that the pilot's focus is both on the current design and on current software. Both might appear eligible for adaptation to the other.

A software package for building an MMS specifically for Atlas GIS is not known. A quick scan has revealed that one software package seems to be promising relative to other packages: GeoKey, supplied by Geodan, an Amsterdam based GIS-house. GeoKey qualifies for the pilot because it:

- is a well-maintained commercial package,
- is already in use at a GIS site contained in this thesis: Rijkswaterstaat - see Section 4.3.2),
- is available for the Windows 3.x platform (which is used for the VSN GIS),
- shows generic design: it can be used for a variety of GIS packages, including Atlas GIS,
- makes use of the MS Access RDBMS, which is already available at the VSN GIS site.

In the pilot it will be demonstrated what - in the framework of our MMS design - is possible with GeoKey and what isn't. The two standard modules of GeoKey and the accompanying manuals (Geodan, 1995b and 1995c) were made available by Geodan for the exclusive purpose of this pilot. Other modules are in development (Geodan, 1995a).

GeoKey Select is the module for actual use of the metadatabase, GeoKey Edit is the module for filling or making alterations to the metadatabase. GeoKey is **generic** software in the sense that it can be used with any GIS package, GeoKey's standard metadatamodel being rather general. It is not generic in the sense that it could be flexibly adapted by the customer to seamlessly fit all file and data structure aspects of any GIS. This is demonstrated in Appendix A.4. What will be described in the pilot is the standard GeoKey software (which can, of course, be adapted by Geodan).

GeoKey's basic objects are the *datasets* on which metadata is available. A dataset is of one of several *types* (e.g. geographic, attribute, software program, map composition, earth observation). Dataset types classification is flexible. Datasets have *values* on *metavariables*[66]. Such a value is the basic metadata element. Variables can be *general* (applying to all dataset types) or *specific* (applying to one or more dataset types). The content of a dataset can be decribed in more detail by its values on *items* (descriptions of the kinds of elements in a dataset), *keywords*, (position in a) *hierarchy*, (position in a) *geographical classification*, and *geographical coordinates* (defining a rectangle in a particular projection system).

GeoKey Select is meant for selecting (and viewing pictures of) one or more datasets and transferring it to the GIS software. Selection is based on a combination of six selection methods: type, keyword, metavariable, hierarchy, geographical classification, and geographical rectangle.

GeoKey Edit is not only for entering or altering metadatabase content, but also for structuring it. There is a large degree of flexibility in using metadata. E.g.: *items* **might** pertain to Atlas GIS map layers, *type* **might** be Atlas Script syntax, *hierachy* **might** pertain to VSN's organisational structure, *geographical classification* **might** pertain to Dutch provinces and/or transport regions, etc.
Furthermore it is possible to define:

- types of metadata shown to user groups,
- user authorisations,
- metavariables ranges,
- pictorial representations of geographically referenced files.

After this first acquaintance with the GeoKey software it is now time to see to what degree our MMS

---

[66] Actually they are called 'variables' by Geodan but the word 'metavariables' allows discrimination between meta- and basic variables.

design would actually fit into these standard MMS tools.

Appendix A.4 contains all documentation on all steps taken to build a MMS that comes close to our design. The use of GeoKey in the pilot can be evaluated as follows underneath.

Among strengths of GeoKey are:

- Software existing, tested, documented, maintained.
- Available for multiple platforms.
- Ample selection procedures.
- Flexible, applicable to different GIS systems.
- Transfer of selected data to applications is possible.
- Authorisation infrastructure.
- Pictures of datasets appear in screen.
- Interface clear and appealing.

Among major weaknesses of GeoKey are:

- Poor (or - at the best - non-rigid) representation of ER diagram from the MMS design (focus is on datasets rather than on elements of datasets).
- No automatic metadata generation.
- Pictures of datasets on the GeoKey Select screen do not follow dataset modifications.
- Batch programming necessary to have GeoKey performing well.

The first two weaknesses deserve strong emphasis, since:

- The ER diagram reflects basic metadata needs.
- Non-rigid ER diagram implementation for dataset elements creates maintenance and reliability problems.
- The pilot has shown that much time is needed to find and enter all metadata manually, while according to this project's aims and objectives[67], the MMS should save labour time.

The conclusion is that, despite of the many advantages of the GeoKey software, there is a preference to develop a new MMS that allows more closely following details of our ER diagram and allows automated metadata generation.

[67] See the end of Section 5.3.

6. Conclusions

This chapter will contain the harvest from the present study. In fhe first section conclusions will be drawn from what has been seen in the previous chapters. The second section comprises a simple plan for the future GIS metadata management system at the VSN group.

6.1 Where do we stand?

This study has been aimed at the design of a GIS metadata management system (MMS) for the VSN group, a large provider of collective personal transport in the Netherlands. The MMS will describe and yield access to data, processes and environments of the Geographic Information System operated at VSN corporate headquarters, currently Atlas GIS, a vector standard package for desktop use. Design of this MMS has been undertaken after a study of GIS metadata end its relevance for a company (in general, and for VSN in particular). The study has culminated in a pilot project where standard MMS tools have been used to try and implement crucial aspects of the chosen design.

This study was intended to serve two aims:

1. Demonstrate the company relevance of GIS metadata in general.
2. Design a GIS metadata system for the VSN group.

ad. 1

The company relevance of GIS metadata was shown as follows. In many Dutch companies GIS is considered to be relevant, especially as an operational tool. Expectation is that GIS will become a strategic tool as well, and that its relevance certainly is not restricted to the Netherlands (Section 2.1).

The GIS house concept can be used to demonstrate the interrelatedness of the various datasets comprising a GIS (Section 2.2).

Metadata (being the ‘glue’ connecting elements of the GIS house) serve to to identify and select *data* on the basis of properties such as content, source and quality. This can be augmented to *processes* and *environments* relevant to GIS use (Section 2.3).

Management of metadata can be organised making use of several underlying conceptual models (Chapter 3).

The possibilities of GIS metadata use were demonstrated from a number of practical implementations, showing also the relevance of metadata standards and providing clues as to how to build an MMS, with special reference to Atlas GIS (Chapter 4 and Section 2.4).

On the basis of all this it was shown in Section 5.1 that GIS metadata can solve and prevent problems arising from discrepancies between company entities related to GIS use. Such discrepancies otherwise could well prevent rational GIS use. If a company relies on GIS as an element essential to its primary process, management of GIS metadata a fortiori calls for an automated MMS (while MMS may - in general - also be manual).

ad 2.

Design of an MMS for the VSN Marketing and Transport Department started with the assessment that documentation of the existing GIS is poor from the viewpoint of corporate interests. The problem was first structured by a graphical representation of major actors, objects and processes. On that basis - as a step toward solving the problem - a rough MMS identification was made in a preliminary model describing transformation activities and in- and outputs. Comparison of this model with current reality provided a basis for further study on the conceptual model and requirements of the intended system. A number of possible future developments and consequences were stated to be taken into consideration (Sections 5.2 and 1.2).

On the basis of this analysis objectives for the VSN MMS project were identified, related to the aims of effectiveness, efficiency and reduced risk (Section 5.3).

By a combination of cost/benefit and risk analysis it was shown that building the MMS seems to constitute an economically viable project (Section 5.4).

All processes and information flows relevant to the MMS were identified, leading to an Entity-Relationship (ER) logical model. The prime reason for using the relational data model and ER-modelling was found in the strong apriori ideas regarding logical structure of the intended system. An attempt was made to implement as much as desirable from the emerging European metadata standard (see also below). This led to lists of functional requirements, interface requirements and technical requirements (Section 5.5).

In a pilot existing MMS tools were tested in an attempt to implement the design. This proved only partly possible with the standard tools. Preference was stated for a specifically developed system that - unlike the pilot MMS tools - would facilitate full representation of the logical model and save labour time by automated metadata generation (Section 5.6).

Two essential topics call for specific attention since they have been pervading this thesis every now and then without full coverage in the conclusions above: (1) the use of metadata standards and (2) the VSN MMS case in comparison with broader developments. These topics are partly related.

Section 5.1 ended with: “While metadata (...) acts as glue between the elements of the GIS house, metadata standards act as glue between different GIS houses" A geographical metadata standard[68] makes data in one GIS house comparable with those in another GIS house elsewhere. This constitutes advantages for companies that fit in several of the following categories:

- companies with standard GIS software (which will - hopefully - be adapted to metadata standards someday);
- companies that buy and/or sell their data (since marketed data and the clearinghouses yielding access are also likely to follow metadata standards);
- companies with GIS as an essential tool in its primary process (since following a standard reduces risk involved in (meta)data structures;
- companies where several departments and many operators make use of geographical data (since there will be a need for a common understanding of (meta)data content and structure);
- companies where the same geographical data is used in different formats and different systems (since data conversion is much more efficient and less risky if the data is described in standard ways);
- companies that share (or compare) data with other companies (which calls for a common basis).

In this chapter use has been made of the word 'metadata', which may elicit the associaton of data about data. In reality metadata may also comprise processes and environments, as we have seen. Metadata standards largely cover the metadata on data, but lack substance on the data on processes and environments. (There are however other - related - standards covering those aspects.) So a company will do wise devoting explicit attention to the question how to provide a common understanding of system processes and environments. Furthermore a company will have to consider the pro's and con's of making alterations and additions to a metadata standard. Advantages can be found in full representation of a company's peculiarities; disadvantages can be found in reduced comparability/exchangeability with the outside world.

In many companies use of a metadata *standard* may in the long run reduce labour time (involved in data searching, in multiple data conversions and in learning time). On the other hand, following a metadata standard may in the short run involve a significant time investment in (e.g. Express) education and providing and entering metadata, as we saw in the pilot (Section 5.6). And that while metadata itself is often used to save labour time. It is up to an individual company to weigh all advantages against all disadvantages, and the short run against the long run.

In the VSN group case following the European metadata standard has been limited by available labour

[68] As such standards have been emerging in e.g. USA and Europe.

time, by cutting out all aspects which aren't covered in our geographical data, and by an attempt to fully represent data and file structures within the particular GIS in use: Atlas GIS. It is interesting to note that this fit to Atlas GIS has been made while it is very well possible that in a few years from now GIS at the VSN group central office will be different from Atlas GIS. This brings us to the nature of the MMS proposed: why has a 'dedicated' MMS been proposed instead of a 'generic' one? Why represent Atlas GIS structures if in a few years from now there may very well be quite a different GIS installed? The answers can be linked to the aims of the VSN MMS project itself (Section 5.3):

- Effectiveness of GIS operating should be increased. To enhance the ability to produce maps and analyses in accordance with management information needs it is important that all relevant Atlas GIS-related aspects of data, processes and environments are covered in metadata, since it is Atlas GIS that will have to do the job (not to speak of the operator).
- Efficiency should be increased. To reduce map/analysis production time it is important that all metadata can be found quickly. This can be achieved only if search for non-MMS-covered Atlas GIS-related aspects is prevented.
- Risk of error and inoperability should be reduced. This can be achieved by eradicating any unnecessary discrepancy between GIS and metadata.

An alternative would have been to write the specification in such a way that the resulting software could support adaptation to specific packages. However, this would call for greater IT skills than available to the author. Expectation is that the *explicit* coverage of data, processes and environments in metadata will further any conversion to any future GIS, since conversion must start with knowing what one has.

What more is typical for the VSN group case? Basically we see a single-user and quite simple application of desktop GIS standard software, while most MMS projects are in companies with larger and more complex application of (often tailor-made) GIS software. This is to say that need for a formal MMS in the VSN group case is rather limited compared to other companies. Yet the investment in such a system was estimated to yield positive financial result after a period of less than three years (Section 5.4).

6.2 What's next?

The pilot led to the conclusion that developing a new MMS following all details of our logical model deserves a preference over the use of existing MMS tools. The next step is the transfer of this study (especially Chapter 5 and Appendix A.3) to an agency that will produce a proposal and a budget for a system to be built according to the full specification. It remains possible that the specification will, in that process, undergo changes. Furthermore there is a chance that due to changing priorities the decision will be made *not* to build the MMS or to wait for new GIS software.[69]

[69] E.g. a new generation of mapping software that might bloom from the 1996 acquisition of Atlas GIS by ESRI, the producer of ArcInfo and ArcView. See Section 4.4 for a brief overview of recent developments with Atlas GIS.

Appendices

A.1. Details of Spatial Metadata Projects

The United States

In 1990 the Federal Geographical Data Committee (FGDC) was established by the Executive Office of the President of the United States (Darman, 1990). The main objective of this "interagency" was the coordination between several federal activities in the field of spatial data, in order to arrive at the development of a national digital "spatial information resource". Among its intended activities were:

- 'Promoting the development, maintenance, and management of distributed data base systems that are national in scope for surveying, mapping, and related spatial data;
- 'Encouraging the development and implementation of standards, exchange formats, specifications, procedures, and guidelines' (Ibid., p.7).

'Currently, within the FGDC, approximately three hundred individuals representing a variety of federal agencies and disciplines comprise a structure of committees, subcommittees, and working groups' (Tosta, 1995, p.7.8).

In 1993 'the concept of a coordinated "spatial information resource" was further explored in a report of the Mapping Science Committee (MSC) of the National Research Council. The MSC coined the phrase "National Spatial Data Infrastructure" (NSDI)' (Ibid., p.7.8).

Famous is President Clinton's Executive Order 12906, wherein officially developments are started toward (among other initiatives) an NSDI and a National Geospatial Data Clearinghouse. NSDI means 'the technology, policies, standards, and human resources necessary to acquire, store, distribute, and improve utilization of geospatial data', and a National Geospatial Data Clearinghouse means 'a distributed network of geospatial data producers, managers, and users linked electronically' (Clinton, 1994, p.17671).

In accordance with Executive Order 12906 four major NSDI activities have evolved[70]. These activities are interdependent (Tosta, 1995, p.7.10-12):

1. National Geospatial Data Clearinghouse
2. Geospatial Data Standards
3. National Digital Geospatial Data Framework
4. NSDI Partnerships.

Ad 1. National Geospatial Data Clearinghouse

---

[70] Or rather: *are* evolving, since Executive Order 12906 has scopes for 1998 and 2000.

From the Internet we adapted following text (FGDC 1995b): ‘As part of the NSDI, which is building on initiatives such as the National Information Structure, the clearinghouse allows data providers to make known what geospatial data exists, the conditions of this data, and instructions for accessing this data. Each data provider describes available data in an electronic form and provides these descriptions (or “metadata”) to the network using a variety of software tools. In addition to these metadata, the provider also may provide access to the geospatial data. The data described in the clearinghouse may be located at the sites of data producers or at sites of designated data disseminators throughout the USA.’

‘The clearinghouse uses the Internet to link computer nodes that contain information about geospatial data. (...) Using the Internet, data users can search the descriptions provided by producers to locate data that are suitable for their applications. A public-domain software tool known as WAIS (Wide Area Information Servers) enables users to perform queries for data over the network. A recent enhancement to WAIS allows users to search on the basis of geographic coordinates. WAIS uses a communications protocol known as Z39.50, which is emerging as an international standard. In addition, the rapidly expanding World Wide Web (WWW) technologies, including browse software such as Mosaic[71], provide gateways to WAIS servers, and allow graphic or forms-based queries for data’ (Ibid.).

As the clearinghouse evolves, additional functions will be supported, such as the capability for producers to publish data that are being prepared or are planned and for users to advertise their data needs. Thereby communication on new product needs and developments will develop and help to form partnerships for data production and to minimize duplication of data collection. This communication is vital to ensure that the NSDI continues to be responsive to the needs of the community (Ibid.).

Ad 2. Geospatial Data Standards

‘The second activity area is the development of standards. Most of the FGDC subcommittees and working groups are developing standards for data collection and content, data presentation, and data management, to facilitate data sharing. (...) The FGDC also formally adopted the Spatial Data Transfer Standard for the transfer of data sets among unlike computer systems. All of the FGDC developed standards are subjected to an extensive public review process that includes nationally advertised comment and testing phases’ (Tosta, 1995, p.7.10-11). The World Wide Web contains a useful overview of geospatial data standards, among which are those developed under FGDC control (Heine 1996).

Ad 3. National Digital Geospatial Data Framework

The third FGDC activity area is conceptualising and testing the development of a digital framework dataset that will minimise redundancy in data collection and facilitate the integration and use of geospatial data. ‘This dataset is envisioned to consist of the most commonly required datasets for most geospatial data applications, including digital orthoimagery, geodetic control, elevation, transportation, hydrology, administrative boundaries, and cadastral or ownership information. These themes of data will likely vary in resolution over different geographical areas and will likely be developed by different organisations, including state, local, and federal government agencies and the private sector’ (Tosta, 1995, p.7.11). The framework is a basic, consistent set of data and services that will (FGDC 1995a):

- provide a geospatial foundation to which an organisation may add detail and attach attribute information,
- provide a basis on which an organisation can accurately register and compile other themes of data,
- orient and link the results of an application to the landscape.

---

[71] Or (more popular:) Netscape Navigator. Future clearinghouse developments include using the new JAVA technology where end user computers simultaneously receive data as well as software from the Internet.

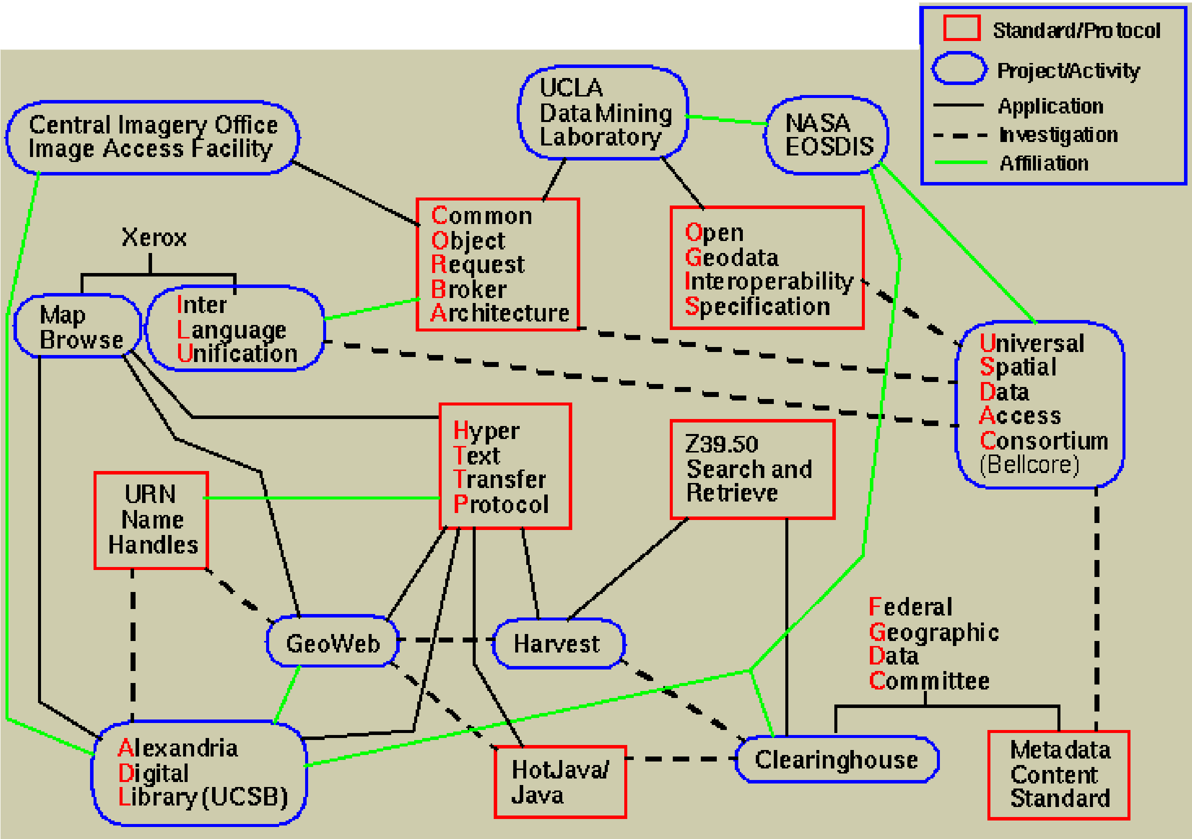
Ad 4. NSDI Partnerships.

