1

METADATA IN GEOGRAPHIC AND ENVIRONMENTAL DATA MANAGEMENT

Oliver Günther* and Agnès Voisard**

**Institut für Wirtschaftsinformatik*
Humboldt-Universität zu Berlin
Spandauer Str. 1, D-10178 Berlin, Germany
`guenther@wiwi.hu-berlin.de`

***Institut für Informatik*
Freie Universität Berlin
Takustr. 9, D-14195 Berlin, Germany
`voisard@inf.fu-berlin.de`

ABSTRACT

Metadata is used increasingly in geographic and environmental information systems to improve both the availability and the quality of the information delivered. The growing popularity of Internet-based data servers has accelerated this trend even further. In this chapter we give an overview of metadata schemes and implementations that are common in this domain. Case studies include the Content Standards for Digital Geospatial Metadata of the U.S. Federal Geographic Data Committee (FGDC), and the Catalogue of Data Sources (CDS) of the European Environmental Agency. Another activity that we will discuss in somewhat greater detail concerns the UDK project, an international software engineering effort to facilitate access to environmental data. The UDK (Environmental Data Catalogue) is a public meta information system and navigation tool that helps users to identify and retrieve environmental data from the government and other sources. In 1995, first versions of the UDK were made available in Austria and Germany; several other European countries are currently evaluating the system. We will present the UDK data model, its implementation as a distributed information system, and its integration into the World Wide Web.

To appear in: W. Klas and A. Sheth (eds.), *Managing Multimedia Data: Using Metadata to Integrate and Apply Digital Data*, McGraw Hill, 1997.

1 INTRODUCTION

The preservation of the environment has become a major public policy goal throughout the world. Governments are concerned more than ever about their environmental resources and are establishing policies to control their consumption. Citizens take a greater interest in the current and future state of the environment and adapt their ways of living accordingly. As a result of these political developments, there is a major demand for environmental information and appropriate tools to manage it. Recent legislation reflects this trend. According to a recent directive of the European Union, for example, almost all environmental data that is stored at public agencies has to be made available to any citizen on demand [4]. As the last few years have shown, the tendency to exert this right is rising steadily. There is, for example, an increasing demand for up-to-date information on air quality in inner cities, on water quality in coastal regions, and so on. In addition, new legislation requires companies to provide an increasing amount of data about the environmental impact of their products and activities.

Given the amount and complexity of environmental data, these new information needs can only be served by using state-of-the-art computer technology. *Environmental information systems* are concerned with the management of data about the soil, the water, the air, and the species in the world around us. The collection and administration of such data is an essential component of any efficient environmental protection strategy. Vast amounts of data need to be available to decision makers, mostly (but not always) in some kind of condensed format. The requirements regarding the currency and accuracy of this information are high. While the information technology required for this task is rarely domain-specific, it is often important to select and combine the right tools among those that are available in principle. This requires a thorough knowledge of related developments in computer science, as well as a good understanding of the environmental management tasks at hand.

A particular need exists for convenient navigation aids that help users to take advantage of network-based, distributed information, regardless of their computer literacy. Starting from some environmental query or problem formulation, such navigation aids should help users to localize the relevant data sets and to retrieve them quickly and in a user-friendly manner. An essential prerequisite for both navigation and data transfer is the availability of appropriate *metadata*, i.e., data about the format and the contents of the data. The key idea is to enhance data sets by concise descriptions of themselves in order to improve both the speed and the accuracy of related search operations. The metadata serves as a kind of online documentation that can be read and utilized by appropriate tools as well as by human users. Note that there is no intrinsic distinction between data and metadata; it is rather a question of context whether a given data item represents metadata or not.

In this chapter we will discuss the question of metadata in geographic and environmental data management in greater detail. Section 2 gives a more elaborate definition of metadata and shows how metadata can be integrated into a traditional data management architecture. Sections 3 through 5 describe several concrete approaches to metadata management. Section 3 presents the U.S. initiative to create a National Spatial Data Infrastructure (NSDI); this includes discussions of the Spatial Data Transfer Standard (SDTS) and the FGDC Content Standards for Digital Geospatial Metadata. Sections 4 and 5 continue with descriptions of two European systems: the Catalogue of Data Sources of the Environment (CDS), and the German and Austrian proposal for a European Environmental Data Catalogue (UDK). Section 6 concludes with a summary and an outlook on future work.

2 METADATA AND DATA MODELING

Our further discussion is based on a *three-way data model* that distinguishes between environmental objects, environmental data objects, and environmental metadata (Fig. 1). The term *environmental object* is used to describe the real-world objects making up the environment. This includes natural entities, such as lakes and biotopes, as well as man-made objects, such as factories or highways. Nesting or overlaps between environmental objects are common. Each environmental object is described by a collection of *environmental data objects*, which are abstract entities that can be handled by computers or directly by decision makers. A typical environmental data object would be a series of measurements that captures the concentration of a certain substance in a river (the corresponding environmental object). Each environmental data object is in turn associated with one or more *metadata objects* that specify its format and contents. The documentation of the measuring series described above would be a typical example. It may include data about the spatial and temporal scale of the measurements, the main objectives of the project, the responsible agency, and so on.

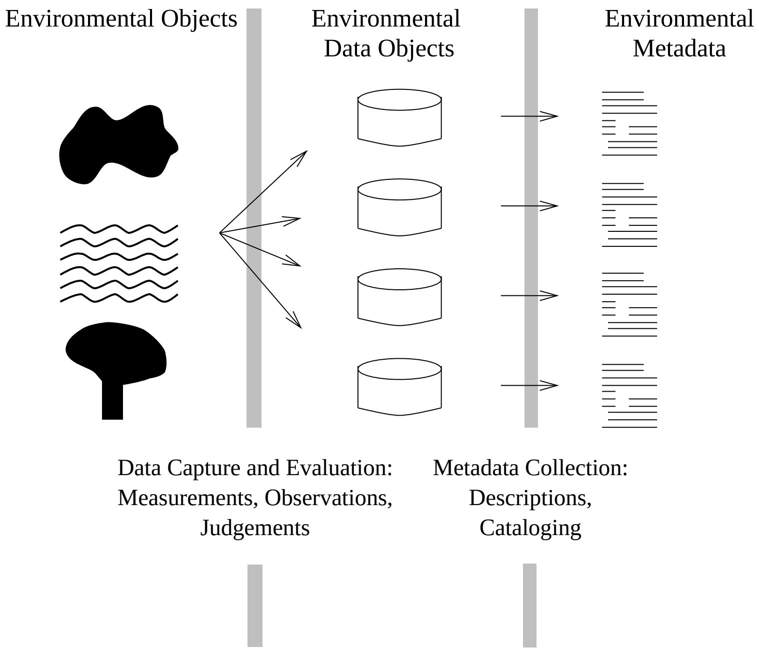


Figure 1 Three-way object model of environmental information systems

The data flow in many environmental applications closely resembles the data flow in classical business applications. It can be structured into four phases: data capture, data aggregation, data storage, and data analysis.