'Development of the NSDI as a networked, distributed enterprise requires new relationships and partnerships among different levels of government and between public and private sector entities. The FGDC is promoting partnerships in a variety of ways' (Tosta, 1995, p.7.11-12):

- encouraging states to coordinate geospatial data activities,
- encouraging use and development of standards and the clearinghouse among non-federal sectors: a maximum of $25,000 per project is available as seed money for creative partnerships to carry out such activities,
- organising fora to discuss partnerships and to coordinate activities among levels of government and between sectors,
- establishing a database of current partnership opportunities related to geospatial data,
- developing the capability within the clearinghouse to identify parties with similar data interests.

Once again we now turn to metadata, the core of this study. The current infrastructural basis for access and retrieval of clearinghouse metadata is depicted in the diagram in Box A-1, that was taken from the Internet.

Box A-1: Spatial Information Access and Retrieval MetaMap (FGDC 1995c)



We can see that the metadata infrastructure is under construction, and that there is a certain amount of complexity involved in being an FGDC Clearinghouse Node (see also Nebert, 1995). It is, of course, not this complexity that the clearinghouse end user must face. The end user's complexity has much more to do with all the hypertext link levels that must be crossed[72] in order to find the data desired. Work is being done to reduce end user complexity (Ibid.).

'Both the spread of technology and ease of access to geospatial data are creating an environment in which thousands of organizations require and produce data for use, sale, and distribution. Numerous data sets are available, which differ in accuracy, completeness, currency, and cost. Users often have great difficulty finding, filtering, and identifying useful data to meet their needs. For many areas within the U.S., current, accurate digital geospatial data do not exist, while for many other regions, duplication is rampant. Better means of coordinating, collecting, organizing, finding, and sharing data are required to maximize investments being made in geospatial data. The NSDI has evolved as a means to begin to address these issues' (Tosta, 1995, p.7.9).

Europe

We cite the EGII (European Geographic Information Infrastructure) Policy Document to get a feel for the background of the EGII project: 'The advantages of the EGII include efficiencies of scale in a unified market, reduced problems for trans-border and pan-European projects, a common European-wide spatial database for all, efficient technical solutions for future growth, increasing use of European skills, improved market position in GI and better results of European-wide planning and decision making'. The EGII project is very interesting because of the need to overcome member state peculiarities and coordinate and adapt national infrastructures and efforts in the field of geographic information (GI2000 Document, 1995).

From two documents of the Dutch Advisory Council for Information on Real Estate (RAVI, 1994a and 1994b) we summarize two central concepts in the CEN/TC 287 standard: the object-oriented approach, and the nominal basis for data specification. These may help in developing VSN's metadata system in (partial?) accordance with leading developments, *before* such will involve capital destruction.

CEN/TC 287 will be based on an object-oriented approach in the sense that elements in reality (*objects*) are leading in defining model elements (*entities*, as abstractions from similar objects). The similarity of objects is judged on the basis of *features*, describing objects. Entities are described by *attributes*, which include a domain of attribute values[73].

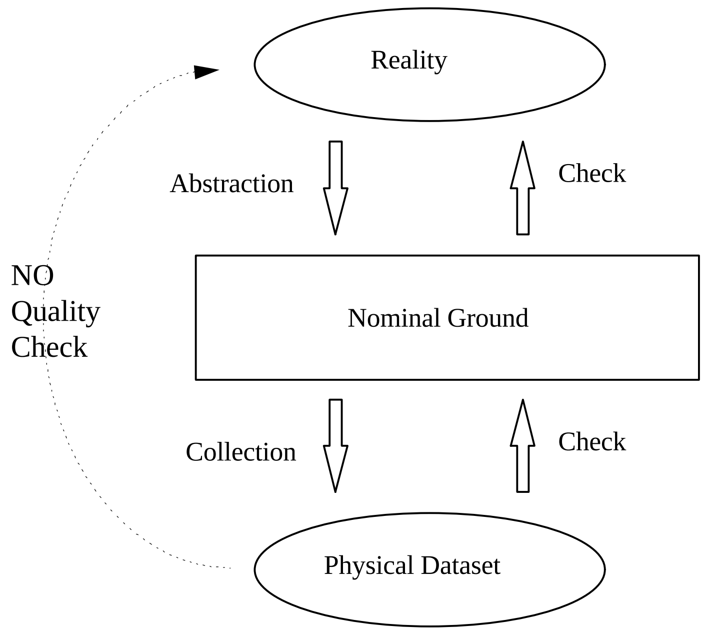
The nominal basis for data specification is a principle for judging the quality of data. According to this principle physical data is not checked against reality, but against a model of reality as a nominal basis. This is deemed necessary since reality shows far too much detail and variation to provide useful quality checks. The nominal basis as a relevant model of reality calls for checking of its own. The following diagram may clarify the concepts mentioned.

---

[72] Sometimes up to 15.
[73] An entity can be simple or complex (i.e. single or multiple/composed). It has a unique identification and at least one attribute as a classification code. In the conceptual schema relations between entities are defined. Each relation has a unique identification, two entity identifications, and minimally one attribute (the name of the relation). An attribute has a type description and a value. Attributes can be divided into (RAVI, 1994a, p.8):

- Identifying attributes: for unique discrimination of one entity against other entities.
- Descriptive attributes: discrete or continuous, quantitative or not.
- Geometrical attributes: to describe location, form and topology of entities.
- Graphical attributes: to describe entities that deviate from standardised structures.
- Meta-attributes: to describe data about entities, attributes and relations between these.

Box A-2: Function of the Nominal Basis for Geo-information (RAVI, 1994b, p.12)



The GI2000 document (1994/95) states a.o. action to enable access to Europe's GI resources: 'Geographic information metadata services should be encouraged, which all potential users can consult in order to prevent unnecessary data collection efforts and to encourage sharing and reusing of existing data sets. Existing sources of metadata should be publicised and made available using the "information superhighways". Each metadata service would provide information about available geographic data sets, whether the data is digital cartography, business demographic data, census data, national statistical data, etc. (...) It has been proposed that co-ordination of current metadata service initiatives would be beneficial, at both national and disciplinary levels, in order to achieve the most effective and efficient EU-wide access to such data'[74].

The Netherlands

Details on the Metasys prototype at the Staring Centrum**:**

After brainstorm sessions, a list of functional requirements was devised, which was used to investigate the consequences of these requirements. We present (in Box A-3) the original list[75] combined with an evaluation later made as to the priority of the requirements.

[74] 'Current initiatives in the metadata arena include the GDDD initiative of MEGRIN, the SNIG metadata service of CNIG in Portugal, and the proposed GIDD project of EUROGI, plus the private sector OMEGA project, which receive partial funding from DG XIII's IMPACT programme' (GI2000, 1994/95).

[75] The project reports contain, of course, much more information than these requirements and their consequences.

Box A-3: Functional Requirements for Metasys (adapted from Bisoen and Heerikhuisen, 1995a/c)

User's requirements	Administrative requirements	General requirements
• **Alphanumerical searching for meta-information (keywords, coordinates)** • **Geographical searching for meta-information** • **Browsing meta-information of selected dataset** • **Metadata export (conversion/ output)** • **Input of meta-information on basic and on derived files** • **Data quality judgement** · *Browsing alphanumerical data* · *Browsing/viewing graphical data* · *Changing meta-information* • Data export (conversion/output) • Sending messages about files between user and administrator • (Searching for data) • (Operations on data)	• **Input of meta-information on basic and on derived files** • *Changing meta-information* • **Moving meta-information** • **File/Map version management** • **User's authorisation** • **Control access to (groups) of files** • **Data quality judgement** • Sending messages about files between user and administrator	· *Online help function (no manual needed)* · *Fast response* · *Available to many working stations at various platforms* · *All meta-information centrally stored* · *User's registering and logging* LEGEND: · **Highest priority** · *Desired* • Less priority • (Dropped)

After this exercise five options and alternatives for the metadata information system construction were formulated (Bisoen and Heerikhuisen, 1995c):

1. Hardcopy (book) format
2. World Wide Web
3. Construction in SQL Forms 3.0
4. Limited CSDGM implementation in Oracle Forms 4.5
5. Full CSDGM implementation in Oracle Forms 4.5

The prototype was developed in Oracle Forms 4.5, whereby the choice between options 4. and 5. was left open. ER diagrams were developed (showing the relations between entity tables and attributes), and screens were devised. The results were laid down in Bisoen and Heerikhuisen 1995d, from which we derive an alphabetical list[76] of entity tables, which may help in developing the metadata information system at the VSN group:

Altitude Resolution
Attribute
Attribute Date
Attribute Domain Values
Grid Coordinate System
Horizontal Coordinate System Definition
Identification Information
Keyword

[76] We omit tables containing only secondary keys but no attributes. We leave the attributes themselve out.

Browse Graphic
Citation Information
Computer Contact Information Network
Contact Address
Contact Electronic Mail Address
Contact Facsimile Telephone
Contact Information
Contact TDD/TTY Telephone
Contact Voice Telephone
Coordinate Representation
Data Quality Information
Depth Resolution
Detailed Description
Digital Form
Digital Transfer Options Offline
Digital Transfer Options Online
Distribution Information
Geographic
Map Projection
Metadata Reference Information
Multiple Dates
Non-digital Form
Online Linkage
Originator
Overview Description
Planar
Point and Vector Object Information
Process Step
Range of Dates
Raster Object Information
Recording Density
Recording Format
Source Citation Abbreviation
Source Information
Spatial Data Organisation Information
Vertical Coordinate System Definition

A.2. Atlas GIS Users' Reports

GIS users in several organisations were asked questions about metadata needs in order to construct a firmer foundation for the VSN metadata system. Since this system will be dedicated to Atlas GIS metadata it makes sense to interview specifically Atlas GIS users. This appendix contains the reports of these interviews and also the results of a questionnaire posted in the Atlas GIS newsgroup on the Internet.

The first contacts for the interviews were made by VSN's GIS supplier Oasis, the criterion being that the respective organisations would effectively make use of Atlas GIS software. Persons agreed to be interviewed on the subject, after which the contact with the VSN group was established and an appointment made. Two weeks before the interview data the interviewees received an introduction letter and a list of issues, to allow them some preparation:

Box A-4: Topics for Interviews about Atlas GIS metadata (translated)

Introduction

- the VSN group
- the project
- purpose of the interview

The Organisation Visited (not into depth)

- markets and products
- organisation structure
- information systems

Data for GIS (not into depth)

- geographical data
- attribute data
- database management systems

Atlas GIS

- types of applications
- organisational aspects
- technical aspects

Is there metadata/documentation for Atlas GIS applications?

If so:

- what metadata/documentation?
- what use made?
- more wishes?

If not:

- is this a problem?
- what metadata/documentation would be useful?
- are metadata developments started?

A.2.1. Interview 1: Province of Noord-Holland: Bureau Verkeersonderzoek en -Studies

On November 7, 1995 a 1½ hour interview was conducted with Mr. L.R.P. de Jong, traffic expert and main GIS user at the Bureau of Traffic Research and Studies (VOS), which is one of the bureaus of the Department of Roads, Traffic and Transport (WVV) of the Province Noord-Holland (which includes the national capital city of Amsterdam). He was assisted by Mr. A. van Loevezijn, databank administrator in the Staff Bureau of Finance, Information, Organisation and Automation (FIOA).

The VOS Bureau is responsible for traffic counts and forecasts in the Province, to provide a basis for road (traffic) policies. There is also an inspection responsibility[77]. VOS applies traffic models (TRIPS applications, and the North Wing Model in cooperation with other organisations).

Information systems at the Province have been organised within the separate Bureaus (island automation). Most of the time this situation is deemed satisfactory, although there is a diversity of platforms. The Province is in the process of adding data warehouse technology to the information systems. Applications remain compartdecentralised as they are, but contents are subject to nocturnal copying to a central RDB database to provide overnight data access for other users, with an interface based on Visual Basic and in future possibly GIS. For the new situation metadata (in data dictionaries) are seen as crucial.

In the current situation two desktop GIS systems are used by the Province: Arc/View by the Department of Environment and Water, and Atlas GIS (for DOS, at the moment) by VOS. Furthermore there is (within the WVV Department) a Survey Bureau where AutoCAD is used. More Bureaus are interested in producing maps from database materials. A working committee is developing a strategy for a future standard for GIS in the organisation. We concentrate on the current application of Atlas GIS in VOS.

Atlas GIS is frequently used as a mapmaking tool to visualise traffic intensities on road sections and on relations between areal units. Within VOS there are 4-5 Atlas GIS users with separate computers. The main geographical (vector) features are:

- Areal units (with surface areas somewhere between 4 and 5 position postcode levels) have been digitised from AutoCAD hardcopy maps.
- Positional reference is offered by the location of urban areas, roads and waterways, which were all derived from AutoCAD as well.
- Seldomly used are municipal and provincial borders.
- Heavily used are relations between areal units. These straight connections between areal unit gravity centers are imported on an ad hoc basis from the TRIPS software, and converted by specially developed conversion tools. In GIS these features are overwritten at every case of a new application.
- The future use of a (main) road network is foreseen.

Attribute data are population and economic characteristics on the areal unit level, and traffic data for the ad hoc areal unit connections. All attribute data are imported and converted from TRIPS. A separate DBMS is not needed for managing these data.

[77] There is a separate Bureau for typical transport affairs. Provincial reponsibility for public transport will augment, due to a national policy shift.

A serious lack of GIS metadata is felt, especially for the version management of urban areas and of areal units. Changes in these coverages are made on a piecemeal basis by several users on separate systems. Every change is made directly to the user-specific Atlas GIS geographical (*.AGF) file on the C-drive of the users' computer. (Atlas GIS for DOS opens only one geographical file, while the Windows 3.x version encourages geodata administration in multiple geofiles). The changes to coverages are therefore impossible to trace and manage without administration by hand. That is one of the reasons why version management is seen as a problem. The problem is said to possibly get out of hand without action.

The awareness of the problem has already urged the (main) Atlas GIS user (Mr. L. de Jong) to make plans for the construction of a file with (meta)data on the various versions of the main GIS coverages (areal units and urban areas). The shift to a geographical file on the network server is seen as a potential further solution to the version problem. The version administration in a separate metadata file (done 'by hand') is clearly regarded as a temporary solution. This file is seen as containing coverage versions as records and the following fields:

- map layer name
- date of latest change
- version code
- description
- number of elements
- type {point, line, area}
- scale
- suitability for visualising traffic model outcomes
- source
- remarks

Metadata for the areal unit connections is not regarded as necessary, since this data (both features and attributes) is not maintained within Atlas GIS[78]. The maintenance within TRIPS is regarded satisfactory. And typically, GIS is seen as an extension of TRIPS.

The future of GIS metadata is related to broader developments within the provincial organisation, in particular:

- strategic developments regarding GIS
- implementation of the data warehouse concept.

A.2.2. Interview 2: Municipality of Den Haag: Dienst Bouwen en Wonen

On November 14, 1995 a 1¼ hour interview was conducted with Mr. Ing. P. Pronk, project engineer and main Atlas GIS user in the Housing Sector of the Construction and Housing Service (DBW) at the Municipality of Den Haag (The Hague, with 450,000 inhabitants the third largest city in the Netherlands).

The DBW Service will in 1996 merge with the Spatial and Economic Development Service (REO) into the new Urban Development Service (DSO). The 1995 Housing Sector contains two departments: (1) Planning and Finance, and (2)

[78] The volume of logically possible connections (and resulting file size) prevents this.

Development and Programming. Mr. Pronk is involved in research and analysis for planning and programming of housing projects in the second department. He has contacts with many other users of spatial data in the Municipality, and is in fact the founder of a municipal Atlas GIS users group.

Among software tools used by Mr. Pronk are Atlas GIS, SPSS and dBaseIII. Information exchange in the Municipality takes place through a large network (now Novell and Unix, later possibly Windows NT), connected to external networks. Atlas GIS (under Windows 3.x) is operating on a single PC of which Mr. Pronk is the main user. There are five other employees who have secondary access to Atlas GIS functionality on this machine. In the near future Atlas GIS will probably be operating on three PC's in a small LAN. Of all Atlas GIS users Mr. Pronk is the one to supply GIS products to others in and outside DBW. Atlas GIS is applied in close cooperation with Mr. J.G.M. Starmans of the Planning and Programming Unit, who is the supplier of most of the attribute data.

Atlas GIS has been in use by DBW since 1½ year. Before that, use has been made for several years of Atlas Graphics (as a displaying tool). Strange as it now may seem: a 1992 user needs study (while Atlas Graphics was already used) arrived at the conclusion that GIS was not really needed in the local authority. At that point in time there were already plans to build a Base System of Buildings. Following the negative study outcome DBW continued the development of the Base System, allbeit not in GIS. This system contains data for each residential unit: number of rooms, room functions, room areas, house number, and codes for street, neighbourhood, urban district and urban sector; every residential unit has a reference to a REO paper map and will shortly receive (x,y) coordinates. DBW moved to Atlas GIS (Windows) for displaying data at urban district level, while REO purchased Atlas GIS (DOS) for traffic studies.

In REO there is a Department for Survey and Properties, which is responsible for (nowadays digital) municipal topography (DigTop): a system with 50 coverages, maintained by input from various cooperating municipal services. The DigTop geographic data is available for the participating services and for external parties. These facts are important since GIS at DBW leans heavily on geographic features from DigTop and on attribute data from the Base System of Buildings, although many data stream operations were (and are) necessary to satisfy specific GIS data needs. For Atlas GIS at DBW to play a significant role in researching and advising on municipal housing matters, following (vector) coverages are mainly used (arranged from large to small):

- Municipal border.
- Urban sectors (#7).
- Urban districts (#40).
- Neighbourhoods (#94).
- Four-position postcode areas (old[79]).
- Building blocks (3000, plus 1500 broader road sections), derived from:
- Line segments, connecting street junctions.
- (in future:) Building units, streets, water, greenland, beach, etc. etc.

Anno 1995 most attributes from the Base System of Buildings are (after aggregation operations) related to urban districts; in future to neighbourhoods. Other attribute data sources are CBS (the National Statistical Bureau), Social Security, Police, Housing Corporations, and the Employment Register. There are plans for developing (Atlas Script) applications that can be

[79] It is interesting to note that 4-position postcode areas especially in Den Haag have little significance apart from post service planning, since they were based on postman routes without proper additional input.

supplied to significant parties in Den Haag as an interface to feature and attribute data on a specific geographic level (e.g. neighbourhoods).