1. The first phase, *data capture*, concerns the collection of environmental raw data, such as measurement time series or aerial photographs. In this phase the great variety of environmental objects is mapped onto a collection of environmental data objects, which have a structure that is much simpler and more clearly defined. There are a variety of ways to perform such a mapping, including measurement and observation, but also value-based judgement.
2. In the second phase, *data aggregation*, this raw data is condensed and enriched in order to extract entities that are semantically meaningful. In the case of image data, for example, this includes the recognition of geometric primitives (such as lines and vertices) in an array of pixels, the comparison of the resulting geometric objects with available maps, and the identification of geographic objects (such as cities or rivers) on the

picture. The information can then be represented in a much more compact format (in this case, a vector-based data format, as opposed to the original raster data). Measurement time series also need to be aggregated and possibly evaluated by means of some standard statistical procedures. The aggregated data is then stored in a file or a database.

3. In the third phase of *data storage*, one has to choose a suitable database design and appropriate physical storage structures that will optimize overall system performance. Because of the complexity and heterogeneity of environmental data, this often necessitates substantial extensions to classical database technology.
4. In the final *data analysis* phase, the available information is prepared for decision support purposes. This may require access to data that is geographically distributed, stored on heterogeneous hardware, and organized along a wide variety of data models. The data analysis is typically based on complex statistical methods, scenarios, simulation and visualization tools, as well as institutional knowledge (environmental legislation, user objectives, etc.). Only the synthesis of these different inputs allows us to judge the state of the environment and the potential of certain measures, both planned and already implemented.

The overall objective of this complex data flow is to provide decision support at various levels of responsibility. Figure 2 uses the symbol of the pyramid to visualize this idea. The last three phases of the data flow correspond to a bottom-up traversal of the pyramid. Data can be used throughout that traversal for decision support purposes. While the data in the lower part of the pyramid tends to be used for local, tactical tasks, the upper part corresponds to strategic decision support for the middle and upper management.

Metadata may be *collected* at any of the four phases of the data flow and built into the corresponding data structures. As Kashyap et al. point out in this volume [18], much of the data produced during data aggregation and storage is already metadata, starting with simple database schema information up to high-level semantic abstractions of the available data sets. While collection has been mostly manual so far, the automatic extraction of metadata is increasingly becoming an option; Drew and Ying give a concrete example in another chapter of this volume [3].

As for the *use* of metadata, it is mainly taking place in the data analysis phase and fulfills a variety of purposes:

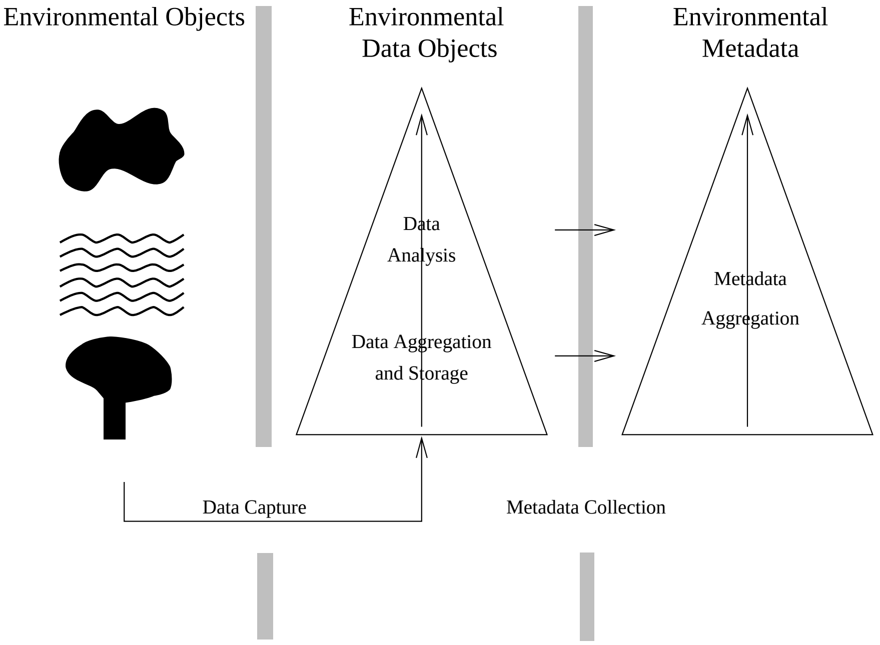


Figure 2 Data flow in environmental information systems

- Computerized environmental information systems are able to collect and process much greater amounts of data than anybody could have thought of only a few years ago. Automatic data capture and measurement results in terabytes of new data per day [2]. Even in processed form, this kind of data is impossible to browse manually in order to find the information that is relevant for a given task. Modern information retrieval tools allow the automatic or semi-automatic filtering of the available data in order to find quickly those data sets one is looking for. Metadata forms an important foundation of these tools by serving as a condensed representation of the underlying data. As such, it supports browsing, navigation, and content-oriented indexing.

- Environmental data management is extremely *heterogeneous*, both in terms of hardware and software platforms. Data is organized according to a wide variety of data models, depending on the primary objectives of the particular agency in charge. Metadata can help to overcome these heterogeneities by specifying the platforms on which a given data item is located. This way, appropriate conversion routines can be introduced (semi-)automatically, wherever necessary.

- Environmental data is frequently *uncertain*. Metadata can be used to specify the accuracy of a data item, so users can judge from the metadata whether the corresponding environmental data objects are relevant for their current needs.
- Metadata can also help to *inventory* existing data holdings, to unify naming schemes, and to record relationships between different data items and data sets. This aspect of metadata has recently become very popular as one of the core functionalities of *data warehouses* [16].

The concept of metadata is not new. Online documentation of programs and data sets has been in common use for many years. Machine-readable metadata has also been known for a long time, in particular in the context of relational databases, where the internal database structure (the *database schema*) is typically represented in a relational format itself. What is new, is the more systematic approach to providing machine-readable metadata, and the trend to standardize metadata in certain application areas.

For the subsequent discussion, it is useful to distinguish between two kinds of metadata [28]. The term *denotative metadata* is used to refer to the kind of metadata that describes the logical structure of a data set; a relational schema would be a typical example. The term *annotative metadata*, on the other hand, is used to describe data that provides content-oriented context information, such as the documentation of the measuring series described above. Following Melton et al. [28], further examples of annotative metadata include "information in scientific notebooks, instrument logs, manuals, and reports that document the platform and instrument conditions, the operational environment, interferring sources of noise, and that uniquely identify the software and computer platforms used for analysis, modelling and simulation." In the remainder of this chapter, we will concentrate on annotative metadata and use the term "metadata" in that sense. It should be noted here that other researchers have presented different ways to classify metadata; Kashyap et al. give an overview of related work in this volume [18].

The relevance of metadata for the management and analysis of complex data sets has been pointed out early on by McCarthy [26] and pursued further in the area of statistical and scientific databases. Siegel and Madnick [36] built on those ideas, concentrating on possible applications in financial data analysis. The IEEE Mass Storage Systems and Technology Committee has sponsored several metadata workshops whose results are available on the Web (URL `http://www.llnl.gov/liv_comp/metadata/metadata.html`). The use of metadata in geographic and environmental information systems is of a more recent nature [34]. Lately, however, there has been broad agreement that metadata are a crucial factor to improve both the quality and the availability of geographic and environmental data. Several conferences on spatial databases and geographical information systems (GIS) have devoted parts of their program to metadata [15, 5, 6], and there has been a variety of workshops dedicated exclusively to metadata management in the geosciences and the environmental sciences [27, 28].

In terms of practical consequences, metadata technology is increasingly being integrated into commercial GIS. Most commercial systems have always maintained some basic metadata on the objects to be administered. *ARC/INFO*, for example, generates and maintains metadata on the spatial registration, projection, and tolerances of a coverage or grid [7]. Every time one creates a coverage, the system creates a set of metadata files, including the TIC file (containing data about the coverage's coordinate registration), the LOG file (tracking all ARC operations performed on the coverage), and the BND file (containing the coordinate values that denote the outer boundary or spatial extent of your coverage). There is also denotative metadata giving some schema information of the INFO tables that contain the non-spatial data components.

The practical use of metadata, however, is extending much beyond this somewhat narrow scope. One trend is to collect more information about the detailed content of the data. Vendors typically choose some bibliography-style format to represent this information; conformity with the FGDC Content Standards (see Section 3.2) is increasingly required. The ARC/INFO component *DOCUMENT.AML* [7] is a typical example of such a tool.

Another trend is to describe the history and quality (also called *lineage*) of data sets and their sources in more detail. *Geolineus* of *Geographic Designs Inc.* is a typical tool for this purpose [12]. Geolineus represents the data in a GIS by means of dataflow diagrams, where coverages and grids are shown as icons. Icons along the top of the diagram represent the *source data* on which the GIS is based. Icons further down represent data layers that were *derived* with spatial analysis operations like *BUFFER* or *INTERSECT*. Finally, icons at the bottom of the diagram represent *products*, i.e., derived data items that represent the final steps in a GIS application. Geolineus shows the type of data in the corresponding layer for each icon and maintains command histories for each coverage. The system allows to store documentation about each layer in a frame-based format.

In this volume, Drew and Ying describe a concrete approach to use metadata in order to provide uniform access to a heterogeneous collection of GIS and spatial databases [3]. Based on metadata about those systems and their contents, their *GeoChange* system serves as a navigation and access tool. To a large extent, it is non-intrusive, i.e., it can be implemented on top of an existing collection of independent systems without major changes to the underlying architectures and implementations.

Other trends in metadata management include the inclusion of more spatial elements in the metadata itself [35] and the use of metadata to describe and access not only other data sets, but also models and algorithms [23].

Parallel to these application developments, metadata management has become a focus in an increasing number of government R&D projects. Besides the efforts described in the following sections, there has been a project by the European Space Agency (ESA) to develop an online geosciences metadata system, called the *ESA Prototype International Directory* [39]. At about the same time, the United Nations Environmental Program (UNEP) has started a project on the *Harmonization of Environmental Measurements (HEM)* [19]. Also actively involved in the harmonization of environmental data in research and monitoring is the International Council of Scientific Unions (ICSU), represented by its Scientific Committee for the Problems of the Environment (SCOPE) and its Committee on Data for Science and Technology (CODATA) [1]. The Norwegian *SAMPO* project uses ARC/INFO's *ArcView* to catalogue its spatial data holdings [29]. The Austrian Ministry of the Environment has developed a *Central European Environmental Data Request Facility (CEDAR)* [33]. Other efforts include the *CIMI* system of the Dutch Ministry of Transportation, Public Works and Water Management [22], the Australian *FINDAR* system [17] and the New South Wales Department of Conservation and Land Management's *Data Directory* [31, 30].

Coordination between this great variety of efforts is difficult. As we will show in Section 4, the newly founded European Environmental Agency will have an important role to play here. One promising effort concerns the development of a common European geodata standard. With strong support from the European Center of Normalization, Germany's and Belgium's Geographic Data Files (GDF) are generally considered the frontrunner [32]. Further standardization is required, however. Environmental phenomena do not stop at national borders. In this domain, international cooperation on a broad scale is essential for making progress.

3 THE U.S. NATIONAL SPATIAL DATA INFRASTRUCTURE (NSDI)

Since the early 1980s, the U.S. Government has been working intensively on creating a National Spatial Data Infrastructure (NSDI). A major motivation for this effort was to abolish the notorious incompatibilities among the internal formats used by various government agencies. Examples include DLG, TIGER/Line, and GRASS of the U.S. Geological Survey, DIGEST and the Vector Product Format (VPF) of the Defense Mapping Agency (DMA), and DX90 of the National Ocean Service. The parallel use of such a variety of standards led to considerable expenses to the taxpayer that could at least in part have been avoided.

Most of the early efforts on NSDI were coordinated by the U.S. Geological Survey, an agency under the supervision of the U.S. Department of the Interior. One of the first major results was the development of the *Spatial Data Transfer Standard (SDTS)*, a Federal Information Processing Standard to facilitate the online exchange of spatial data [37]. The goal is to accommodate different spatial data models, to preserve topologies, and to maintain even complex relationships, as data is transferred across different computer platforms and software systems. Other than many existing standards (such as VPF), the SDTS is not an exchange format. It rather provides guidelines that need to be translated into a native application-specific format before they can be used. Most GIS vendors provide interfaces and tools for that purpose [8].