Documentation of available DBW spatial data is perceived as a problem in the sense that GIS application is now rather dependent on the knowledge of the main user. This main user himself feels a need for information on subjects, locations and reliability of data. A threat is perceived in the contingency that a secondary GIS user pressed in the absence of the primary user would compile a map from the wrong data or using incorrect procedures. In future Atlas GIS will probably be used by more users and in close links with other software, sharing data. For all these reasons a need of GIS metadata is recognised.[80]

Two actions have already been taken:

1. A person has been hired temporarily to produce a manual containing documentation on all GIS operations and applications.
2. Several employees are cooperating to devise standards for more essential municipal data, among which spatial data.

Additional thought is given to a metadata management system. It will answer questions as: What data is there? What can be done with it? What is permitted? How does it work? This metadata will be available to all Atlas GIS users at DBW. More information on the future metadata system is not available yet.

A.2.3. Survey Results from the Atlas GIS Discussion List on the Internet

With the results of these two interviews in mind a questionnaire was devised and posted in the Atlas GIS (AGIS-L) newsgroup on the Internet. This shift from personal interviewing to a posted questionnaire was made for four reasons:

- with personal interviewing an unwishful limitation to the Netherlands was felt,
- personal interviewing seemed not to be the only appropriate data collection technique,
- personal interviewing and the writing of reports took more time than was available,
- the author felt a certain curiosity to explore research opportunities at the Internet.

The AGIS-L discussion list is an Internet listservice offered by a non-profit center company, CIESIN, to the Atlas GIS users (and the rest of the interested) community. Posted messages are distributed to the e-mail boxes of the 490 participants worldwide. The following message was posted on November 27, 1995.

Box A-5: Atlas GIS Metadata Questionnaire, posted in the AGIS-L Discussion List

Hi all,
I am an Msc (GIS) student in the UNIGIS network of European Universities. My thesis' subject is: 'The Company Relevance of GIS Metadata, with special emphasis on the Design of a Metadata System for GIS at the VSN group'.
With metadata is meant 'data about data'. GIS at the VSN group (a large Dutch provider of collective personal transport) is Atlas GIS 3.01. The thesis has both a theoretical and a practical part.

[80] Data maintenance is not perceived as a problem: attribute maintenance is in the hand of other specialists, while geographic features (changes) can be taken from DigTop.

```
This questionnaire was designed to gather information about the metadata (needs) of other
Atlas GIS users, in order to generate ideas for VSN's metadata system.
All Atlas GIS users (in organisations) who respond by fully, seriously and briefly completing
the following questionnaire in English will receive a free (hard) copy of my thesis (in 1996
or 1997). Your answers may be included in an appendix to the thesis, unless you mention
objections. Please keep the questions' text in your reply.
PLEASE E-MAIL your reply to: clebruns@pi.net  (don't post it in AGIS-L).

Thank you very much for your response. You won't regret it...
------------------------------------------------------------------
Clemens van Brunschot                            (clebruns@pi.net)
VSN groep    P.O. Box 19222     NL-3501 AH Utrecht     Netherlands
Tel +31 30 2971800                              Fax +31 30 2938245
------------------------------------------------------------------

Q1.  Name and address of your organisation (don't forget the country)
Q2.  A short description of your organisation's mission, goals and/or objectives
Q3.  Your name
Q4.  Your department, function, and main activities
Q5.  What Atlas GIS version do you use? Is it a network version?
Q6.  Since when does your organisation use Atlas GIS at all?
Q7.  How many Atlas GIS users are working on how many machines?
     Are these machines connected in a LAN?
Q8.  Is there a network connection to the broader organisation?
Q9.  What is GIS used for?
Q10. Describe the geographical features (objects) in your GIS
Q11. What attribute data are in/around your GIS?
Q12. What database and statistical software surrounds your GIS?
Q13. Describe the sources of your main data.
     Do you cooperate with other departments to obtain GIS data?
Q14. Is there metadata/documentation for Atlas GIS applications?
     If yes:  ---> Q15.
     If no:   ---> Q19.
Q15. What is the purpose of this metadata/documentation?
Q16. Describe this metadata/documentation
Q17. What use is made of it?
Q18. What more wishes for metadata/documentation are there?
     (Thank you, please skip Q19.- Q21.)
Q19. Is the lack of metadata/documentation seen as a problem?
Q20. What metadata/documentation would be useful?
Q21. Are metadata/documentation developments started? If yes, please describe.
Q22. Do you agree with including your answers in the appendix to my thesis?
```

Only 7 completed questionnaires were returned. This may be due to the length of the questionnaire, to a low level of interest, or perhaps to the time lag between completion and reward. The (slightly edited) reports received are included below. Only questions completed are mentioned.

```
Q1.  Name and address of your organization (don't forget the country)
North Carolina Division of Water Resources
PO Box 27687
```

Raleigh, NC, USA 27687

Q2. A short description of your organization's mission, goals and/or objectives
The Division of Water Resources (DWR)

Q3. Your name
Reid Campbell, PE

Q14. Is there metadata/documentation for Atlas GIS applications?
If yes: ---> Q15. Yes
If no: ---> Q19.

Q16. Describe this metadata/documentation
DWR metadata efforts are aimed at meeting the US Federal Standard for geospatial data documentation. DWR served as a non-Arc/Info non-Unix site to provide metadata on GIS layers.

Q17. What use is made of it?
Internal to DWR, metadata provides documentation of databases: who did the work, who maintains it, where did the data came from, what fields are in the database, and such. DWR also serves metadata over the internet as part of the NC GIS Clearinghouse. In this manner, others can review the data description to see if it meets their needs.

Q18. What more wishes for metadata/documentation are there?
(Thank you, please skip Q19.-Q21.)
While the items to include in metadata are well established through the US Federal Standard, just how to create metadata can still be a problem. This is especially true for datasets in Atlas*GIS. It would be very helpful to have a script that reviewed the data layers and produced an output file, such as a .dbf file, with basic descriptive information. This could include items like: number of features, geographic extents of features, attribute field names, type, and size. It could be similar to the DOCUMENT AML for Arc/Info.

In addition, metadata should be tied to the actual data. Several files are included in an Atlas*GIS layer (.dbf, .agf., & etc). The metadata would be an additional file, and perhaps an additional table (in .dbf format).

Q1. Name and address of your organisation (don't forget the country)
Bureau of Labor Statistics
2 Massachusetts Ave. N.E. suite 5095
Washington, D.C. 20212
U.S.A.

Q2. A short description of your organisation's mission, goals and/or objectives
Collect housing data for the CPI index (Consumer Price Index).

Q3. Your name
Marcelo Thome Caminha (contractor from Brazil)

Q4. Your department, function, and main activities
U.S. Department of Labor
Bureau of Labor Statistics
Consumer Price and Consumption Studies Division
Housing Branch
Senior computer specialist.
Developing a mapping application in VB script to simplify the making of 50,000 maps for the CPI revision in 1996.

Q5. What Atlas GIS version do you use? Is it a network version?
version 3.01. Not networked.

Q6. Since when does your organisation use Atlas GIS at all?
I think one year (I was not in the project in the beginning).

Q7. How many Atlas GIS users are working on how many machines?
Are these machines connected in a LAN?
Under development: 4
In production there will be: 10
All machines are on a LAN.

Q8. Is there a network connection to the broader organisation?
Yes, our local network connects to the main BLS network.

Q9. What is GIS used for?
To draw maps that will help the field specialists find the houses they are going to visit plus help managers assign work to their staff.

Q10. Describe the geographical features (objects) in your GIS
Mainly blocks, streets, water features, green areas and non-visible boundaries but it includes all typical layers (hospitals, schools...).

Q11. What attribute data are in/around your GIS?
Only street names that may be updated by the field.

Q12. What database and statistical software surrounds your GIS?
We get the sampled blocks that will be mapped from our statistical division. I have no idea what they have in there but it is probably Sybase database with SAS applications.

Q13. Describe the sources of your main data. Do you cooperate with other departments to obtain GIS data?
We bought all GIS data from SMI (Atlas GIS owner). I think the data comes from US census data.

Q14. Is there metadata/documentation for Atlas GIS applications?
Not that I'm aware of.

Q19. Is the lack of metadata/documentation seen as a problem?
Not to me at least. Huzil_J is the geographer in the project, she may give you more information on her answer.

Q20. What metadata/documentation would be useful?

Not apply.

Q21. Are metadata/documentation developments started? If yes, please
     describe.
Don't know.

Q1. Name and address of your organisation (don't forget the country)
Global Mapping International
7899 Lexington Dr., Suite 200 A
Colorado Springs, CO 80920

Q2. A short description of your organisation's mission, goals and/or objectives
Our desire it to see the worldwide Christian community better able to make informed ministry decisions. To this end, we seek to enable ministry decision makers to acquire, manage, analyze, apply, communicate, and share strategic information. To do this we develop and supply global databases to provide foundational geographic and statistical information and provide technical consultation and training.

Q3. Your name
Loren Muehlius

Q4. Your department, function, and main activities
Computer Cartographer. Provide technical support for users of Atlas GIS software. Assist with development of a detailed computerized map of the world for use with Atlas GIS. Answer mapping related questions from Christian workers.

Q5. What Atlas GIS version do you use? Is it a network version?
3.01 - Not a network version

Q6. Since when does your organisation use Atlas GIS at all?
Feb. 1990

Q7. How many Atlas GIS users are working on how many machines?
     Are these machines connected in a LAN?
Five on five machines. Three are on a LAN

Q8. Is there a network connection to the broader organisation?
Yes

Q9. What is GIS used for?
Making maps of statistics, paper and overhead transparencies. We are developing detailed geo files to serve as base maps for other users.

Q10. Describe the geographical features (objects) in your GIS
Points:
Cities, towns
Languages
Lines:
Geo-political borders, international and 1st level administrative borders.

Major roads, railroads, rivers, coastlines
Polygons:
1st level and some 2nd level administrative borders.
Indigenous languages

Q11. What attribute data are in/around your GIS?
Population, religious, linguistic, social and economic

Q12. What database and statistical software surrounds your GIS?
Foxpro

Q13. Describe the sources of your main data. Do you cooperate with other departments to obtain GIS data?
World Data Bank II
Digital Chart of the World
Rand McNally
Summer Institute of Linguistics

Q14. Is there metadata/documentation for Atlas GIS applications?
If yes: ---> Q15.
If no: ---> Q19.
A little (Yes)

Q15. What is the purpose of this metadata/documentation?
Show status of our detailed computerized map (the level of detail of administrative border for each country - international border only, 1st or 2nd level of administrative border; progress in mapping languages of country), define fields in attribute tables

Q16. Describe this metadata/documentation
Status map; dBase files of 1st level administrative names, languages, list of finished project files, project file descriptions, codetable of various international standard country codes (ISO 3166, FIPS 10-3, etc.), descriptions of fields in .COL files; text files describing fields in attribute tables

Q17. What use is made of it?
Included with documentation of product so users will know data included with product

Q18. What more wishes for metadata/documentation are there?
It would be nice if project files would display attribute tables used and date they were last updated. Could Atlas GIS be given a feature of making an Index from Project files and contain the above information?

Q1. Name and address of your organisation (don't forget the country)
REMOTE SENSING RESEARCH GROUP, DEP. OF GEOGRAPHY, UNIV. OF BERN, SWITZERLAND

Q2. A short description of your organisation's mission, goals and/or objectives
SUPPORTING REST OF DEPARTMENT
NOAA/AVHRR & METEOSAT RECEIVING STATION
PROJECTS ABOUT REMOTE SENSING

Q3. Your name

BEAT BURRI

RUOPIGENRING 43

CH-6015 REUSSBUEHL

SWITZERLAND

Q4. Your department, function, and main activities

REMOTE SENSING RESEARCH GROUP, STUDENT, THESIS "MAPPING THE SNOW COVER IN THE ALPS WITH NOAA/AVHRR DATA"

Q5. What Atlas GIS version do you use? Is it a network version?

NOT NETWORK

ATLAS WIN 3.0

ATLAS DOS 1.2 FOR DIGITIZING AND EDITING MAP PROJECTIONS

Q6. Since when does your organisation use Atlas GIS at all?

SINCE ABOUT 1988

Q7. How many Atlas GIS users are working on how many machines?
Are these machines connected in a LAN?

YES, CONNECTED IN A LAN

IN THE REMOTE SENSING RESEARCH GROUP: 2 MACH, 1 USER

IN THE DEP. OF GEOGRAPHY: CA. 2 MACH., CA. 2 USERS (?)

Q8. Is there a network connection to the broader organisation?

YES, ALL UNIVERSITIES

Q9. What is GIS used for?

MAPPING THE SNOW COVER VARIATIONS IN THE ALPS

Q10. Describe the geographical features (objects) in your GIS

RIVERS, LAKES, BASIN, STREAMGAUGE, CLIMATE STATIONS, ELEVATION ZONES

SNOW COVER ON DIFFERENT DAYS

Q11. What attribute data are in/around your GIS?

CLIMATE CATA, STREAMGAUGE DATA

PERCENT CALCULATIONS OF SNOW COVER PER ELEVATION ZONE

Q12. What database and statistical software surrounds your GIS?

DBASE DOS

DBASE WIN 5.5

SNOW RUNOFF MODEL

Q13. Describe the sources of your main data. Do you cooperate with other departments to obtain GIS data?

HYDROLOGIC ORGANISATIONS

HYDROPOWER COMPANIES

SWISS METEOROLOGICAL AGENCY (SMA)

Q14. Is there metadata/documentation for Atlas GIS applications?
If yes: ---> Q15.
If no: ---> Q19.

YES

Q15. What is the purpose of this metadata/documentation?
ADDITIONAL INFORMATION
DBASE PATH -> LINK

Q16. Describe this metadata/documentation
QUANTITIES OF WATER STORAGE IN STORAGE LAKES
GENERAL INFORMATION
OTHER COLUMNS

Q17. What use is made of it?
PRESENTATION
DBASE PATH
INFORMATION

Q18. What more wishes for metadata/documentation are there?
(Thank you, please skip Q19.-Q21.)
NONE

Q1. Name and address of your organisation (don't forget the country)
Quebec Ministry of the Environment and Wildlife
Aquatic Ecosystem Branch
930, Ste.Foy Road, 2nd floor
Quebec, Quebec, G1S 2L4
Cabada
Phone: 418-644-3297 Fax: 418-646-8483 email: jdupont@mediom.qc.ca

Q2. A short description of your organisation's mission, goals and/or objectives
The main goal of the Quebec Ministry of the Environment and Wildlife is to
protect and restore the environment, apply regulations and programs to
minimize adverse effects of man on the ecosystem. The goal of my division is
to monitor the quality of the aquatic ecosystems (acid rain, toxics, water
quality, biological diversity, etc.) and predict the outcome of future
actions and evaluate the benefits of actual actions.

Q3. Your name
Jacques Dupont

Q4. Your department, function, and main activities
Aquatic Ecosystem Branch
Water Science Analyst and Numeric technology specialist
GIS and model development and applications

Q5. What Atlas GIS version do you use? Is it a network version?
Atlas GIS 3.0 (no network)

Q6. Since when does your organisation use Atlas GIS at all?
We had Atlas Gis since last January but we had Atlas Pro since 1993 before
that. We also use SPANS GIS for OS/2 5.31 and IDRISI for Windows.

Q7. How many Atlas GIS users are working on how many machines?
Are these machines connected in a LAN?
Two atlas users are working on two machines.

Q8. Is there a network connection to the broader organisation?
Not actually but there will be one next year.

Q9. What is GIS used for?
The GIS are used to create thematic maps of surface water quality (point and vector data) with presentation of trend series of water quality. The GIS are also used to do spatial analysis in order to relate watershed (altitude, geology, pedology, landuse, etc.) and anthropogenic characteristics (agriculture, industries, etc.) to water quality.

Q10. Describe the geographical features (objects) in your GIS
There are maps at scales ranging from 1:20000 to 1:2000000. We have also administrative maps (counties), agriculture data, pedology data, water quality data (point data), topographic maps, eoclogical maps, etc.

Q11. What attribute data are in/around your GIS?
Water quality, physical attributes of the sampling station, toponomy, etc.

Q12. What database and statistical software surrounds your GIS?
We are mostly working with SAS for Windows 6.10 and Microsoft Access 2.0. We also work with Atlas Script/VB, Delphi, Excel, DB2 and Visual Basic. We will soon look at Oracle.

Q13. Describe the sources of your main data. Do you cooperate with other departments to obtain GIS data?
The water quality data are gathered by us and all spatial maps are from other departments

Q14. Is there metadata/documentation for Atlas GIS applications?
If yes: ---> Q15.
If no: ---> Q19.
Not actually

Q19. Is the lack of metadata/documentation seen as a problem?
Not much but we have some problems with the analytical methods used for the laboratory analysis. If there were metadata on these we would have less difficulty interpreting the results and discerning what data was measured with what method.

Q20. What metadata/documentation would be useful?
It could be.

Q21. Are metadata/documentation developments started? If yes, please describe.
Not yet, but we will look into it in the near future.

Q1. Name and address of your organisation (don't forget the country)
Center for Urban Policy and the Environment
342 N Senate Ave
Indianapolis, IN 46256
USA

Q2. A short description of your organisation's mission, goals and/or

objectives
Mission: To work with state and local governments and their associations, neighborhood and community organizations, community leaders, and business and civic organizations in Indiana to identify issues, analyze options, and develop the capacity to respond to challenges. Goal: To empower local citizenry and community leaders to enhance the quality of their lives and the decisions that affect them.

Q3. Your name
Matthew M. Rummel

Q4. Your department, function, and main activities
Research Assistant, GIS consultant and primary user

Q5. What Atlas GIS version do you use? Is it a network version?
DOS, 2.1 (soon to be Windows 3.0) yes it is a network

Q6. Since when does your organisation use Atlas GIS at all?
1991-92

Q7. How many Atlas GIS users are working on how many machines?
Are these machines connected in a LAN?
2-3 users, 2-3 machines, not a LAN

Q8. Is there a network connection to the broader organisation?
yes

Q9. What is GIS used for?
census data analysis, creating maps for reports, land use analysis, parcel information

Q10. Describe the geographical features (objects) in your GIS
????

Q11. What attribute data are in/around your GIS?
Census data, user-defined data

Q12. What database and statistical software surrounds your GIS?
??????

Q13. Describe the sources of your main data. Do you cooperate with other departments to obtain GIS data?
Census data, plat maps, zoning maps, utility maps. Yes, we cooperate.