Since 1990, the NSDI efforts have been coordinated by a working group called the *Federal Geographic Data Committee (FGDC)*, which is composed of representatives of the Departments of Agriculture, Commerce, Defense, Energy, Housing and Urban Development, the Interior, State, and Transportation; the Environmental Protection Agency; the Federal Emergency Management Agency; the Library of Congress; the National Aeronautics and Space Administration; the National Archives and Records Administration; and the Tennessee Valley Authority. The committee is chaired by the Department of the Interior, represented by the U.S. Geological Survey.

In May 1994, the FGDC published a draft for the new *Content Standards for Digital Geospatial Metadata* [11], which was later approved by the National Institute of Standards and Technology as a Federal Information Processing Standard. The implementation of the standard is based on the Executive Order 12906, "Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure," which was signed on April 11, 1994, by President

Clinton [38]. In addition to providing a long-needed political foundation for the NSDI, the order requires all government agencies to use the FGDC Content Standards for documenting all new geospatial data it collects or produces as of April 11, 1995.

While both the SDTS and the FGDC Content Standards refer to metadata about spatial data, they have distinctly separate functions. The SDTS is a language for communicating spatial data across different platforms without losing any structural or topological information. The FGDC Content Standards, on the other hand, specify the kind of annotative metadata that federal agencies are required to collect on a spatial data set they maintain. The only two sections that both standards have in common concern data quality and the data dictionary information; we will discuss this in detail later on.

3.1 The Spatial Data Transfer Standard (SDTS)

The Spatial Data Transfer Standard (SDTS) [37] was designed to facilitate the online transfer of the full range of geographic and cartographic data. Both vector and raster data of a large variety of data models can be exchanged across heterogeneous hardware and software platforms using the SDTS. The standard is structured into three main parts; the subsequent presentation follows the overview of Fegeas et al. [10].

1. *Logical Specification*

   This part contains the logical specification of the entities and data objects used to describe different GIS data models. It consists of three major sections in turn and provides guidelines on how spatial and nonspatial objects (simple or composite) are to be organized, named, and structured.

   The first section presents a conceptual model of spatial data. It describes the real world as a set of "entities" (cities, rivers, factories, etc.), each characterized by attributes, which are assigned attribute values. The model then goes on to define a set of zero-, one-, and two-dimensional spatial objects (such as points, lines, and polygons) and the relationships between entities and spatial objects.

   The reader should take notice of the particular use of the term "entity," which in this standard has been chosen to describe a real-world phenomenon, whereas the term "object" is reserved for the digital representation of an "entity." In analogy to the standard entity-relationship literature, the term "entity type" is used to describe a set of similar entities; in that context the single entities are also called "entity instances." The term "feature," finally, which is still very common in the geoscientific community, is here defined as both a real-world entity and its object representation, i.e., as the superclass of the classes "entity" and "object."

The second section of part 1 is devoted to data quality. It specifies five portions of a data quality report: lineage, positional accuracy, attribute accuracy, logical consistency, and completeness. The lineage portion describes source and update material (with dates), methods of derivation, transformations, and other processing history. Positional accuracy is concerned with how closely the locational data represent true locations. Attribute accuracy is similarly concerned with non-locational descriptive data. Logical consistency refers to the fidelity of encoded relationships in the structure of the spatial data (e.g. the degree to which topological relationships have been verified). The completeness portion includes information about geographic area and subject matter coverage. Note that large parts of this second section of part 1 are replicated in feature group 2 (data quality information) of the FGDC Content Standards.

The third section of part 1 constitutes the largest portion of the whole standard; it specifies detailed logical transfer format constructs and specifications for SDTS transfer data sets. An SDTS transfer is organized into modules with records, fields, and subfields. Thirty-four module types are specified as detailed field and subfield record layout specification tables, designed to include many kinds of information: global, data quality, feature and attribute data dictionary, coordinate reference, spatial object, and associated attribute and graphic symbology information. The data dictionary portion, which conveys the meaning and structure of entity and attribute data, is divided into three module types: definition, domain, and schema. Parts of the data dictionary portion are replicated in feature group 4 (entity and attribute information) of the FGDC Content Standards.

2. *Data Content Registry*

   This part provides data content standards by specifying a model for the definition of spatial entity types, attributes, and attribute values. The underlying idea of this part of the standard is that there is a need for common definitions of spatial features (resp. entities). In that sense, this part is nothing but a thesaurus. It contains a list of about 200 topographic and hydrographic entity types with 244 attributes, plus a list of about 1200 terms that are in a synonym or subtype relationship to any of those standard or primary terms. It is foreseen by the designers of the standard that this section will be subject to continuous updates and extensions.

3. *Physical Structure*

   This part specifies the implementation of the transfer using the ISO 8211 international standard for information interchange. The ISO standard itself is embedded into the SDTS to ensure that data can be transferred to any computing environment. The U.S. Geological Survey has developed a public domain software function library to assist in encoding and decoding SDTS data into ISO 8211 format.

   It is important to keep in mind that the SDTS and ISO 8211 are separate standards. ISO 8211 is an international data exchange format that can be used to transfer any type of data, not just spatial data. ISO 8211 provides a means of transferring data records and their description across heterogeneous hardware and software platforms. It requires, however, that the *content* and the *meaning* of the data records are defined by the user. In that sense, the SDTS can be considered a user of ISO 8211.

   The SDTS is designed such that Parts 1 and 2 are independent of part 3, which is specific to ISO 8211. If necessary, the SDTS could replace part 3 by another version that uses a different implementation format without affecting parts 1 and 2. ISO 8211 was chosen so that the SDTS could use an existing general-purpose transfer standard rather than having to develop a new SDTS-specific format. It is designed to work for any media, including communication lines. ISO 8211 is self-describing. An ISO 8211 file (called a *Data Descriptive File (DDF)*) contains both data and the description of the data. The *Data Descriptive Record (DDR)* is fixed; it contains the structure and description of the data. The *Data Records (DRs)* are of variable size; they contain the actual data. There is always one DDR in a file, and one or more DRs.

Given the great complexity of the standard, the designers also introduced a concept called *profile*, which is a kind of customization of the standard for a particular data model. If a new data model is to be supported, the interested parties may specify those options of the standard that are needed to support that data model. This subset of options can then be submitted for approval as its own Federal Information Processing Standard and, once approved, is added to the SDTS as a new SDTS profile.

Currently, there exists a *Topological Vector Profile (TVP)* for vector data with full and explicit topology. Another profile that is about to be approved is a raster profile for image and gridded data. Under consideration are further vector profiles for network/transportation data, for nontopological nautical chart and hydrographic data, and for CAD data.