Q14. Is there metadata/documentation for Atlas GIS applications?
If yes: ---> Q15.
If no: ---> Q19.
no

Q19. Is the lack of metadata/documentation seen as a problem?
not yet

Q20. What metadata/documentation would be useful?
where to find specific data, large source bank of data

Q21. Are metadata/documentation developments started? If yes, please describe.
No

Q1. Name and address of your organisation (don't forget the country)
Department of Geography
PH Erfurt
P.O.Box 307
99006 Erfurt, Germany

Q2. A short description of your organisation's mission, goals and/or objectives
Geographical research and education

Q3. Your name
Thomas Ott, Diplom-Geograph

Q4. Your department, function, and main activities
Human geography (population and urban geography) research

Q5. What Atlas GIS version do you use? Is it a network version?
Windows 3.0

Q6. Since when does your organisation use Atlas GIS at all?
September 1993

Q7. How many Atlas GIS users are working on how many machines?
Are these machines connected in a LAN?
3 users, LAN

Q8. Is there a network connection to the broader organisation?
yes

Q9. What is GIS used for?
mapping, spatial analysis

Q10. Describe the geographical features (objects) in your GIS
urban districts, administrative and statistic regions

Q11. What attribute data are in/around your GIS?
census data, poll data

Q12. What database and statistical software surrounds your GIS?
Q&A (Symantec), SPSS

Q13. Describe the sources of your main data. Do you cooperate with other departments to obtain GIS data?
governmental and urban offices, own studies, yes we cooperate

Q14. Is there metadata/documentation for Atlas GIS applications?
If yes: ---> Q15.
If no: ---> Q19.
yes, sometimes

Q15. What is the purpose of this metadata/documentation?
enable future users to understand purpose and "quality" of data

Q16. Describe this metadata/documentation
description of source and possible faults

Q17. What use is made of it?
no use yet

Q18. What more wishes for metadata/documentation are there?
(Thank you, please skip Q19.-Q21.)
none

A.3. Details of System Specification

This appendix contains the complete system specification of the desired VSN metadata management system (MMS). Essentials and relevant considerations have been entered in Section 5.5. The MMS is intended to operate in conjunction with the VSN GIS software: Atlas GIS (described in Sections 2.1 and 5.2).

A.3.1. Processes

In this section a process is an activity (or: transformation of inputs) in GIS use. The word 'activity' will in this appendix be used in two ways:

1. in a general sense, describing all the processes that will make up the MMS, including processes that surround the MMS,
2. and in a narrow sense, describing GIS functionality and - therefore - a type of entity in the MMS conceptual model.

The description of processes, now to follow, will be used as a basis for all of the MMS design. This description is based on five years of personal experience with Atlas GIS and on the study reflected in the first part of this thesis. The method of picturing the structure of these processes in an information flow diagram was adapted[81] from de Backer (1982).

- A circle stands for a process (or: function). It has a unique identifier (1, 2, ...).
- A square stand for an external[82] source or destination of data, information or hardware. It has a unique identifier (a, b, ...).
- An right-opened rectancle stands for a stored set of entities. Is has a unique identifier (S1, S2, ...).
- An arrow stands for a flow of data/information (input and/or output). Each flow is described by the essence of its content. Input from and output to stored sets need no further description.

Following data flow diagram describes the working of the Metadata Management System needed by the VSN group for their GIS. It can be seen as an elaboration of the (provisional) conceptual model presented in Section 5.2. Yet, the level of detail has been kept low, to further the value of the data flow diagram. Elements of the intended MMS have been described in bold typeface and included within a dotted line: this holds for both functions and in- and output.

Some remarks need to be made:

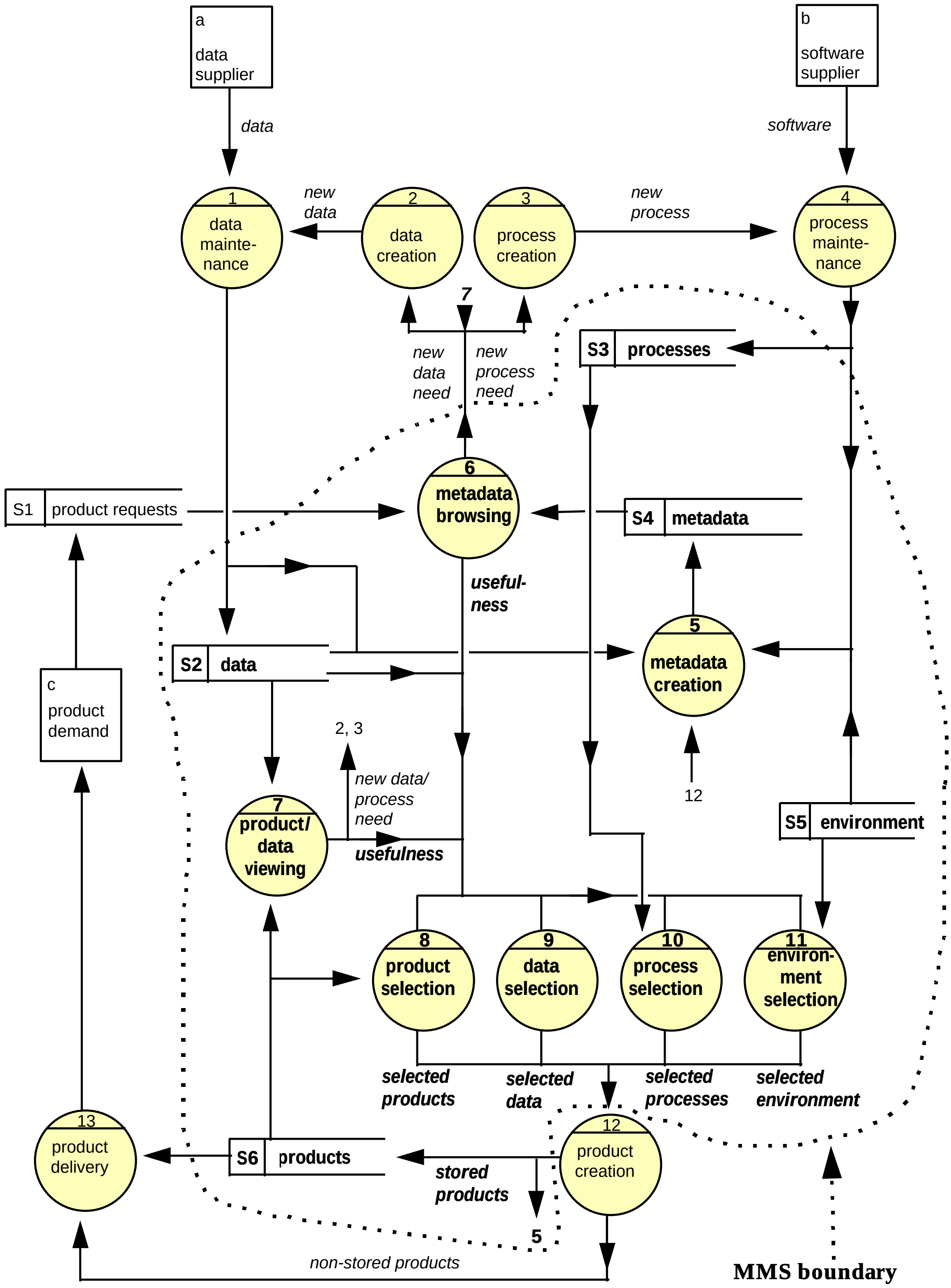
1. Four types of meta-entities can be distinguished:
    - metadata about data
    - metadata about processes
    - metadata about the environment (originators, users, hardware)
    - metadata about (previous) GIS products (maps, analyses, files)
2. Hardware environment is seen as relatively static compared to the other meta-entities.
3. Selection of products, data, processes and environment show interdepence. In fact, products and data may in some cases

[81] The adaptatations are in:
- information and data rather than only data;
- the physical location of each function has been omitted, since our system will show no complexity in this regard;
- arrows have been combined where clarity was enhanced by such;
- we speak of 'stored sets' rather than 'stored data', to include physical entities and information.

[82] External to the MMS and GIS.

be hard to separate.

Box A-6: Information Flow Diagram



4. A boundary has been drawn around what is seen as the MMS, which entails management of metadata and preparations for product creation.
5. The MMS itself does *not* entail:
   - data searching otherwise than through its metadata (e.g. by queries on coordinates or attribute values);
   - exporting of data to other (e.g. transfer) formats.
6. Neither are system maintenance and data auditing (Cornelius 1991) entailed in the MMS's functionality. However: these important activities will be greatly facilitated by the intended system.

Box A-7 briefly describes the MMS process elements and their structure.

Box A-7: MMS process description

process name	**Metadata creation**
reference	5
input	Data from data maintenance and from stored dataset. Created (stored) products. Processes from process maintenance. Environment.
output	Metadata (to be stored).
activities	Definition (and storage) of metadata. Partly actively by the user, partly automated..
circumstances	Must not slow down production, otherwise the total MMS will tend to be neglected.
frequency	1-10 Times per week.
acceptable lead time	2-5 Minutes for each (data, product, process, environment) set.
present situation	Only fragmented notes created.

process name	**Metadata browsing.**
reference	6
input	Product requests (1 or more). Metadata.
output	Assessed usefulness of existing data, previous products, GIS processes and environmental elements, for the creation of a requested product. Assessed need of new data and new processes.
activities	Determining of and which (existing or new) GIS elements will facilitate the creation of the desired product(s).
circumstances	A user may not know his/her way around in the data, products, GIS processes and GIS environment. Metadata browsing is meant to speed up the production time for each user, including the experienced one.
frequency	1-10 Times per week.
acceptable lead time	5-10 Minutes, depending on the user and the complexity of the problem.
present situation	No metadata available except some fragmented notes and a number of printed maps. No MMS available.

process name	**Product and data viewing**
reference	7
input	Stored data. Stored products.
output	Assessed usefulness of stored data and stored products, for the creation of a desired product. Assessed need of new data and new processes.
activities	Having a realistic view on how existing data are structured, how previously produced maps look, and what previous analyses yield.
circumstances	A user may not know his/her way around in the data, products, GIS processes and GIS environment.
frequency	1-10 Times per week.
acceptable lead time	1-5 Minutes.
present situation	Already done, using Atlas GIS/Excel and non-Windows DBMS functionality. Particularly the non-Windows FoxPro software is a drawback here, since it requires the closing of Windows operations.

process name	**Product selection**
reference	8
input	A stored product, after the assessment of its usefulness (by metadata browsing and product/data viewing) for new product creation.
output	Product selected for GIS processing.
activities	Picking one or more previously produced (and stored) products and making them available for GI S (if so desired).
circumstances	Most of the time there is a stored map (project file) or analysis results (fields from attribute tables) that will be helpful in creating a new product.
frequency	1-10 Times per week.
acceptable lead time	< 1 Minute.
present situation	Now integrated in product/data viewing, because using Atlas GIS a viewed map (project file) will be automatically selected.

process name	**Data selection**
reference	9
input	Stored data, after the assessment of its usefulness (by metadata browsing and product/data viewing) for new product creation.
output	Data selected for GIS processing.
activities	Picking one or more items from previously created (and stored) data and making them available for GIS (if so desired).
circumstances	Most of the time there is stored data (fields from attribute tables) that will be helpful in creating a new product.
frequency	1-10 Times per week.
acceptable lead time	1-2 Minutes.
present situation	Now partly integrated in product/data viewing, because using Atlas GIS a viewed map makes related attribute tablers automatically available.

process name	**Process selection**
reference	10
input	Stored processes, after the assessment of their usefulness (by metadata browsing and perhaps product/data viewing) for new product creation.

output	Processes selected for GIS processing.
activities	Picking one or more items from previously created (and stored) GIS and DBMS processes and making them available for GIS (if so desired).
circumstances	Most of the time there will be stored processes (syntax in various software packages, or simply instructions how do do things) that will be helpful in creating a new product.
frequency	1-5 Times per week.
acceptable lead time	1-5 Minutes.
present situation	Sometimes already there is syntax or instructions that facilitate new product creation.

process name	**Environment selection**
reference	11
input	Existing environmental elements: users, system (documentation) and devices, after the assessment of their usefulness (by metadata browsing) for new product creation.
output	Environmental elements selected for GIS processing.
activities	(Partly automatically) picking environmental elements and making them available for GIS (obligatory).
circumstances	The environment is there, but needs to be fitted to the particular application of GIS.
frequency	1-10 Times per week.
acceptable lead time	< 1 Minute.
present situation	No recording of user. Standard system documentation. Printer selection in standard Windows screens (however: switching to local colour printer currently requires the closing of Windows operations.

A.3.2. Entities, Attributes and Relationships

The full conceptual model will br presented grossly in compliance with the demands of the Third Normal Form[83] of the well-known relational model. The conceptual model is not an end in itself: it serves in bringing aims and objectives (see Section 5.3) closer to reality, or vice versa. The objectives call for a conceptual model (and, of course, the resulting system) to help in the creation of new GIS products and maintaining the GIS system and the eventual migration to a new GIS system. Throughout, use will be made of Reeve (1993) and a text describing the Oracle CASE*METHOD: Barker (1990). In fact, the ER diagram follows the conventions of this method.

The contents of the ER diagram can grossly be divided into three parts:

1. the upper half refers to metadata on data and products;
2. the lower left quadrant refers to metadata on processes;
3. the lower right quadrant refers to metadata on the environment.

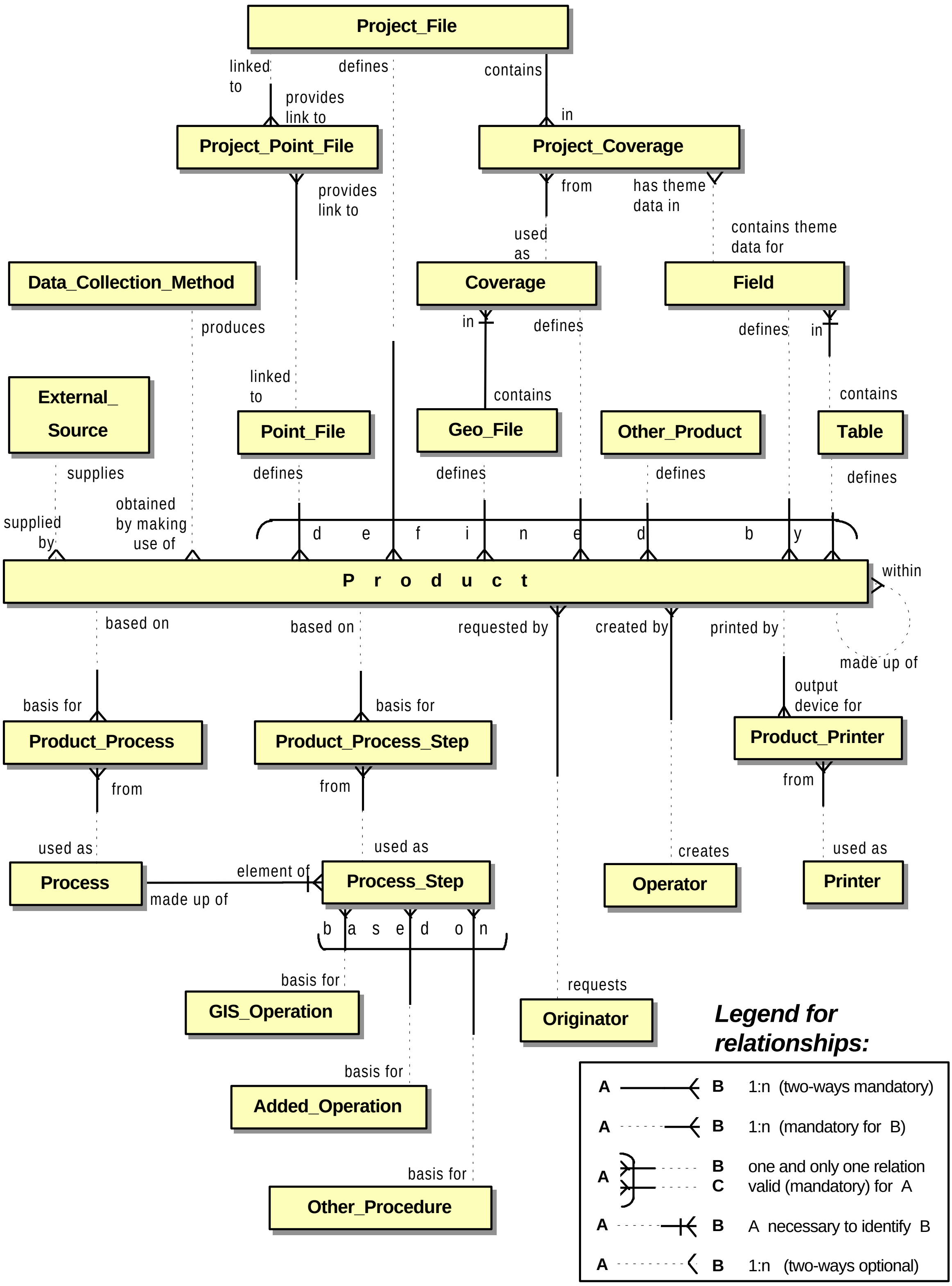
A central position is taken by the entity ‘Product’. A Product may consist of several elemental Products. Each Product must have one link to one and only one of 7 information/data entities (a file - including a map - , an element of a file - e.g. a coverage or a field - , or another Product). In fact, products and data come close to one another, more so than suggested in Box A-6 (Information Flow Diagram). The link between data and its source, data collection method, lineage, originator,

[83] Minor deviations can be found in date attributes pertaining to the validity of other attributes in the same table.

operator, and output medium is achieved through the central Product entity[84].

[84] One of the reasons for this was the wish to keep the design straightforward, and therefore not too expensive to build.

Box A-8: Entity-Relationship Diagram for the VSN GIS Metadata Management System



Now a list of all intended attributes in each of the entities (tables) of the RDBMS will be presented. The so-called 'intersection' tables will be treated last. Because this system will be a metadata system, its (meta)data volumes will not be extensive. Therefore there has not been any hesitation to use character strings as keys. Change with time is dealt with in relation to data/products and metadata itself, but not in relation to originators and operators. Thus, if an originator's profile or operator's profile changes, the system will be able to maintain corrects links to the old profiles only by retaining the old profiles and entering a new operator or originator with a new profile. Such is possible since those primary keys have been set to ID's rather than names.

There is some amount of repetition of data values (e.g. the attribute Map Projection in Geo_File, with a default). This is not considered to be a problem since the number of tuples (cardinality) will be low.

Box A-9: Attribute Listing, by Entity[85]

Entity / Attribute Name	Notes / Definitions	Format	Key	NotNull	Unique	Mandatory/ Optional	Domain	Default	In CEN/TC 287 metadata standard
Project_File									
Name	Unique (within MMS) name of the file, incl. extension	Char 12	Primary	NotNull	Unique	Mandatory			'Dataset title'
Description		Char 40				Optional			
File location	Full path of the file	Char 40				Mandatory			
Date of latest change	Assigned by system	Date				Mandatory	ó Today		
Coverage									
<u>Name</u> Name of Geo_File	A coverage is a map layer in Atlas GIS. It is always part of a Geo_File	<u>Char 15</u> Char 12	<u>Primary</u> Foreign	<u>NotNull</u> NotNull	} Unique }	<u>Mandatory</u> Mandatory			'Feature type name'
Description	Description of the features in the coverage, or reference to a standard definition	Char 40				Optional			'Feature type definition'
Number of features	Exact number of coverage features	Num 5.0				Mandatory	> 0; integer		'Occurrences'
Feature type		Char 6				Mandatory	{Point, Line, Region}		'Geometric primitive'
Extent date	Date at which the status and description of the planar and temporal extent is valid	Date				Mandatory	ó Today		'Extent date'
Extent status	Degree to which the coverage meets its proposed final geographic coverage; any geographic variation in classification	Char 40				Mandatory			'Extent status'
Min X	The minimum x coordinate of the coverage planar extent	Num 10.0				Optional			'Bounding XY'
Max X	The maximum x coordinate of the coverage planar extent	Num 10.0				Optional			'Bounding XY'
Min Y	The minimum y coordinate of the coverage planar extent	Num 10.0				Optional			'Bounding XY'
Max Y	The maximum y coordinate of the coverage planar extent	Num 10.0				Optional			'Bounding XY'
From date	The earliest date to which the coverage corresponds	Date				Optional	ó Today		'From date'
To date	The latest date to which the coverage corresponds	Date				Optional	ó Today		'To date'
Positional accuracy	Accuracy of geographical position within the dataset. It may be measured by the accuracy relative to a reference system, between close points, etc.	Char 40				Optional			'Positional accuracy'
Completeness	Expressed by the difference between the coverage and its specifications	Char 40				Optional			'Completeness'
Process history	A summary of the processes which the coverage has gone through	Char 320				Mandatory			'Process history'

[85] The exact naming of the attributes is dependent upon the software used, and needs further elaboration.