3.2 The FGDC Content Standards for Digital Geospatial Metadata

The FGDC Content Standards define metadata as *data about the content, quality, condition, and other characteristics of data.* They structure the spatial metadata into the following seven groups of features. Only the first (identification information) and the last feature group (metadata reference information) are obligatory; the remaining ones are optional.

1. *Identification Information*

   This feature group contains the basic meta information about a given data set, including:

   - Textual description
   - Information about the time period described
   - Spatial reference: A minimum bounding rectangle is required. Optionally, one can provide a more detailed polygonal description.
   - Keywords: They can be freely chosen, but need to be associated with a term from the relevant thesaurus. One keyword about the theme of the data set is obligatory. Optionally, one can provide further keywords that refer to the theme, the space, or the time corresponding to the data set in question.
   - Person or organization to contact for more information about the data set (optional)
   - Access constraints and security information (optional)
   - Information about the technical representation of the data set: special software, operating system, file name, data set size (optional)

2. *Data Quality Information*

   This feature group contains general information about the quality of the data set. In addition to an assessment of the accuracy and consistency of the data, this includes metadata about the data source ("lineage") and about completeness.

Note that this feature group replicates the content (but not the structure) of the SDTS's data quality section (part 1, second section).

3. *Spatial Data Organization Information*

   This feature group contains information on which mechanism was used to represent spatial information in the data set. At this point the standard supports a generic mechanism to represent raster data, and SDTS and VPF to represent vector data. The SDTS section is based on part 1 of the SDTS specification.

   The fact that both SDTS and VPF were included explicitly shows how the designers of the standard sometimes had to sacrifice conciseness and clarity in order to obtain approval from all participants. It was not possible to move all government agencies towards a single standard for representing vector data. Among other reasons, this is mainly due to large amounts of essential legacy data, whose conversion would exceed the available resources of the respective agencies.

4. *Spatial Reference Information*

   This feature describes the projection and coordinate system used (e.g. *Mercator* or *Miller_Cylindrical*).

5. *Entity and Attribute Information*

   This feature group allows the user to describe the information content of the data set using the entity-relationship model. The SDTS's data dictionary information is captured in this feature group. There is common agreement that this section of the standard is too superficial and should be redesigned in future versions of the standard.

6. *Distribution Information*

   This feature group contains information about the distributor of the data set and about options for obtaining it. The distributor usually corresponds to the contact person/organization listed in the identification information (see 1.). The order information includes data about the possible modes of communication (modem, e-mail, etc.) and about the transfer formats used (e.g. the ARC/INFO Export format, the Initial Graphics Exchange Standard (IGES), or ASCII).

7. *Metadata Reference Information*

   This obligatory feature group serves for storing what could be called "meta-metadata." This includes information about the last update of the metadata, the latest and the next review of the metadata, the party responsible for the metadata, as well as access and security constraints.

In summary, the FGDC Content Standards represent an impressive effort to establish a uniform way to document digital geospatial data sets. While mainly targeted at the description of geographic data, it also provides a solid basis for an environmental metadata system. Such an extension would entail a more detailed semantic framework, especially with regard to theme-related information.

4 THE CATALOGUE OF DATA SOURCES FOR THE ENVIRONMENT (CDS)

The European Union (EU) has been working on similar issues, especially since the 1994 foundation of its European Environmental Agency (EEA), located in Copenhagen. In comparison to the American activities, the EEA efforts have a wider focus, concentrating not only on spatial data, but on environmental data in a more general sense. On the other hand, the results obtained so far are not nearly as concrete as the FGDC recommendations described above.

The ultimate goal of the EU activities is the implementation of an integrated European environmental information system. Based on the results of a previous project called *CORINE CDS* (1985-1989), the EU recently commissioned a study entitled "Catalogue of Data Sources for the Environment - Analysis and Suggestions for a Meta-Data System and Service for The European Environment Agency" [9]. An essential result of this study was the (hardly surprising) insight that the construction of a European environmental information system from scratch is neither economically feasible nor politically viable. Many member countries already have some kind of national environmental information system. A European system should take advantage part of these developments and attempt a bottom-up integration of the systems that are already functional. Devised as a *meta* information system, CDS would only store descriptions of data sets that are locally available.

The study recommends the simultaneous realization of the following two architectures:

- a standalone variant that is updated periodically based on current information from the member countries;
- a networked variant, which has on-line connections with a variety of national catalogues and which is only usable in connection with those.

Since the study was written (1993), the percentage of computers that are networked, usually including some connection to the Internet, has increased considerably. The first architecture option seems therefore obsolete. In turn, it should be made sure that the central catalogue provides some base functionalities independently of the current state of the national catalogues and the connections to them. This can easily be achieved by making local copies of a subset of the metadata periodically. For distribution and update purposes, the study recommends the usage of CD-ROMs. Once again, the usage of the Internet instead will probably be a matter of course by the time a CDS system will be operational. The data should be stored in a relational database system, with text fields playing an important role. The language problem shall be alleviated, if not solved, by using a multilingual thesaurus. The GIS functionalities of the proposed CDS system are only rudimentary; more complex spatial functionalities are referred to an external GIS instead.

The study does not propose a concrete format for the metadata, comparable to the detailed specifications of the FGDC or the UDK (see Section 5). The authors suggest instead to form some synthesis of the existing proposals of the member countries and of the United States. Of course, such a fusion is bound to produce semantic discrepancies and even incompatibilities. To minimize those, the study proposes to focus the synthesis on eight major classes of entities. The three most important classes listed in the study are

1. Institution
2. Activities/Projects
3. Products

The remaining five entity classes serve to represent secondary information about the entities in classes 1, 2, and 3:

4. Addresses
5. Stations
6. Communication
7. People/Persons
8. Data Sets

It seems somewhat questionable whether a single-layer taxonomy like the one above would ever be able to capture the extreme heterogeneity that resulted from a synthesis of the environmental data and metadata schemes throughout Europe. On the one hand, there will always be entity types that do not fit into the given scheme. On the other hand, there has to be a formal mechanism to refine a given entity class in order to serve the local requirements of a particular agency in an optimal manner. A multi-layer taxonomy, i.e., a class hierarchy with an inheritance mechanism seems to be much better suited for this purpose. The UDK system described in the following section is an example where such a class hierarchy approach has been introduced successfully.

5 THE UDK - A EUROPEAN ENVIRONMENTAL DATA CATALOGUE

The *UDK (Umwelt-Datenkatalog = Environmental Data Catalogue)* is a meta information system and navigation tool that documents collections of environmental data from the government and other sources. These data sets may be available either online or by request to the responsible data administrator. Potential users of the system include government agencies, industry, as well as the general public. The UDK helps them to get answers to the following questions:

- Which relevant information is principally available for a given problem?
- Where is this information stored?
- How can this information be retrieved?