Entity / Attribute Name	Notes / Definitions	Format	Key	NotNull	Unique	Mandatory/ Optional	Domain	Default	In CEN/TC 287 metadata standard
Geo_File									
Name	Unique (within MMS) name of the file, incl. extension	Char 12	Primary	NotNull	Unique	Mandatory			'Dataset title'
Description		Char 40				Optional			
Number of coverages		Num 3.0				Mandatory	> 0 ; < 1000 ; integer		
Spatial reference year	Reference year for ellipsoid and map projection	Char 10				Optional	ó Today		'Spatial reference datum'
Ellipsoid	The reference ellipsoid	Char 40				Mandatory		Bessel 1842	'Ellipsoid'
Map projection and coordinate system		Char 80				Mandatory		No project-ion; rectang-ular grid (Rijksdrie-hoekmeting)	'Map projection'
File location	Full path of the file	Char 40				Mandatory			
Date of latest change	Assigned by system	Date				Mandatory	ó Today		
Backup	Information about available file backup	Char 40				Optional			
Field (i.e.: *thematic* field)									
Name Name of Table	A field name is inherently linked to one database table	Char 10 Char 12	Primary Foreign	NotNull NotNull	} Unique }	Mandatory Mandatory			'Attribute type name'
Description	Description of the field, or reference to a standard definition	Char 40				Optional			'Attribute type definition'
Format	A combined description of field data type / length / decimals, e.g.: F7.2 ; A12 ; Boolean	Char 10				Optional			
Thematic accuracy	The correctness of the classification of tuples (records) by the field	Char 40				Mandatory			'Thematic accuracy'
Temporal validity	The currency of the data in the field	Char 40				Mandatory			'Temporal accuracy'
Table									
Name	Unique (within MMS) name of the file, incl. extension	Char 12	Primary	NotNull	Unique	Mandatory			'Dataset title'
Description		Char 40				Optional			
Number of records	Number of tuples in the Table	Num 8.0				Mandatory	> 0 ; integer		'Occurrences'
Overall thematic accuracy	A summary of the correctness of the classification of attributes within the Table	Char 40				Optional			'Overall thematic accuracy'
Overall temporal validity	A summary of the closeness of temporal values of attributes within the Table to values accepted to be true	Char 40				Optional			'Overall temporal accuracy'
Overall completeness	Expressed by the gross difference between the attribute values in the Table as compared to its specifications	Char 40				Optional			'Overall completeness'
File location	Full path of the file	Char 40				Mandatory			
Process history	A summary of the processes which the table has gone through	Memo				Mandatory			'Process history'
Software used	Types and version of software used in deriving table elements	Char 20				Optional			
Date of latest change	Assigned by system	Date				Mandatory	ó Today		
Backup	Information about available file backup	Char 40				Optional			
Key columns	Names of Table fields that may be used as as key columns to create links with Coverages, plus a description of the indirect spatial reference system in which the references to a position are given, e.g. Municipality, 4-position Postcode, Ad hoc region, etc.	Char 40				Optional			'Type of indirect spatial reference system', 'Name of areal unit'

Entity / Attribute Name	Notes / Definitions	Format	Key	NotNull	Unique	Mandatory/ Optional	Domain	Default	In CEN/TC 287 metadata standard
Point_File									
Name	Name of the point file, inclusing extension	Char 12	Primary	NotNull	Unique	Mandatory			'Datset title'
Description	Description of the features in the point file, or reference to a standard definition	Char 40				Optional			
Number of points	Exact number of points with valid coordinates	Num 8.0				Mandatory	> 0; integer		'Occurrences'
Extent date	Date at which the status and description of the planar and temporal extent is valid	Date				Mandatory	ó Today		'Extent date'
Extent status	Degree to which the point file meets its proposed final geographic point file; any geographic variation in classification	Char 40				Mandatory			'Extent status'
Min X	The minimum x coordinate of the point file planar extent	Num 10.0				Optional			'Bounding XY'
Max X	The maximum x coordinate of the point file planar extent	Num 10.0				Optional			'Bounding XY'
Min Y	The minimum y coordinate of the point file planar extent	Num 10.0				Optional			'Bounding XY'
Max Y	The maximum y coordinate of the point file planar extent	Num 10.0				Optional			'Bounding XY'
From date	The earliest date to which the point file corresponds	Date				Optional	ó Today		'From date'
To date	The latest date to which the point file corresponds	Date				Optional	ó Today		'To date'
Positional accuracy	Accuracy of geographical position within the dataset. It may be measured by the accuracy relative to a reference system, between close points, etc.	Char 40				Optional			'Positional accuracy'
Completeness	Expressed by the difference between the point file and its specifications	Char 40				Optional			'Completeness'
Process history	A summary of the processes which the point file has gone through	Memo				Mandatory			'Process history'
Spatial reference year	Reference year for ellipsoid and map projection	Char 10				Optional	ó Today		'Spatial reference datum'
Ellipsoid	The reference ellipsoid	Char 40				Mandatory		Bessel 1842	'Ellipsoid'
Map projection and coordinate system		Char 40				Mandatory		No project-ion; rectang-ular grid (Rijksdrie-hoekmeting)	'Map projection'
File location	Full path of the file	Char 40				Mandatory			
Date of latest change	Assigned by system	Date				Mandatory	ó Today		
Backup	Information about available file backup	Char 40				Optional			
Other_Product									
Name	Unique name of the product	Char 12	Primary	NotNull	Unique	Mandatory			'Dataset title'
Description		Char 40				Optional			
Copy type		Char 10				Mandatory			
Completeness	Expressed by the difference between the product and its specifications	Char 40				Optional			'Completeness'
Process history	A summary of the processes which the product has gone through	Memo				Mandatory			'Process history'
File location	Full path of the file	Char 40				Mandatory			
Software used	Types and version of software used in deriving table elements	Char 20				Optional			
Product									
Product ID	Unique ID-code, assigned by the system	Num 6.0	Primary	NotNull	Unique	Mandatoty			
Name of (exclusively) Project_File Point_File Geo_Fille Coverage Table Field Other_Product	Crucial: the system ***must*** guard that one and only one of these source entities is represented in a Product entity record. Through these foreign keys a view is offered to all attributes in the source entity. This needs to be worked out, since the attrubutes differ between source entities. List ALL possible attributes?	 Char 12 Char 12 Char 12 Char 15 Char 12 Char 10 Char 12	Foreign	NotNull (1)	Unique	Mandatory (1)			

Entity / Attribute Name	Notes / Definitions	Format	Key	NotNull	Unique	Mandatory/ Optional	Domain	Default	In CEN/TC 287 metadata standard
Name of Data_Collection_ Method		Char 20	Foreign			Optional			
Name of External_Source		Char 20	Foreign			Optional			
Name of Product_Process		Char 15	Foreign			Optional			
Name of Product_Process_ Step		Char 15	Foreign			Optional			
ID of Originator		Num 3.0	Foreign			Mandatory			
ID of Operator		Num 3.0	Foreign			Mandatory			
Name of Product_Printer		Char 15	Foreign			Optional			
Parent Product ID	Product of which the present Product is an element	Num 6.0	Foreign			Optional			
Purpose	A summary of the intentions with wich the product was developed	Char 40				Mandatory			'Purpose'
Production date	Date at which the product was created. Automatically assigned.	Date				Mandatory	ó Today	Today	
Saving date	Date after which a product may be destroyed.	Date							
Metadata entry	Answer to the question if product metadata were entered manually on production date	Logical				Mandatory	{T,F}		'Metadata entry date'
Metadata check	Answer to the question if existing product metadata were checked on production date	Logical				Mandatory	{T,F}	F	'Last check date'
Metadata update	Answer to the question if existing product metadata were updated manually on production date	Logical				Mandatory	{T,F}	F	'Last update date'
Data_Collection_Method									
Name	Unique name of the data collection method	Char 15	Primary	NotNull	Unique	Mandatory			
Description	A characterisation of the data collection method	Char 40							
External_Source									
Name	Unique name of the external data source	Char 15	Primary	NotNull	Unique	Mandatory			
Description	A characterisation of the external data source	Char 40				Optional			
Address information	Address of the external data source	Char 40				Optional			
Contact information	Contac person(s) at the external data source organisatuon	Char 40				Optional			
Process									
Name	Unique name of the process	Char 15	Primary	NotNull	Unique	Mandatory			
Description	A summary of what happens in a Process	Char 40				Optional			
Purpose	The reason for utilizing the Process	Char 40				Optional			
Process_Step									
Name Name of Process	A Process_Step name is inherently linked to one Process	Char 15 Char 15	Primary Forign	NotNull NotNull	} Unique }	Mandatory Mandatory			
Name of (exclusively) GIS_Operation, Added_Application, Other_Procedure	Crucial: the system ***must*** guard that one and only one of these elementary process entities is represented in a Process_Step entity record.	Char 15	Foreign	NotNull	Unique (1)	Mandatory (1)			
Description	A summary of what happens in a Process_Step	Char 40				Optional			
Purpose	The reason for utilizing the Process_Step	Char 40				Optional			
Requirements	Hardware and software requirements of the Process_Step	Char 40				Optional			
GIS_Operation									
Name	Name of the GIS_Operation	Char 15	Primary	NotNull	Unique	Mandatory			
Description	A summary of what a GIS_Operation does and how it should be used	Memo				Optional			

Entity / Attribute Name	Notes / Definitions	Format	Key	NotNull	Unique	Mandatory/ Optional	Domain	Default	In CEN/TC 287 metadata standard
Source	The person or organisation which represents the origin of the GIS_Operation	Char 40				Optional			
Software	The software with which to apply the GIS_Operation	Char 20				Mandatory			
Location	Physical location of the GIS_Operation	Char 20				Mandatory			
Added_Application									
Name	Name of the Added_Application	Char 15	Primary	NotNull	Unique	Mandatory			
Description	A summary of what a Added_Application does and how it should be used	Memo				Optional			
Source	The person or organisation which represents the origin of the Added_Application	Char 40				Optional			
Software	The software with which to apply the Added_Application	Char 20				Mandatory			
Location	Physical location of the Added_Application	Char 20				Mandatory			
Other_Procedure									
Name	Name of the Other_Procedure	Char 15	Primary	NotNull	Unique	Mandatory			
Description	A summary of what a Other_Procedure does and how it should be used	Char 80				Optional			
Source	The person or organisation which represents the origin of the Other_Procedure	Char 40				Optional			
Software	The software with which to apply the Other_Procedure	Char 20				Mandatory			
Location	Physical location of the Other_Procedure	Char 20				Mandatory			
Printer									
Name	Printer name (e.g. type)	Char 12	Primary		Unique	Mandatory			
Description	Details of printer, if not covered by other fields	Char 30				Optional			
Port	Computer port to whcih the printer is attached	Char 10				Mandatory			
Quality information	General quality information about the printer	Char 20				Optional			
Network address	If applicable, network address of the printer	Char 12				Optional			
Media information	Enumeration of the media that the printer can handle	Char 20				Optional	{paper; trans-parency}		
Print size information	Print sizes that the printer can handle	Char 20				Optional			
Operator									
Operator ID	Assigned by the system	Num 3.0	Primary		Unique	Mandatory			
Name	Name of the operator	Char 20				Mandatory			
Profile	A description of experience and authorities of the operator	Char 40				Optional			
Originator									
Operator ID	Assigned by the system	Num 3.0	Primary		Unique	Mandatory			
Name	Name of the originator/principal of the product	Char 20				Mandatory			
Profile	A description of general requirements of the originator/principal	Char 40				Optional			
Project_Point_File									
Name of Project_File		Char 12	Foreign			Mandatory			
Name of Point_File		Char 12	Foreign			Mandatory			
Project_Coverage									
Project_Coverage ID	Assigned by the system	Char 6.0	Primary		Unique	Mandatory			
Name of Project_File		Char 12	Foreign			Mandatory			
Name of Coverage		Char 15	Foreign			Mandatory			

Entity / Attribute Name	Notes / Definitions	Format	Key	NotNull	Unique	Mandatory/ Optional	Domain	Default	In CEN/TC 287 metadata standard
Name of Field + Name of Table	The Field (from Table ...) that contains thematic data which is actually used in a Project_Coverage	Char 10 Char 12	Foreign		} Unique }	Optional			
Language		Char 15				Optional			
Product_Process									
Product ID		Num 6.0	Foreign			Mandatory			
Name of Process		Char 15	Foreign			Mandatory			
Product_Process_Step									
Product ID		Num 6.0	Foreign			Mandatory			
Name of Process_Step		Char 15	Foreign			Mandatory			
Product_Printer									
Product ID		Num 6.0	Foreign			Mandatory			
Name of Printer		Char 12	Foreign			Mandatory			
Print Size	Size of the print	Char 12				Optional			
Orientation	Print orientation	Char 12				Optional	{portrait; landscape}		
Medium	Material on which the print is made	Char 12				Optional	{paper; trans-parency}		
Quality information	Specific quality information on the print, if not covered by general Printer quality information	Char 20				Optional			

Below is a list of all deviations from the CEN/TC 287 metadata standard. It is organised by metadata standard section, and contains five types of deviation:

- non-used mandatory elements
- mandatory elements used in a different manner
- mandatory elements used as optional
- revised naming
- revised data format

The rationale for each deviation will be given.

Box A-10: Deviations from the European metadata standard[86]

Section / Element	Devation	Rationale for deviation
DATASET OVERVIEW		
Abstract	omitted	no global datasets; prevent redundant information
Purpose	used in a different manner	preferred application to atomic elements ('Products')
Usage	omitted	can be derived from combinations of tables
Geometry sub-schema	omitted	invariant: spaghetti network
Spatial reference system	omitted	invariant: 'Tables' all indirect; 'Coverage' and 'Point_File' all direct

[86] CEN 1996a. A later version of the metadata standard (CEN 1996b) was received November 26, 1996. Decision is to stick to the earlier version since the general idea is quite similar across versions and all of the MMS design and pilot was already finished.

Section / Element	Devation	Rationale for deviation
Language	used only for Project_Coverage; optional	language is largely invariant: Dutch
DATASET QUALITY PARAMETERS		
Overall thematic accuracy	made optional	mandatory (detailed) fields elsewhere in metadatabase
Overall temporal accuracy	made optional	mandatory (detailed) fields elsewhere in metadatabase
Overall completeness	omitted	(detailed) fields elsewhere in metadatabase
Overall logical consistency	omitted	not applicable (no structures in datasets)
Overall positional accuracy	omitted	(detailed) fields elsewhere in metadatabase
SPATIAL REFERENCE SYSTEM		
Spatial reference date	'Year' as a character string	exact date seldomly known
Map projection	augmented with coordinate system	map projection is one thing, coordinate system another
Height reference system	omitted	no heights in our GIS
GEOGRAPHIC AND TEMPORAL EXTENT		
Bounding XY	decomposed into 4 optional elements	decomposition allows more useful information; optional fields because they are not so important (since practically all geographical data pertains to the Netherlands)
Bounding area	omitted	bounding XY seems to be enough
Type of indirect spatial reference system	'Key columns', made optional	naming more clear to the GIS user; optional because a Table may lack an indirect spatial reference system
Name of areal unit (xN)	omitted	far too much detail
ID code of areal unit (xN)	omitted	far too much detail
Coverage (xN)	omitted	far too much detail
Vertical extent (Minimum, Maximum)	omitted	not applicable
Temporal extent (From, To date)	made optional	often unknown
DATA DEFINITION		
Feature type name	'Coverage' name and 'Geo_File' name	adapted to Atlas GIS file structures
Feature type definition	'Description', made optional	often pretty obvious
Geometric primitive	'Feature type'	adapted to Atlas GIS conventions
Structure primitive	omitted	not applicable
Attribute type definition	made optional	since introduced for every Field; care taken not to augment administrative burden
Association type (name, definition, from, to, cardinality)	omitted	not applicable
CLASSIFICATION		
Thesaurus (name, administrator, terms, definitions, synonyms, related terms, broader terms, narrower terms)	omitted	not applicable

Section / Element	Devation	Rationale for deviation
ADMINISTRATIVE METADATA		
(All)	omitted (except for 'File location' similar to 'On-line access')	not relevant within the organisation, and since there is no variation in formats
METADATA REFERENCE		
Entry date	logical field	production date is already present in metadatabase; it suffices to make a logical association with that date
Last check date	logical field	production date is already present in metadatabase; it suffices to make a logical association with that date
Last update date	logical field	production date is already present in metadatabase; it suffices to make a logical association with that date

A.3.3. Functional Requirements

In Section A.3.1 the Information Flow Diagram was followed by a brief description of each process element (activity) within the metadata management system (MMS). Box A-7 therefore already contains the essence of what can be named 'Functional Requirements'. Yet, the current Section A.3.3 is necessary to augment and detail the process description with a number of specific issues. Matters pertaining to interface design and technical matters will be mentioned in separate sections. Now all that matters is, what the system is supposed to do. Underneath an additional (to Box A-7) listing of matters pertaining to functionality will be given. Reference is to MMS process elements in Boxes A-6 and A-7, entities in Box A-8, and attributes in Box A-9. 'Product' is an entity within the MMS, while 'product' is the physical or virtual thing in GIS.

RELATIONSHIPS BETWEEN MMS AND GIS

- GIS data and products are input to Metadata Creation and Product/Data Viewing.
- Metadata Browsing can lead to an assessed need of new data and/or processes.
- The MMS facilitates GIS operation: selected data, products, processes and environment are marked for use by GIS.
- The MMS facilitates data portability to a different, new GIS system.
- When Atlas GIS is closed the user is asked if Metadata Creation is to take place.

RELATIONSHIPS BETWEEN MMS ACTIONS AND USER ACTIONS

- Data are input to Metadata Creation only at user's discretion.
- Metadata Creation can be done automatically, if and when wished by the user, collecting as many attributes as possible (for *new metadata* and/or for *updates*):
    a) represent a Project File in the MMS
    b) represent a Table in the MMS
    c) represent a GeoFile in the MMS
- Data lineage is not generated automatically, although part of lineage is reflected in the MMS data structures.
- Selected data, products, processes and environment are visible to the GIS user, but not opened automatically by GIS.
- Metadata Creation is not mandatory upon closing Atlas GIS.
- Metadata Creation can only store a Product with all mandatory fields in Product and related entities completed.

- The MMS reminds the user of the need of Metadata Creation as soon as (within GIS) stored GIS data (Point_File, Project_File, Geo_File, Coverage, Field or Table) is opened that is not represented in the MMS.
- The MMS reminds the user of the existence of products that may be discarded since the saving date has expired.
- Metadata Creation suggests default values and texts where applicable.
- Intersection tables are automatically filled.

MISCELLANEOUS

- The MMS will have preprogrammed queries, as well as a possibility to make tailor-made queries (see Section A.3.4: Interface Requirements).
- The MMS will offer functionality to print what is shown on the monitor, as well as the results of some of the queries.
- The MMS will offer a Help Facility (per preprogrammed screen).
- The MMS will use the Dutch language in all screens and printout.
- The MMS will have functionality to assign operators to three levels of authorisation:
    1. MMS Administrator: access to all functionality
    2. GIS Operator: access to all functionality, *excluding*:
        - authorisation of operators
        - violations of database referential integrity
        - changes to the structure of the MMS relational database (entities, attributes, relationships) GIS
    3. Operator: access to all functionality, *excluding*:
        - authorisation of operators
        - tailor-made queries
        - any changes to metadata
- Each user will have a password, assigned by the MMS Administrator.

A.3.4. Interface Requirements

The metadata management system (MMS) will only be useful if it establishes a firm relationship between what a (GIS) user does and what the system does. Therefore, an interface is crucial, and deserves a requirements section of its own. This section will go into matters as queries and screens.

Queries in our MMS can have following properties (sometimes in combinations):

1. Queries that are preprogrammed (‘*programmed*’).
2. Queries that may be defined by the MMS administrator on an ad hoc basis (‘*ad hoc*’).
3. Queries resulting in all tuples/attributes of one entity Y (‘*entity*’).
4. Queries resulting in the attributes of tuples in one or more entities Y, related to one specific tuple in one entity X (‘*XY*’).
5. Queries that allow to mouse-click on a tuple in entity Y in order to launch another query into tuples/attributes in entities Z, related to the Y-tuple (‘*YZ*’).

The table in Box A-11 lists the programmed queries that seem necessary. For each query reference is made to:

- the entity where it starts from (i.e. the tuple for which related tuples are sought)

- the entity (or entities) that are being queried
- MMS process element(s) for which it is important
- query properties (see above)
- the question(s) answered by the query

Box A-11 Queries and their properties

Query #	Starts from X-tuple	Entities queried (Y)	MMS Process elements	Query properties	Question answered	Remarks
1	X entity	All X tuples	Metadata Creation/ Browsing, Product/ Data/ Process/ Environment Selection	programmed, entity, YZ	which tuples are there in entity X?	Shows a.o. all Product ID's; possibiliy to click through to query#2-5
2	Product	External_Source, Data_Collection_ Method	Metadata Creation/ Browsing	programmed, XY	Product X has which (general) properties?	Show X and Y attributes; Modify and Save functions; Saving date expired?
3	Product	Point_File, Project_File, Coverage + Geo_File, Field + Table, Other_Product	Metadata Creation/ Browsing, Data Selection	programmed, XY	Product X has which (data) properties?	Show X and Y attributes; Modify and Save functions; alternative Y's
4	Product	Process, Process_Step + GIS_Operation / Added_Operation / Other_Procedure	Metadata Creation/ Browsing, Process Selection	programmed, XY	Product X has which (process) properties?	Show X and Y attributes; Modify and Save functions; 1X : Many Y
5	Product	Originator, Operator, Printer	Metadata Creation/ Browsing, Environment Selection	programmed, XY	Product X has which (environ-ment) properties?	Show X and Y attributes; Modify and Save functions; 1X : Many Printers
6	Project_File	Point_File, Coverage + Geo_File, Field + Table	Metadata Creation/ Browsing, Data/ Product Selection	programmed, XY	X is related to which Y?	Modify and Save functions
7	Geo_File	Coverage	Metadata Creation/ Browsing, Data Selection	programmed, XY	X contains which Y?	Modify and Save functions
8	Table	Field	Metadata Creation/ Browsing, Data Selection	programmed, XY	X contains which Y?	Modify and Save functions
9	Process	Process_Step + GIS_Operation / Added_Operation / Other_Procedure	Metadata Creation/ Browsing, Process Selection	programmed, XY	X consists of which Y?	Add a Process_Step, Modify and Save functions; alternative Y's
10	Process_Step	GIS_Operation / Added_Operation / Other_Procedure	Metadata Creation/ Browsing, Process Selection	programmed, XY	X is based on which Y?	Modify and Save functions, alternative Y's
11	Process_Step	Process	Metadata Creation/ Browsing, Process Selection	programmed, XY	X is an element of which Y?	Modify and Save functions
12	Product	Product	Metadata Creation/ Browsing	programmed YZ	X consists of which Y?	Add a Product; as X or as Y; possibility to click through to query#2-5
13	Coverage + Geo_File	Project_File	GIS Maintenance	programmed, XY	X has been used for which Y?	

Query #	Starts from X-tuple	Entities queried (Y)	MMS Process elements	Query properties	Question answered	Remarks
14	Point_File	Project_File	GIS Maintenance	programmed, XY	X has been used for which Y?	
15	Field + Table	Project_File	GIS Maintenance	programmed, XY	X has been used for which Y?	
16	Project_File	(map view)	Product/Data Viewing	programmed	What does X look like?	Use GIS functionality
17	Table	(table view)	Product/Data Viewing	programmed	What does X look like?	Use GIS/DBMS functionality
18	X entity	Y entity	All	ad hoc	X is related to which Y?	

Query results will default be presented ordered by *(Name)* primary or - in the case of the Product entity- secondary key. Queries will be organised into a few screens which show logical structure within and in between. The details of these screens will have to be elaborated with the person who will actually build the system (and who may be contracted), since he/she is likely to have far more experience in designing a good interface. We assume it suffices to mention here the global structure between screens and some general and specific requirements of screens needed.

General requirements of screens[87]

The screens should:
... show options
... be easy to grasp to novice users
... allow shortcuts to experienced users
... show consistency across screens
... establish a sound labor division between human and computer: let each do what he/she/it is best in
... provide help

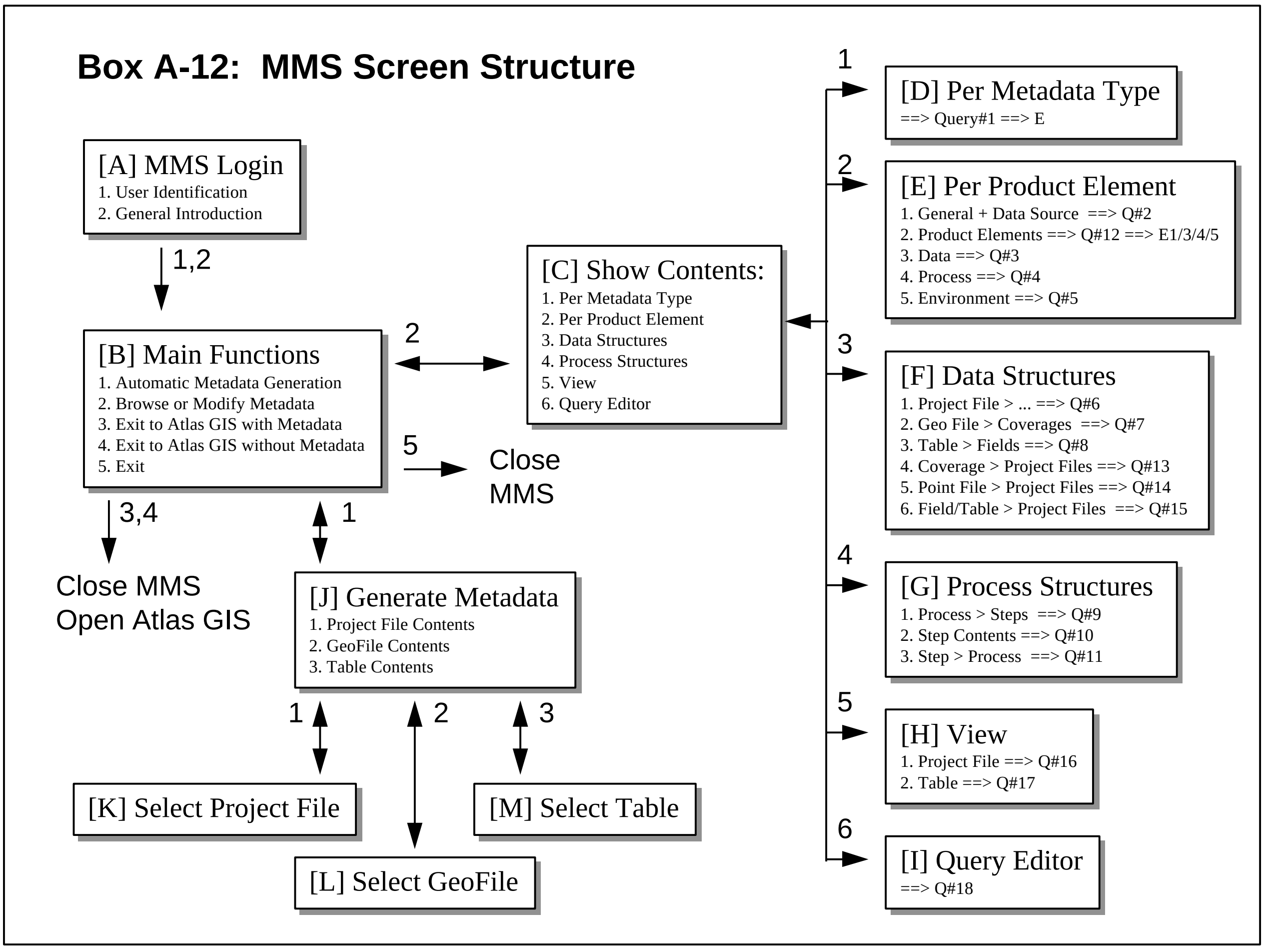
Global structure between screens

The MMS can be activated *within* and *outside* GIS. When it is closed certain screen elements are gathered in one window that remains visible as long as Atlas GIS is opened: the elements selected for use in GIS.

Box A-12 contains the global structure between screens.

[87] General screen requirements were distilled from Petch (1993b).

Box A-12: MMS Screen Structure



Some specific requirements of screens

Screens will show segments, containing (where applicable):

- Metadata Contents
- Primary Options (from Box A-12)
- Secondary Options (see below)

Secondary Options' domain is:

- select element for continued search
- select element for GIS
- add a metadata element
- sort by name
- sort by date
- cancel
- save
- previous screen
- back to Main Functions (screen B)
- print (sufficient is query result copying to the Windows clipboard)
- help

Special attention is needed for screens and queries that show attributes depending on alternative relationships (from Product and from Process_Step). The printouts need to be elaborated in case it is impossible to print all results of any query. All screens will be in Dutch.

A.3.5. Technical Requirements

Although we have touched on rather technical issues in some of the past sections, typically technical issues will be gathered in this section that concludes our MMS requirements chapter. Underneath is a list of rules that should in the *ideal* case be observed by the MMS developer.

1. The MMS will be able to operate under Windows 3.x/DOS, Windows 95 and Windows NT.
2. The MMS will be installable from diskette or CD-ROM.
3. The MMS will support the use of Access 2.0 or higher (including Access 95) as an RDBMS.
4. The MMS may make use of ODBC, OLE and/or DDE.
5. The MMS may make use of Visual Basic or another scripting language for Access and/or Atlas GIS.
6. Notwithstanding 3/4/5 every effort will be made to write such syntax that the MMS will be easily portable to a different GIS (especially the next generation of desktop GIS that after ArcView 3.0 will come from ESRI - who owns both ArcView and Atlas GIS) and/or RDBMS (especially what will combine easily with ESRI’s next generation desktop GIS).
7. The MMS will allow use on a (a.o. Novell) network.
8. Simultaneous use by multiple users is supported.
9. The MMS will guard referential integrity of the metadatabase (in case tuples are added/modified/deleted).
10. The MMS will perform as a ‘fast’ system, e.g. screen filling by query within 3 seconds in case of 1000 Product tuples on a 90 Mhz Pentium computer with 24 MB total memory.
11. The MMS documentation (also available as Help facility) will a.o. go into all following matters:
    - Introduction
    - MMS structure
    - How to use the system
    - How to conduct various system adaptations

A.4. Details of Pilot

To accomplish the pilot described in Section 5.6 following steps were taken:

Step 1: Raw metadatabase

Using GeoKey Edit a raw and empty Access metadatabase 'metamap.mdb' is created. This database contains 25 metatables, whose relations are known only within the GeoKey software. The relations don't show up if the metadatabase is inspected in Access 2.0.

Step 2: Dataset types

Our ER diagram is compared to what is possible in GeoKey, in order to find out which dataset types should be defined in the metadatabase. Several concessions turn out to be necessary:

- The ER diagram contains a central *Product* entity which may refer to a data element (Point_File, Project_File, Geo_File, Coverage, Field, or Table) or to an Other_Product (e.g. a physical map or bitmap picture). With GeoKey the central entity is a *Dataset* which is of one and only one dataset type. So the entity *Product* was dropped, along with *non-file tuples of the entity Other_Product*.
- Another consequence is that *Coverage* and *Field* are dropped as entities, since these are not datasets but rather elements of datasets Geo_File and Table, respectively. It is possible to describe them as *Items* in datasets, but this procedure is not very strict, since all intersection tables of the ER diagram are dropped as well. We also drop the possibility to devise metadata fields per item, since there is no standard way to differentiate item metadata structure along with dataset type.
- Entities *Data_Collection_Method*, *External_Source*, *Originator*, *Operator* and *Printer* are dropped as well, since these are not represented by datasets. It may be possible, however, to represent this metadata by values of metavariables rather than by entities.
- *Process* entities can exist only as program (syntax) files. Therefore Process_Step is dropped. GIS_Operation, Added_Operation and Other_Procedure might be represented by values of metavariables.

We arrive at following six dataset types:

1. ***Project_File*** (metatable 'project'),
2. ***Geo_File*** (metatable 'geografisch'),
3. ***Point_Table*** (metatable 'punten'),
4. ***Attribute_Table*** (metatable 'attribuut'),
5. ***Syntax*** (metatable 'programma'),
6. ***Picture*** (metatable 'afbeelding').

Where necessary these dataset types are (using Access 2.0) added to the metadatabase structure by adding appropriate metatables with these names, and entering corresponding names (and codes) in metatable *typetable*.

Step 3: Structuring the other metadata

Dataset-type is one vehicle for metadata. GeoKey offers more vehicles, which all can be used in the selection of datasets. Upon inspection of GeoKey's structure (summarized above) following use is made of GeoKey's possibilities:

- As *general (meta)variables* will be used: **datacode**, **type**, **filename**, **datatype**[88], **description**, **purpose**, **sources**, **overall_quality, location**, **backup**, **date_of_latest_change**. There is also a mandatory metafield '**bestandstitel**'. In metafields **regions**, **lines**, **points**, **VSN_internal**, and **VSN_external** a Yes / No value is put to indicate the contents of a dataset.
- As *specific (meta)variables* will be used:
    1. *Project_File*: **projection**, **geo_files_opened**, **point_tables_opened**, **attribute_tables_opened**
    2. *Geo_File*: **projection**, **overall_pos_accuracy**, **overall_completeness**
    3. *Point_Table*: **number_of_points**, **projection**, **positional_accuracy**, **overall_them_accuracy**, **overall_temp_validity**, **completeness**, **process_history**
    4. *Attribute_Table*: **number_of_records**, **overall_them_accuracy**, **overall_temp_validity**, **overall_completeness**, **process_history, key_columns**
    5. *Syntax*: **software_used**, **cave_at**
    6. *Picture*: **file_type, pixels**
- As *items* will be used:

    Geo_File: **coverages** (map layers) contained within the Geo_File

    Point_Table and *Attribute_Table*: **attribute fields**

    Syntax: **process steps**
- As *keywords* mention can be made of simple general aspects of datasets, e.g. **is_old**, **is_backup**, **is_OK**, **is_not_OK**. The Saving Date of a product could be implemented, facilitating specific queries for datasets that may be discarded.
- No *hierarchy* is implemented. There is a preference to use general metavariables instead, since these offer more flexibility in combining criteria for the selection process.
- The following *geographical classification* is defined: **regional**, **Netherlands**, **other_countries_in_Europe**, **Europe**, **other_countries_outside_Europe**, **other_continents**, **World**, **(none)**
- *Geography* will be entered by **minimum and maximum X- and Y-coordinates** all according to the 'Rijksdriehoekmeting'-projection (for Dutch datasets that is).

Not all metavariables from the design are implemented. Limited congruence of the proposed GeoKey metadata structure and our ER diagram is accountable for the non-strict relation between Project Files and Geo Files/Point Tables/Attribute Tables, between Geo Files and coverages, and between Pointy Tables/Attribute Tables and fields. The limited design of this pilot is accountable for not entering e.g. originator, operator, data collection method, external source details, user authorisation, metavariable ranges (except for the boolean general variables, having default No), metadata entry data, and systematic bitmap representations of data types 1.-3. (although one specimen will be prepared for demonstration purposes). No metafields will be mandatory in the pilot system (except datatype), nor will they have explanations. Projection will be 'Rijksdriehoekmeting' unless otherwise specified. All coordinates will be in meters (USER_3) unless specified otherwise.

Step 4: Entering the metadata structure

All *metavariables* are (using Access 2.0) entered in the metatables: 'dataset', 'project', 'geografisch', 'punten', 'attribuut', 'programma', and 'afbeelding'.

Items can be entered while metadata entry takes place. In the pilot we do not include specific information on items (e.g. details of fields or coverages), since there is no automated way to do maintenance on such quite specific information. The

[88] We have not found an elegant way to have the text equivalent of the data **type** code automatically entered in a display group. Therefore there is now a need to enter the data type twice: upon entering a dataset and again (mandatory) when entering the general metavariables.

saving date is not implemented since this variable entered the functional specification after the pilot was finished.

Keywords are (using Access 2.0) entered in metatable ‘treftext’. In GeoKey Edit they can be linked to particular datasets.

The *geographical classification* is entered in metatables ‘nivtext’ (containing the information that there is only one classification), and ‘vgitext’ (containing the actual classification). Links to datasets in ‘vgidataset’ are produced by GeoKey Edit.

For *geography* no special arrangements have to be made. There is already a selmap.bmp specified with minimum and maximum X and Y values according to the ‘Rijksdriehoekmeting’. Geographic selection will - in this pilot case - be implemented for the Dutch area only.

Some specific problems that are encountered (and solved):

- Memo fields for GeoKey must in Access be entered as large integers, and after that set as memo fields in GeoKey Edit.
- Logical fields for GeoKey must in Access be entered as text variables, and after that given range limits (e.g. “No” and “Yes” in GeoKey Edit.

Step 5: The display of metadata

GeoKey Edit facilitates display of metadata in GeoKey Select in a number of display groups. Metavariables names are simply used as labels. Following arrangements are made:

Display Group	**in Dutch**	**Contents**
Overview	Overzicht	datacode, filetitle, filename, datatype, description, purpose, location
General	Algemeen	all general metavariables, except regions, lines, points, VSN_internal, VSN_external
Access	Toegang	(not used; metadata contained in general metavariables)
Contents	Inhoud	datacode, type, regions, lines, points, VSN_internal, VSN_external
Specific	Specifiek	datacode, type, all specific metavariables
Items	Items	only item names, no details
Print single	Print enkel	same as with Overview
Print all	Print allemaal	all general and specific metavariables

Step 6: Entering metadata

Thought has been given to enter all datasets into GeoKey. However, neither is this necessary in the pilot situation, nor is it feasible for mere pilot purposes, since there are 54 project files, 70 geofiles and more than 132 attribute tables. Therefore it has been decided to select and enter 5 project files, 5 geofiles, (all) 4 point tables, 5 attribute tables, (all) 3 syntax (script) files and (the only) 1 picture. Input was done manually after inspecting and taking notes of the metadata of datasets. GIF- or BMP-files of the project files and pictures were included in the metadata system.