The UDK design presented in this section is the result of several years of research and development [24, 25]. In 1990, the Environmental Ministry of the State of Lower Saxony launched a research project with funding from the German Federal Environmental Protection Agency. Two years later, an international working group was formed to oversee the UDK design and its further development into a practical software tool. In 1994, Austria passed an Environmental Information Law that introduced the UDK as the official navigation tool for all environmental information on record. In 1995, the first version of the UDK was made available in Austria and the German states of Baden-Württemberg and Lower Saxony; other German states will follow. The UDK is currently also under evaluation by several other European countries, including Switzerland, Italy, Sweden, and Norway.

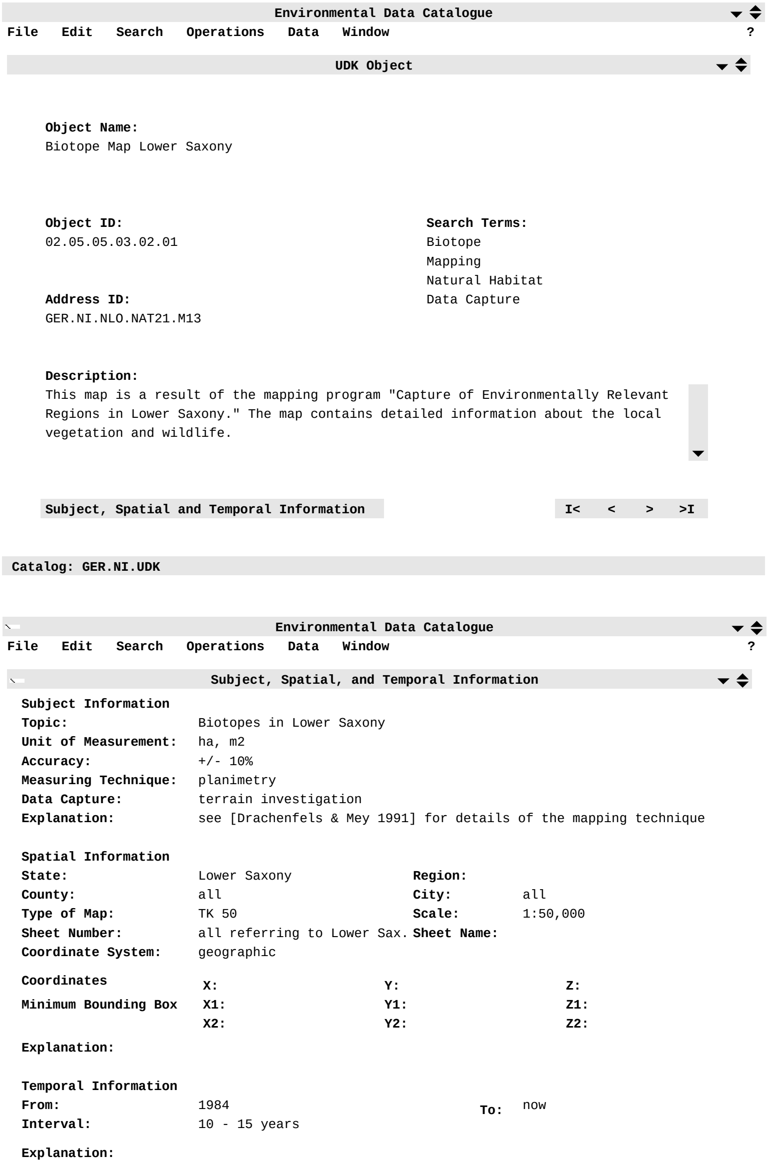
5.1 The UDK Object Model

The UDK is based on a three-way object model that is very similar to the data model described in the introduction (Fig. 1). In the UDK we distinguish between environmental objects, environmental data objects, and UDK (meta) objects. Each real-world environmental object is described by a collection of environmental data objects. Each environmental data object is in turn associated with exactly one metadata object that specifies its format and contents.

On the screen, each such *UDK object* is represented by one or more screen layouts; see Figure 3 for an example. The first screen layout contains some administrative information (object name, object ID, and keywords), a text description, and the address of the agency that is responsible for the maintenance of this UDK object and the underlying environmental data object. The second screen layout contains some more technical information about the environmental data object. This includes detailed data about the information content, the capturing method and its accuracy, the spatial extent, and the validity of the object. Spatial information can be specified using either coordinates, or (as in this example) denominations of administrative entities.

UDK objects may exist for environmental data objects at various aggregation levels simultaneously. Consider, for example, a national groundwater database that contains a large number of measurements from all over the country. There is one UDK object representing this database as a whole. In addition, however, there may be one UDK object each for the measurements from a certain county, there may be UDK objects representing the measurements from a particular station, and there may even be UDK objects that represent single measurements. There may also be UDK objects for groupings that are orthogonal to this primary aggregation hierarchy, such as UDK objects representing the measurements that were taken in a given month.

There are two reasons for this great flexibility in defining UDK objects at various levels of aggregation. First, powerful aggregation facilities are crucial for improving the usability and acceptance of a system like the UDK. Empirical studies have shown that the overwhelming number of queries in such a context refer to aggregated data rather than detailed source data. For example, citizens may be concerned about the ozone concentration in their neighborhood on a certain day; it is rather unlikely that they would want to know the exact concentration at a certain measuring station at an exact time. Second, aggregation semantics differ greatly between different user communities. Some people may have to aggregate over time, others over space, and yet others by topic. In



Object Name:
Biotope Map Lower Saxony

Object ID:
02.05.05.03.02.01

Address ID:
GER.NI.NLO.NAT21.M13

Search Terms:
Biotope
Mapping
Natural Habitat
Data Capture

Description:
This map is a result of the mapping program "Capture of Environmentally Relevant Regions in Lower Saxony." The map contains detailed information about the local vegetation and wildlife.