Step 7: Link to Atlas GIS

GeoKey contains a feature that allows to pass on a selected dataset to an application. This feature is achieved by writing a script file GEOKEY.SCR which is read when GeoKey Select is started. The script file contains (per datatype) a command (or series of commands) that calls the application(s) of choice, as well as parameters derived from metavariables.

In this pilot this feature is used in part. For project files the feature is implemented (invoking Atlas GIS). Atlas GIS can however not be invoked for plainly reading a geofile, an attribute table or a point table. Transferring attribute and point tables to a database program would be possible if the metavariable ‘filename’ included the full file path or if a batch file were invoked to concatenate metavariables ‘location’ and ‘filename’.

Furthermore the script file is arranging such that for datatype syntax Atlas GIS is simply started. For datatype picture LViewPro 3.1 is started, but - again - without directly loading the picture file itself.

Step 8: Working with GeoKey Select

Some exercises[89] with Geokey Select are undertaken to evaluate the way of working facilitated by the medata system against the following sets of requirements contained in the system specification:

- functional requirements
- interface requirements
- technical requirements

There is no need to go into the Processes nor into the Entities, Attributes and Relationships, since these are either contained in the requirements, either dealt with in Steps 1-7 above. Not all requirements are entered into this evaluation: only important requirements for which a clear conclusion can be reached and which have not been treated above. It should be reminded that the requirements were devised for the development of a new system and not for the evaluation of an existing one.

GeoKey Select conforms to our requirements regarding following aspects:

- Metadata creation, metadata browsing, dataset selection, (though static) dataset viewing.
- Suggestion of default values and texts.
- Intersection tables are automatically filled.
- Tailor-made metadata queries.
- Metadata printing.
- Dutch language possible.
- Permit levels of authorisation.
- Passwords.
- Consistency across screens.
- Help available in screens.
- Datasets with expired saving date may be queried.
- Sound labour division between operator and computer.
- Portability to a different GIS system.
- Fit for networking, with simultaneous use by multiple users.
- Available for Windows 3.x, Windows 95 and UNIX.
- Support Access RDBMS.
- Guarding referential integrity of the metadatabase.
- Fast.
- Documentation.

GeoKey Select (without extra programming) does not conform to our requirements regarding following aspects:

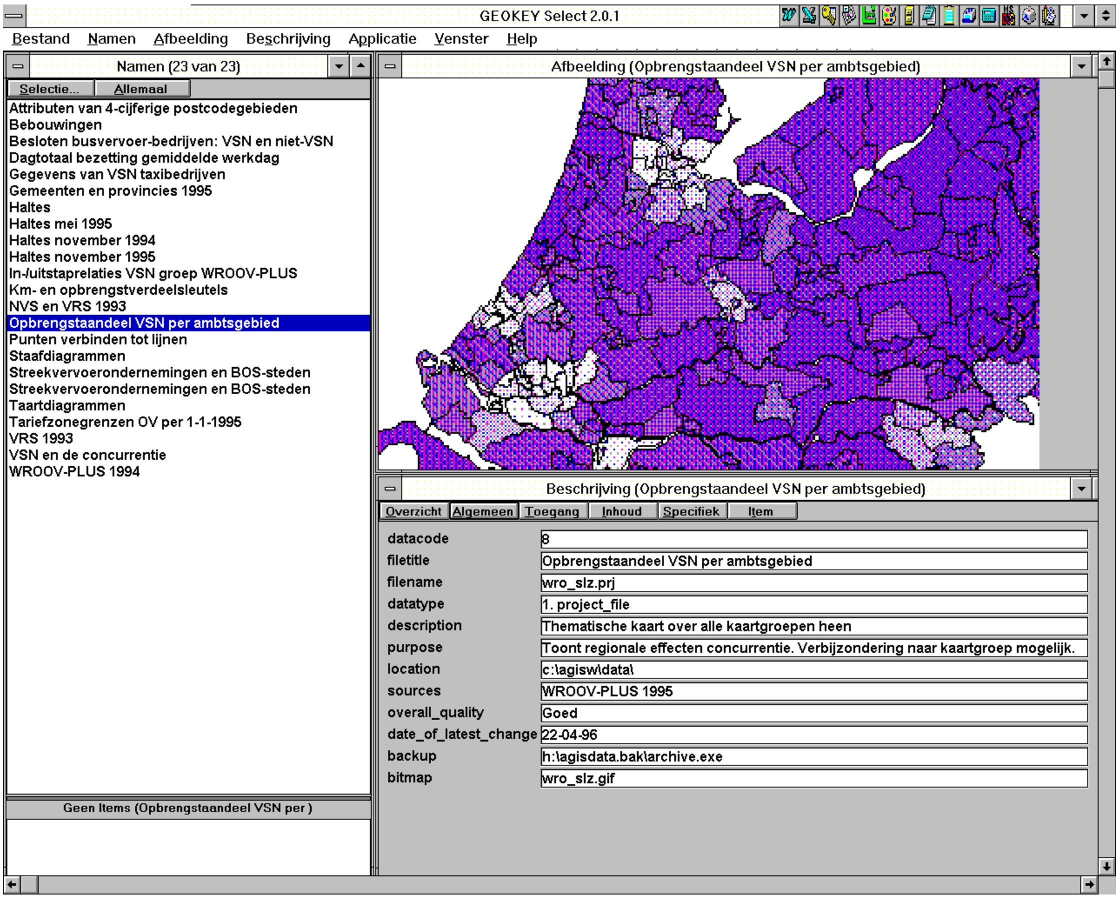
- Automatic metadata creation.[90]
- Prompting from GIS data creation to metadata creation.

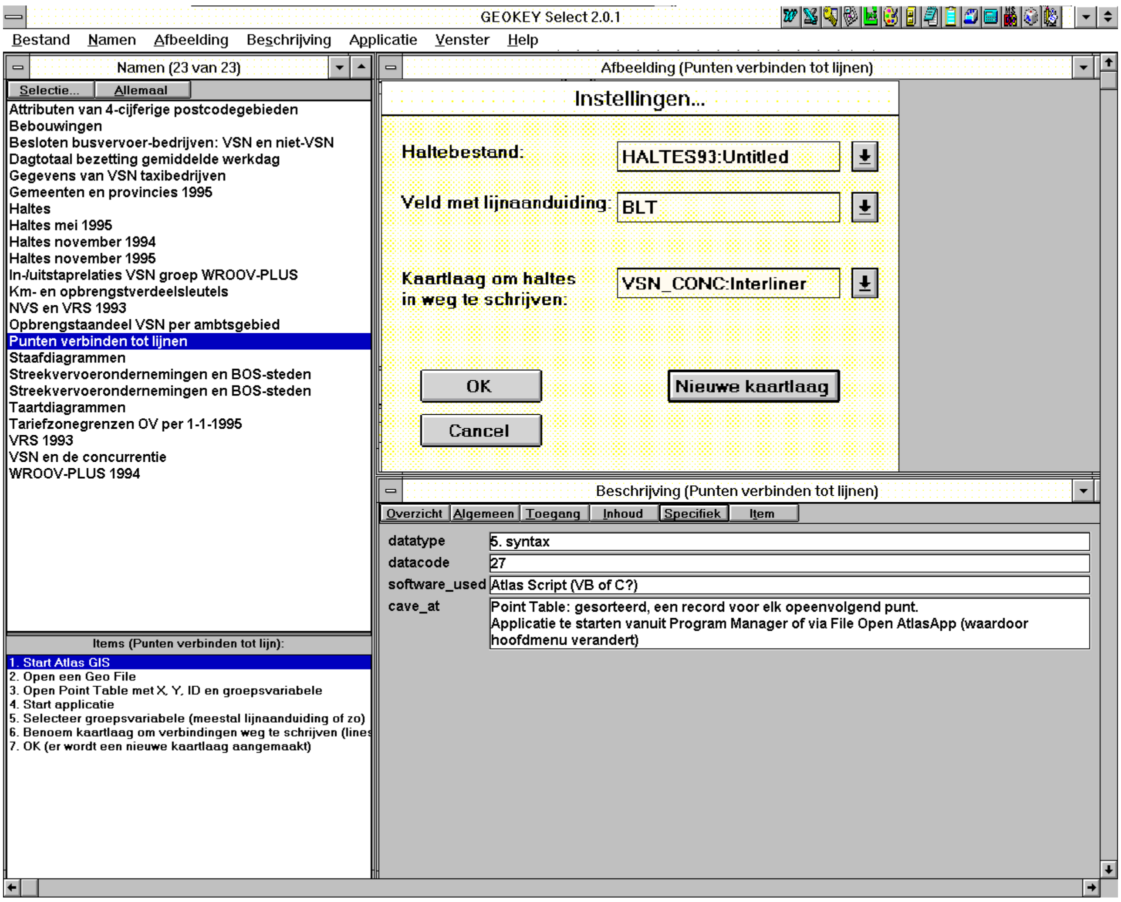
---

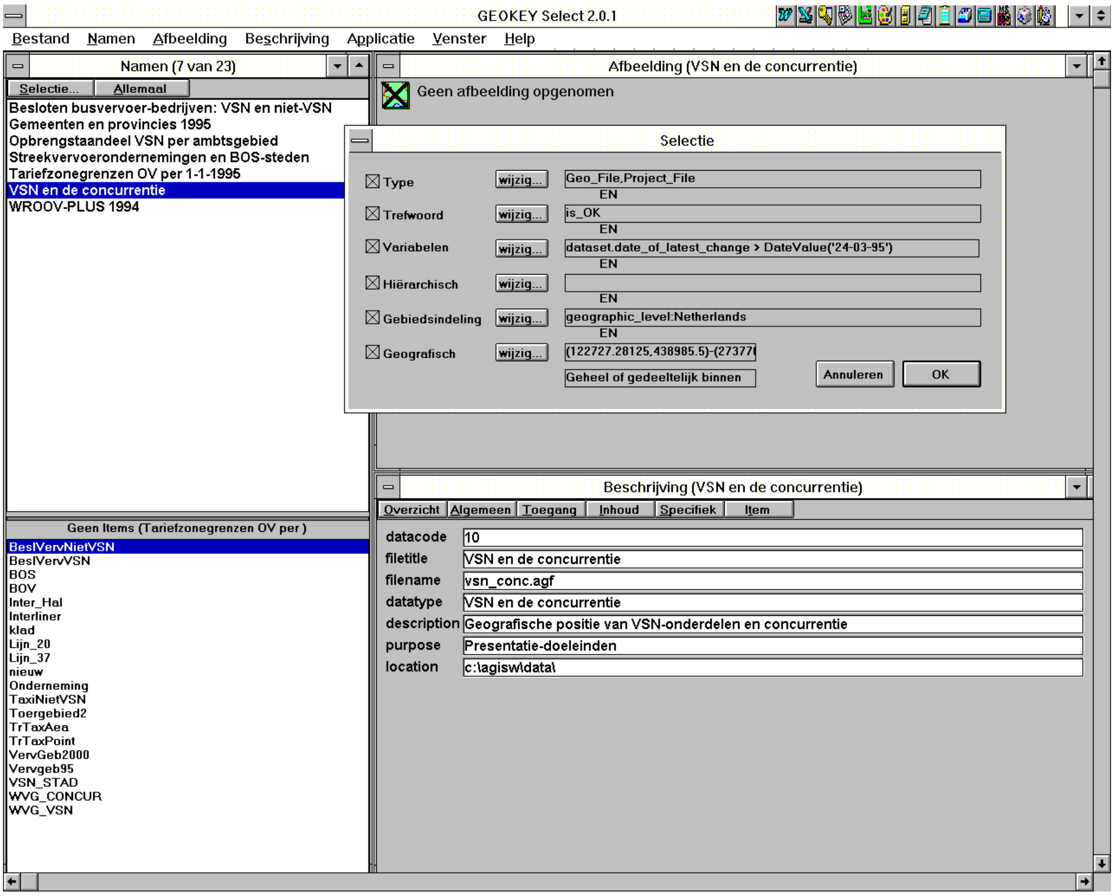
[89] Chiefly data selection and metadata browsing/printing.

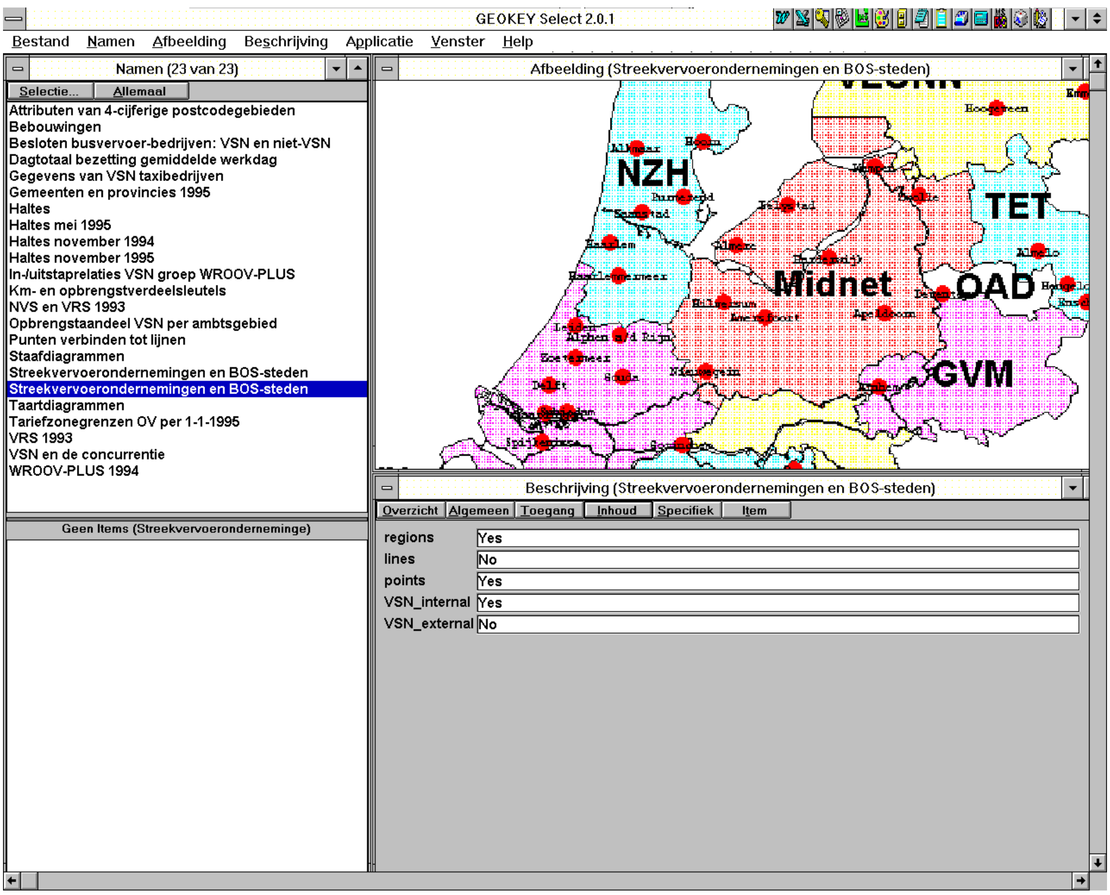
[90] It took two days to find and enter metadata on only 23 datasets; linear extrapolation gives 23 days, though a learning curve might suggest a lower number of days. So much time is not available.

- GIS environment selection.


GEOKEY Select 2.0.1
Bestand Namen Afbeelding Beschrijving Applicatie Venster Help
Namen (23 van 23)
Selectie... Allemaal
Attributen van 4-cijferige postcodegebieden
Bebouwingen
Besloten busvervoer-bedrijven: VSN en niet-VSN
Dagtotaal bezetting gemiddelde werkdag
Gegevens van VSN taxibedrijven
Gemeenten en provincies 1995
Haltes
Haltes mei 1995
Haltes november 1994
Haltes november 1995
In-/uitstaprelaties VSN groep WROOV-PLUS
Km- en opbrengstverdeelsleutels
NVS en VRS 1993
Opbrengstaandeel VSN per ambtsgebied
Punten verbinden tot lijnen
Staafdiagrammen
Streekvervoerondernemingen en BOS-steden
Streekvervoerondernemingen en BOS-steden
Taartdiagrammen
Tariefzonegrenzen OV per 1-1-1995
VRS 1993
VSN en de concurrentie
WROOV-PLUS 1994
Geen Items (Opbrengstaandeel VSN per)
Afbeelding (Opbrengstaandeel VSN per ambtsgebied)
Beschrijving (Opbrengstaandeel VSN per ambtsgebied)
Overzicht Algemeen Toegang Inhoud Specifiek Item
datacode 8
filetitle Opbrengstaandeel VSN per ambtsgebied
filename wro_slz.prj
datatype 1. project_file
description Thematische kaart over alle kaartgroepen heen
purpose Toont regionale effecten concurrentie. Verbijzondering naar kaartgroep mogelijk.
location c:\agisw\data\
sources WROOV-PLUS 1995
overall_quality Goed
date_of_latest_change 22-04-96
backup h:\agisdata.bak\archive.exe
bitmap wro_slz.gif


GEOKEY Select 2.0.1
Bestand Namen Afbeelding Beschrijving Applicatie Venster Help
Namen (23 van 23)
Selectie... Allemaal
Attributen van 4-cijferige postcodegebieden
Bebouwingen
Besloten busvervoer-bedrijven: VSN en niet-VSN
Dagtotaal bezetting gemiddelde werkdag
Gegevens van VSN taxibedrijven
Gemeenten en provincies 1995
Haltes
Haltes mei 1995
Haltes november 1994
Haltes november 1995
In-/uitstaprelaties VSN groep WROOV-PLUS
Km- en opbrengstverdeelsleutels
NVS en VRS 1993
Opbrengstaandeel VSN per ambtsgebied
Punten verbinden tot lijnen
Staafdiagrammen
Streekvervoerondernemingen en BOS-steden
Streekvervoerondernemingen en BOS-steden
Taartdiagrammen
Tariefzonegrenzen OV per 1-1-1995
VRS 1993
VSN en de concurrentie
WROOV-PLUS 1994
Items (Punten verbinden tot lijn):
1. Start Atlas GIS
2. Open een Geo File
3. Open Point Table met X, Y, ID en groepsvariabele
4. Start applicatie
5. Selecteer groepsvariabele (meestal lijnaanduiding of zo)
6. Benoem kaartlaag om verbindingen weg te schrijven (lines
7. OK (er wordt een nieuwe kaartlaag aangemaakt)
Afbeelding (Punten verbinden tot lijnen)
Instellingen...
Haltebestand: HALTES93:Untitled
Veld met lijnaanduiding: BLT
Kaartlaag om haltes in weg te schrijven: VSN_CONC:Interliner
OK
Nieuwe kaartlaag
Cancel
Beschrijving (Punten verbinden tot lijnen)
Overzicht Algemeen Toegang Inhoud Specifiek Item
datatype 5. syntax
datacode 27
software_used Atlas Script (VB of C?)
cave_at Point Table: gesorteerd, een record voor elk opeenvolgend punt. Applicatie te starten vanuit Program Manager of via File Open AtlasApp (waardoor hoofdmenu verandert)


GEOKEY Select 2.0.1
Bestand Namen Afbeelding Beschrijving Applicatie Venster Help
Namen (7 van 23)
Selectie... Allemaal
Besloten busvervoer-bedrijven: VSN en niet-VSN
Gemeenten en provincies 1995
Opbrengstaandeel VSN per ambtsgebied
Streekvervoerondernemingen en BOS-steden
Tariefzonegrenzen OV per 1-1-1995
VSN en de concurrentie
WROOV-PLUS 1994
Afbeelding (VSN en de concurrentie)
Geen afbeelding opgenomen
Selectie
Type wijzig... Geo_File,Project_File
EN
Trefwoord wijzig... is_OK
EN
Variabelen wijzig... dataset.date_of_latest_change > DateValue('24-03-95')
EN
Hiërarchisch wijzig...
EN
Gebiedsindeling wijzig... geographic_level:Netherlands
EN
Geografisch wijzig... (122727.28125,438985.5)-(27377
Geheel of gedeeltelijk binnen
Annuleren OK
Geen Items (Tariefzonegrenzen OV per)
BeslVervNietVSN
BeslVervVSN
BOS
BOV
Inter_Hal
Interliner
klad
Lijn_20
Lijn_37
nieuw
Onderneming
TaxiNietVSN
Toergebied2
TrTaxAea
TrTaxPoint
VervGeb2000
Vervgeb95
VSN_STAD
WVG_CONCUR
WVG_VSN
Beschrijving (VSN en de concurrentie)
Overzicht Algemeen Toegang Inhoud Specifiek Item
datacode 10
filetitle VSN en de concurrentie
filename vsn_conc.agf
datatype VSN en de concurrentie
description Geografische positie van VSN-onderdelen en concurrentie
purpose Presentatie-doeleinden
location c:\agisw\data\


GEOKEY Select 2.0.1
Bestand Namen Afbeelding Beschrijving Applicatie Venster Help
Namen (23 van 23)
Selectie... Allemaal
Attributen van 4-cijferige postcodegebieden
Bebouwingen
Besloten busvervoer-bedrijven: VSN en niet-VSN
Dagtotaal bezetting gemiddelde werkdag
Gegevens van VSN taxibedrijven
Gemeenten en provincies 1995
Haltes
Haltes mei 1995
Haltes november 1994
Haltes november 1995
In-/uitstaprelaties VSN groep WROOV-PLUS
Km- en opbrengstverdeelsleutels
NVS en VRS 1993
Opbrengstaandeel VSN per ambtsgebied
Punten verbinden tot lijnen
Staafdiagrammen
Streekvervoerondernemingen en BOS-steden
Streekvervoerondernemingen en BOS-steden
Taartdiagrammen
Tariefzonegrenzen OV per 1-1-1995
VRS 1993
VSN en de concurrentie
WROOV-PLUS 1994
Geen Items (Streekvervoeronderneminge)
Afbeelding (Streekvervoerondernemingen en BOS-steden)
NZH
TET
Midnet
OAD
GVM
Hoogeveen
Alkmaar
Purmerend
Zaanstad
Haarlem
Almere
Lelystad
Haarlemmermeer
Hilversum
Amersfoort
Apeldoorn
Leiden
Alphen a/d Rijn
Zoetermeer
Delft
Gouda
Nieuwegein
Beschrijving (Streekvervoerondernemingen en BOS-steden)
Overzicht Algemeen Toegang Inhoud Specifiek Item
regions Yes
lines No
points Yes
VSN_internal Yes
VSN_external No

Step 9: Evaluating GeoKey

What conclusions may be derived from the observations above? What are major strengths and weaknesses of GeoKey relative to Atlas GIS at the VSN group? To what degree could GeoKey be the MMS that the VSN group needs for its Atlas GIS operation?

These questions are answered in Section 5.6

A.5. List of references

1. *Atlas GIS Reference Manual*, 1994, Strategic Mapping Inc., Santa Clara (USA).
2. Backer, C. de, 1982, *Kennismaking met Informatica*, De Sikkel, Malle (Belgium).
3. Barker, R., 1990[2] *CASE*METHOD Entity Relationship Modelling*, Oracle Corporation UK Limited, Addison-Wesley Publishing Company, Wokingham, United Kingdom.
4. Berends, J., 1995, Clearinghouse: zowel Albert Cuyp als Albert Heyn, *Vi-MATRIX*, Vol. 3, No. 7, p.48-49.
5. Bisoen, M.R. and Heerikhuisen, A.R., 1995a, *Vooronderzoek 'SC Metadata-informatiesysteem' - Consequenties Functie-eisen*, Staring Centrum/Dienst Landbouwkundig Onderzoek, Wageningen (the Netherlands).
6. Bisoen, M.R. and Heerikhuisen, A.R., 1995b, *Vooronderzoek 'SC Metadata-informatiesysteem' - Onderzoek metadata attributen voor SC-DLO*, Staring Centrum/Dienst Landbouwkundig Onderzoek, Wageningen (the Netherlands).
7. Bisoen, M.R. and Heerikhuisen, A.R., 1995c, *Vooronderzoek 'SC Metadata-informatiesysteem' - Projectvoorstel 'Bouw SC Metadata-informatiesysteem'*, Staring Centrum/Dienst Landbouwkundig Onderzoek, Wageningen (the Netherlands).
8. Bisoen, M.R. and Heerikhuisen, A.R., 1995d, *Vooronderzoek 'SC Metadata-informatiesysteem' - Technisch Document*, Staring Centrum/Dienst Landbouwkundig Onderzoek, Wageningen (the Netherlands).
9. Bisoen, M.R. and Heerikhuisen, A.R., 1995e, *Vooronderzoek 'SC Metadata-informatiesysteem' - Afstudeerverslag*, Staring Centrum/Dienst Landbouwkundig Onderzoek, Wageningen (the Netherlands).
10. Brand, March 1996, M.J.D., *The standardisation debate*. Open letter to the EUROpean Geographic Information community, World Wide Web, http://www.frw.ruu.nl/eurogi/forum/open02.html (August 15, 1996).
11. Bregt, 1995, A.K. et al., *Landsdekkende Ruimtelijke Informatie bij DLO-Staring Centrum*, Geographical Information in the Netherlands, First Joint European Conference and Exhibition on Geographical Information (Proceedings Stream 7), The Hague, p.7.81-7.93.
12. Briggs, D. et al.., 1993, *Data Quality*, Course Notes of Module 8 of the International Distance Learning GIS Diploma Programme, 2nd ed. (edited by Petch, J.), InterGIS (place not mentioned).
13. Brunschot, C.J.M. van, 1994, *Tutor Assessed Assignment B* (Soft Systems Analysis) for Module 10 (GIS in Organisations) of the InterGIS International Diploma Course in GIS, Eindhoven (The Netherlands).
14. Burnhill, P., 1991, Metadata and Cataloguing Standards: One Eye on the Spatial, in: Medyckyj-Scott, D. et al. (eds.), *Metadata in the Geosciences*, Papers derived from a symposium sponsored by the UK Association for Geographic Information, Midlands Regional Research Laboratory/Loughborough University of Technology/University of Leicester, Group D Publications, Loughborough (United Kingdom), p.13-37.
15. Cann, H. van, 1995a, Metadatamodellen lopen niet ver uiteen, *Vi-MATRIX*, Vol. 3, No. 6, p.8-11.
16. Cann, H. van, 1995b, 'We hebben geen GIS meer, maar een geo-informatie infrastructuur', *Vi-Matrix*, Vol. 3, No. 4, p.22-23.
17. CEN, 1996a, Draft prEN 00000-0, *Geographic Information - Data Description - Metadata*, Draft V2 - For 2nd Informal Vote by W2, Working Group 2 of CEN/TC 287, Produced by PT01, CEN, Brussels.
18. CEN, 1996b, prEN 287009, *Geographic Information - Data Description - Metadata*, CEN, Brussels (February 1996).
19. Clinton, W.J., 1994, Executive Order 12906, Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure, *Federal Register*, Vol. 59, No. 71, p.17671-17674.

20. Cornelius, S., 1991, Spatial Data Auditing, in: Medyckyj-Scott, D. (ed.): *'How do I know what data I've got'*, a Symposium on 'Approaches to the Handling of Spatial Metadata', Education, Training & Research Publication Number 1, AGI, London.
21. Dangermond, J., 1996, What Atlas GIS Means to ESRI, *ArcNews*, Vol. 18, No. 3, p.6.
22. Darman, R.G., 1990, *Circular A-16, Coordination of Surveying, Mapping, and Related Spatial Data Activities*, U.S. Office of Management and Budget, October 19, Washington.
23. *Deelnemende Pilots in het Nationale Clearinghouse Geo-informatie*, 1996, RAVI, World Wide Web, http://www.euronet.nl/80/users/ravi/pilot.html (February 28, 1996).
24. EDRM (Standardisation WG/CEN TC278 WG7 SWG7.2), 1994, *Draft GDF 2.2 Specification*, European Digital Road Map II, Commission of the European Communities (place not mentioned).
25. *EGII Policy Document* CPB/RAL Version 20.9.95, World Wide Web, http://www.echo.lu/impact/projects/gis/en/gi2000p5.html (January 25, 1996).
26. ESRI (1996), ESRI Outlines Plans for Atlas GIS Software (1996), *ArcNews*, Vol. 18, No. 3, p.9.
27. EUROGI, 1994/95, *European Umbrella Organisation for Geographical Information*, World Wide Web, http://www.frw.ruu.nl/eurogi/eurorg.html#wp9495 (January 25, 1996).
28. FGDC, 1994, *Content Standards for Digital Geospatial Metadata* (draft), Federal Geographic Data Committee, Washington, World Wide Web, http://geochange.er.usgs.gov/pub/tools/metadata/standard/metadata.html (January 16, 1996).
29. FGDC, 1995a, *The National Digital Geospatial Data Framework*, Federal Geographic Data Committee, Washington, World Wide Web, http://fgdc.er.usgs.gov/framework.html (January 16, 1996)
30. FGDC, 1995b, *National Geographic Data Clearinghouse*, Federal Geographic Data Committee, Washington, World Wide Web, http://fgdc.er.usgs.gov/clearover2.html (January 16, 1996)
31. *Information Access and Retrieval MetaMap,* Federal Geographic Data Committee, Washington, World Wide Web, http://fgdc.er.usgs.gov/clearinghouse/clearing.metamap.gif (January 16, 1996).
32. FGDC, 1995d, *Federal Geographic Data Committee Data Standards,* Federal Geographic Data Committee, Washington, World Wide Web, http://fgdc.er.usgs.gov/metaover2.html (January 16, 1996).
33. *GI2000 Document* dated 19.06.95, World Wide Web, http://www.echo.lu/impact/projects/gis/en/gi2000t13.html (January 25, 1996).
34. Gartzen, P. and Hale, K., 1995, *The European GIS Users: A Very Unhappy Lot*, User and Distribution Series Studies CCAM-EU-UA-9501, Dataquest Europe Ltd. (place not mentioned).
35. Geodan, 1995a, *Ge-O-Key*, Het visuele zoekprogramma voor geografische databases, Amsterdam.
36. Geodan, 1995b, *Ge-O-Key Select Handleiding*, Het visuele zoekprogramma voor geografische databases, Amsterdam.
37. Geodan, 1995c, *Ge-O-Key Edit Handleiding*, Het visuele zoekprogramma voor geografische databases, Amsterdam.
38. GIS users agree to disagree about standards, 1996, *GIS Europe*, Vol. 5, No. 8, p.8.
39. Grothe, M. et al., 1994, *GIS, noodzaak of luxe?* Een verkenning naar het gebruik van geografische informatiesystemen bij private ondernemingen in Nederland, Nederlandse Geografische Studies, Nr. 183, Koninklijk Nederlands Aardrijkskundig Genootschap/Vakgroep Ruimtelijke Economie Vrije Universiteit Amsterdam, Utrecht.
40. *Handling Geographic Information,* 1987, Report to the Secretary of State for the Environment, Committee of Enquiry into the Handling of Geographic Information, Chairman: Lord Chorley, Department of the Environment, Her Majesty's Stationery Office, London.
41. Heine, G., 1996, *OII - Geographical Information Standards*, ECSC-EC-EAEC, Brussels-Luxembourg, World Wide Web, http://www.echo.lu/impact/oii/gis.html#GIS (January 16, 1996).

42. Hobson, S.A., 1991, Methodology Issues in GIS Introduction, *AM/FM European Conference VII* (Proceedings), Montreux, p.207-212.
43. Inmon, W.H., 1992, *Building the Data Warehouse*, Wiley/QED Publication, John Wiley & Sons, New York.
44. Kuggeleijn, P.M., 1995, Managing Data about Data, *GIS Europe*, Vol. 4, No. 2, p.32-33.
45. Kuggeleijn, P.M. and Padding, P., 1995, GIS Meta-Informatiebeheer bij Rijkswaterstaat, Geographical Information in the Netherlands, *First Joint European Conference and Exhibition on Geographical Information* (Proceedings Stream 7), The Hague, p.7.37-7.46.
46. Lanter, D.P., 1990a, The Problem of Lineage in GIS, in: Lanter, D.P., *Lineage in GIS: The Problem and a Solution,* Technical Report 90-6, National Center for Geographic Information and Analysis, University of California, Santa Barbara.
47. Lanter, D.P., 1990b, Design of a Lineage-Based Meta-Database for GIS, in: Lanter, D.P., *Lineage in GIS: The Problem and a Solution,* Technical Report 90-6, National Center for Geographic Information and Analysis, University of California, Santa Barbara.
48. Lanter, D.P. and Surbey, C., 1994, Metadata Analysis of GIS Data Processing: a Case Study, in: Waugh, T.C. and Healey, G., *Advances in GIS Research* (Proceedings of the Sixth International Symposium on Spatial Data Handling, Volume 1), Edinburgh, p.314-324.
49. *NAP-aanvraag Nationaal Clearinghouse Geo-information*, 1995, RAVI, World Wide Web, http://www.euronet.nl/users/ravi/secretar.html (February 28, 1996).
50. Nebert, D., 1995, *What does it mean to be an FGDC Clearinghouse Node?*, Draft Position Paper for FGDC Clearinghouse Working Group, Federal Geographic Data Committee, World Wide Web, http://fgdc.er.usgs.gov/nsdi/nsdinode.html (January 16, 1996).
51. Newman, I., 1991, Data Dictionaries, Information Resource Dictionary Systems and Metadatabases, in: Medyckyj-Scott, D. et al. (eds.), *Metadata in the Geosciences*, Papers derived from a symposium sponsored by the UK Association for Geographic Information, Midlands Regional Research Laboratory/Loughborough University of Technology/University of Leicester, Group D Publications, Loughborough (United Kingdom), p.69-83.
52. Oogen, J.H. van, 1995a, *CEN.TC 287*, (work program), World Wide Web, http://ilm.425.nlh.no/gis/cen/tc287 (January 25, 1996).
53. Oogen, J.H. van, 1995b, *Nationaal Clearinghouse Geo-informatie: Meta gegevens / Keuze standaard*, RAVI, Amersfoort (the Netherlands).
54. Oogen, J.H., 1996a, *Nationaal Clearinghouse Geo-informatie: Architectuur*, RAVI, Amersfoort (the Netherlands).
55. Petch, J., 1993a (3rd Ed.), *Concepts for Spatial Thinking*, Course Notes of Module 2 of the International Distance Learning GIS Diploma Programme, InterGIS, The Manchester Metropolitan University, Manchester.
56. Petch, J., 1993b (1st Ed.), Visualisation, Course Notes of Module 9 of the International Distance Learning GIS Diploma Programme, InterGIS, The Manchester Metropolitan University, Manchester.
57. RAVI, 1992, *Structuurschets Vastgoedinformatievoorziening*, RAVI, Rapport nr. 29, Apeldoorn.
58. RAVI, 1994a, *Referentiemodel Kwaliteit van Geo-informatie*, Rapport nr. 94-1, RAVI, Amersfoort (the Netherlands).
59. RAVI, 1994b, *Kwaliteit in Perspectief*, Overzicht van de behoefte aan registratie van kwaliteitskenmerken over geo-informatie in Nederland, Rapport nr. 95-1, RAVI, Amersfoort (the Netherlands).
60. RAVI, 1995a, Nationaal Clearinghouse Geo-informatie, *Vi-MATRIX*, Vol. 3, No. 8, p.19.
61. RAVI, 1995b, *Europese Standaardisatie met betrekking tot Geo-Informatie*, RAVI participatie in de Europese normalisatie commissie: CEN/TC 287, Rapport nr. 95-7, RAVI, Amersfoort (the Netherlands).
62. RAVI, 1995c, *Nationale Geo-informatie Infrastructuur (NGII)*, RAVI, Amersfoort (the Netherlands).

63. RAVI, 1996a, Nationaal Clearinghouse Geo-informatie, *Vi-MATRIX*, Vol. 4, No. 2, p.37.

64. Reeve, D., 1993, *Attribute Data*, Course Notes of Module 4 of the International Distance Learning GIS Diploma Programme, 3rd ed., InterGIS, The Manchester Metropolitan University, Manchester.

65. Reeve, D., and Cornelius, S., 1993. GIS in Organisations. Course Notes of Module 10 of the International Distance Learning GIS Diploma Programme, 1st ed., InterGIS, The Manchester Metropolitan University, Manchester.

66. Scholten, H. et al., 1994, Towards a spatial meta information system of Europe, *GIS in Business 94 Europe* Amsterdam (Proceedings), Longman GeoInformation, Cambridge, p.190-206.

67. Sjoberg, G. and Nett, R., 1968, *A Methodology for Social Research*, Harper & Row, New York.

68. Stein, C., 1995, *Metadata Validation Service*, World Wide Web, http://www-mel.nr/mry.navy.mil/mel-bin/meta-val (January 16, 1996).

69. Schweitzer, P., 1995, *Frequently Asked Questions*, World Wide Web, http://www.its.nbs.gov/nbs/meta/faqa.htm (January 16, 1996).

70. Thewessen, T.J.M. et al., 1995a, *Plan van Aanpak Pilot Clearinghouse* - "*Idefix*", Meetkundige Dienst Rijkswaterstaat/RIVM/Provincie Gelderland/Staring Centrum - DLO/RAVI (the Netherlands).

71. Thewessen, T.J.M. et al., 1995b, *GeoBase 2.1, Een gegevensmanagement- en meta-informatiesysteem voor Milieuonderzoek*, RIVM, Bilthoven (the Netherlands).

72. Tosta, N., 1995, National Spatial Data Infrastructure Activities within the United States, Geographical Information in the Netherlands, *First Joint European Conference and Exhibition on Geographical Information* (Proceedings Stream 7), The Hague, p.7.7-12.

73. Trivedi, N. and Smith, T.R., 1991, *A Conceptual Framework for Integrated Metadata Management in Very Large Spatial Databases,* Technical Report 91-2, National Center for Geographic Information and Analysis, University of California, Santa Barbara.

74. Wertz, C.J., 1986, *The Data Dictionary: Concepts and Uses*, Elsevier/North-Holland/QED Information Sciences, Wellesley (Massachusetts, USA).