Catalog: GER.NI.UDK

Subject Information

Topic:	Biotopes in Lower Saxony
Unit of Measurement:	ha, m2
Accuracy:	+/- 10%
Measuring Technique:	planimetry
Data Capture:	terrain investigation
Explanation:	see [Drachenfels & Mey 1991] for details of the mapping technique

Spatial Information

State:	Lower Saxony	**Region:**	
County:	all	**City:**	all
Type of Map:	TK 50	**Scale:**	1:50,000
Sheet Number:	all referring to Lower Sax.	**Sheet Name:**	
Coordinate System:	geographic		

Coordinates	**X:**	**Y:**	**Z:**
Minimum Bounding Box	**X1:**	**Y1:**	**Z1:**
	X2:	**Y2:**	**Z2:**

Explanation:

Temporal Information

From:	1984	**To:**	now
Interval:	10 - 15 years		

Explanation:

Figure 3 Two screen layouts representing a UDK object

order to appeal to a large user community, the UDK system must be able to accommodate those different needs.

Although it is therefore desirable to handle the creation (and deletion) of UDK objects with great flexibility, the decision to create a new object has to be based on a cost/benefit analysis, depending on the particular applications a user has in mind. The effort to create and maintain a UDK object is not negligible. Recent empirical data suggests that creation takes one person-day in the average. Maintenance involves not only the occasional update of attributes but also the dynamic tracking of semantic associations between UDK objects and the corresponding environmental data objects; see Section 5.3 for further details. At this time, most of the related work is performed by specialized personnel from higher-level government agencies or consulting firms, and therefore relatively expensive. It is unlikely that the work can be delegated to less qualified support staff in the near future. The idea to leave the creation of UDK objects to local domain experts (biologists, chemists, etc.) is also unrealistic at the present time. The process is still too technical and time-consuming for someone who is not a UDK expert.

Up to now, UDK objects have been identified by their position in the *primary tree*, a directed graph whose nodes correspond to the UDK objects and whose edges represent responsibilities of agencies and departments for particular sets of UDK objects, as well as part-of-relationships between large data collections (e.g. a groundwater database) and their components (e.g. the data sets corresponding to particular measuring stations). This approach to identify objects is unsatisfactory for a variety of reasons. Most importantly, UDK objects may lose their identity when they are relocated in the primary tree due to some reorganization (such as the transfer of a department from one ministry to another). In this case, the objects that were relocated have to be recreated under a new ID at the new location. As an alternative, we are currently investigating the possibility of using *object identifiers (OIDs)*, a concept well-known from the domain of object-oriented databases. OIDs are created by the system; they are usually not visible to the user. To guarantee universal uniqueness, the generation of the OID is usually based on the CPU number, as well as the current date and time-of-day.

5.2 UDK Object Classes and Inheritance

To structure the wide variety of UDK objects, and to facilitate both their capture and their administration, we recently presented the first proposal for a *UDK class concept* [14]. There we distinguish between seven classes of environmental data objects:

1. project data (construction projects, environmental impact studies, etc.)
2. empirical data (measuring series, laboratory data, etc.)
3. data about facilities (factories, buildings, etc.)
4. maps
5. expertises and reports
6. product data
7. model data (simulations, etc.)

For each of these seven classes of environmental data objects there is a corresponding UDK class that contains the UDK objects describing them. Each UDK class corresponds to a screen layout that is used for the capture and administration of the corresponding UDK objects. The basis for this pragmatic proposal were the user requirements that were stated during the first few months of UDK data capture. Obviously, this classification needs to be reviewed and possibly extended from time to time to reflect changes in user requirements. We feel it is important, however, that the above top-level classification reflects a consensus of all UDK participants.

Another extension that is currently planned concerns the *vertical* structure of this classification. In particular, we intend to turn this flat class structure into an object-oriented class hierarchy that allows the inheritance of object attributes. The hierarchy should be structured as follows.

- The root of the hierarchy (level 0) consists of the generic class *UDK_Object* with four obligatory attributes: the unique object identifier (OID), the object name, the date when the object was last modified, and the agency (or the person) that is responsible for the object. Optional attributes, such as a textual description, may be included as well. Note that this generic class is not an abstract class, i.e., it may contain objects that are not included in any of its subclasses.
- Level 1 contains a relatively small number of classes that represent a consensus between all UDK participants. Currently, this level corresponds to the seven classes described above. Changes at this level are subject to negotiation between the UDK member countries.
- On the subsequent levels of the hierarchy, participating countries or agencies are free to introduce additional subclasses depending on their particular requirements. This kind of flexibility is important not only for

> efficiency reasons but also for gaining acceptance throughout the intended UDK user community, especially in government agencies at the national and local levels.

Class attributes are inherited along this class hierarchy in an object-oriented manner. This includes the possibility to upgrade selected attributes from being optional to being required. It also means that attributes that are specific to a certain subclass, but not to its superclass(es), can be masked out when looking only at the superclass. For example, consider a particular topographic map m and its UDK object U_m. m is an element of the class *topographic_map*, which is a subclass of the class *map*. If one now looks at the UDK object U_m through the screen layout corresponding to the class *map*, one only sees the attributes of *map*. The additional attributes that may have been introduced to describe *topographic* maps (as a special case of general maps) are not visible in this case.

This feature, which is typical for object-oriented environments, is a crucial element of standardization in the presence of application-specific extensions on the class hierarchy levels 2 and below. Any tool that is supposed to work at the national (or international) level across particular agencies or user communities can rely on the availability of the attributes defined at level 1. Maintenance and version management are other issues that need to rely on a stable class and attribute structure at the higher levels of the object hierarchy. It is therefore important to take organizational and technical precautions to make sure that users observe this principle throughout user-specific extensions and increasingly complex class structures. The technical details of the implementation of these lower hierarchy levels are still under discussion.

5.3 Semantic Associations Between UDK Objects

Orthogonal to the class hierarchy described in the previous section, the UDK offers users the ability to connect *concrete UDK objects* with each other in a hypertext fashion. The resulting structures are directed graphs whose nodes correspond to UDK objects and whose edges represent semantic associations between them or between their respective environmental data objects. The semantics of those edges may vary; we will later propose a type system for edges to make this aspect more explicit. Note that those semantic nets are completely independent of the class hierarchy described in the previous section.

While the nodes of the class hierarchy are *UDK object classes*, the nodes of the structures described in the following represent *concrete UDK objects*.

The most important graph structure is the *primary tree* or *primary catalogue*. Each UDK object corresponds to exactly one node of this tree structure, i.e., there is a 1:1 relationship between primary tree nodes and UDK objects. The links in the upper part of the tree serve to represent responsibilities of agencies and departments for particular sets of UDK objects. The agency that is in charge of a UDK object has to make sure that its information is correct and up-to-date. It is also responsible for the creation and deletion of UDK objects in the associated subtree(s). In the lower part of the tree, the links are used to represent part-of relationships between large data collections (e.g. a groundwater database) and their components (e.g. the data sets corresponding to particular measuring stations). The example given in Figure 4 depicts the UDK objects related to a groundwater database. Here the solid arrows make up the primary tree; their semantics varies between "is-responsible-for" (in the upper part of the tree) and "is-an-aggregation-of" (in the lower part).

Depending on particular user requirements, there may also be *secondary catalogues* to represent other semantic associations. Like the primary tree, a secondary catalogue is a directed graph whose nodes each correspond to exactly one UDK object. Other than in the case of the primary tree, however, the resulting structure does not have to be a tree. Note also that a UDK object can be referenced by any number of secondary catalogues. There is a 1:n relationship between UDK objects and secondary catalogue nodes: each UDK object can be a node in any number of secondary catalogues, but each secondary catalogue node refers to exactly one UDK object.

A typical application of a secondary catalogue concerns the representation of additional aggregation relationships that are not represented in the primary tree. In Figure 4 these kind of associations are pictured as dotted arrows. These kind of links are often useful to refer users to relevant aggregated data sets first before, upon request, giving them access to more detailed data. Another application of secondary catalogues is the construction of personal association structures. The "debate" association in Figure 4 (dashed line) is an example of such a structure. For such structures the system does not require users to restrict themselves to a tree structure. Similar to the freedom one has for linking pages in the World Wide Web, any directed graph structure is permitted, including graphs with cycles. The idea is to give UDK users maximum flexibility to connect and associate the various information items making up their working environment. With an attractive user interface, this option should be of great interest to a large group of users. What is important is that it has to be reasonably easy to create personal UDK objects and links. Furthermore, it is essential that those "personal" structures can be isolated from the public part of the UDK, so users can build confidential structures that are visible just for them or for their team.

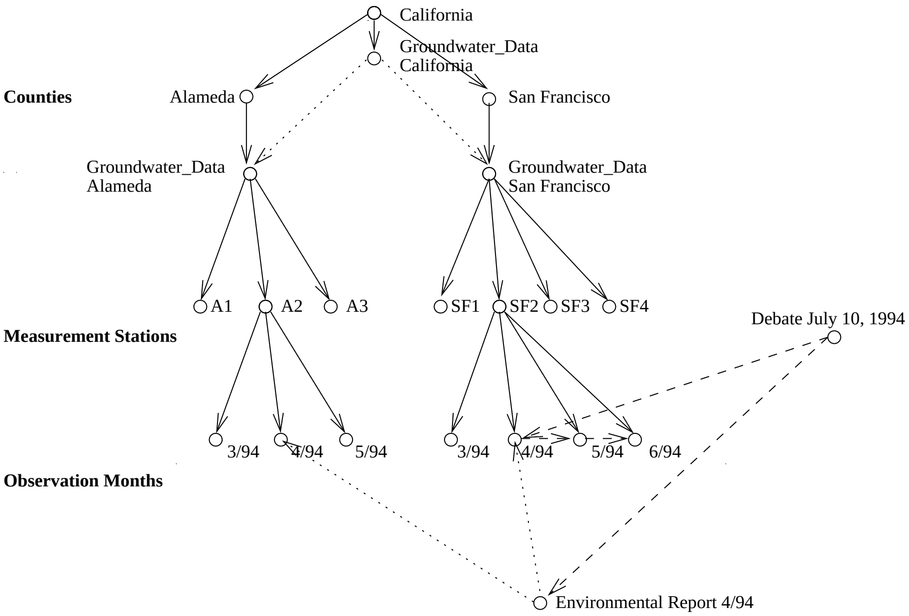


Figure 4 A selection of UDK objects and associations to represent a groundwater database

In summary, it is important to note that the links connecting UDK objects may have a great variety of semantics. These different types of links need to be made explicit in the UDK by a labeling scheme. Users should have the option to choose the types of links they want to see at a given time. This would allow them to see a UDK object in a variety of contexts and to switch back and forth between those different representations. On the screen this could be supported, for example, by different colors and drawing modes for different types of links (Fig. 4).

5.4 The Future of the UDK

The UDK is a meta information system and navigation tool that documents collections of environmental data from the government and other sources. Given the extreme success of the World Wide Web (WWW), we expect a significant amount of this kind of data to be available via WWW in the very near future. At this point there is no question that the Web is the most promising option to follow the spirit of the EU guideline and to make environmental information *really available* to anybody who is interested. The UDK could play a major role in helping users to navigate in this overwhelming information pool, to identify which data is relevant for a given query, and to retrieve it fast and in a user-friendly manner.

Austria and several German states have recently released WWW implementations of the UDK (see `http://www.wiwi.hu-berlin.de/~guenther/udk.html` for URLs). Access is mainly keyword-based. The result of a search is a list of relevant UDK objects. More details on a particular object are available by checking it and sending the marked-up form back to the server. A CGI script then retrieves the corresponding additional attributes. Partly due to backlogs in data entry, however, most UDK objects are much less elaborate than the detailed example given in Fig. 3.

HTML links between UDK objects or to environmental data objects are rarely used in those implementations. Instead one can request the ancestor and descendants of a given UDK object in the primary catalogue. This is done by means of another form-based mark-up mechanism, similar to the one described above. Further details of the implementation and related issues have been described by Kramer et al. [20, 21].

A somewhat different approach for a WWW implementation of the UDK was suggested in [14]. Here, each UDK object corresponds to exactly one Web page. HTML links are used to implement primary and secondary catalogues and to establish connections to environmental data objects. In our view this architecture leads to a much more flexible and user-friendly implementation. A corresponding realization is currently under consideration for the German *Federal* version of the UDK.

6 CONCLUSIONS

The goal of this article was to show how metadata is becoming increasingly popular in geographic and environmental information systems. It can improve both the availability and the quality of the information delivered. The growing popularity of Internet-based data servers has accelerated this trend even further. After a general discussion of the term metadata and of the question how to integrate metadata into traditional information system architectures, we have discussed several case studies in detail. Particular emphasis has been put on the U.S. efforts to build a National Spatial Data Infrastructure, and on several European projects to integrate environmental information processing at the national and international levels.

Despite the remaining heterogeneities and inefficiencies, the outlook seems positive. The ubiquitous trend towards open systems as well as the rise of the World Wide Web are two recent developments that will greatly improve the way we manage geographic and environmental information. Users will have faster and more comfortable access to ever greater amounts of information, and metadata will be an essential component of the underlying software architectures.

Finally, we envision an increasing number of applications where metadata is used to administer not only simple data sets but also complex software tools, such as domain-specific aggregation methods or environmental simulation models. In those applications, the metadata will be used for two purposes: (i) to find the appropriate software tool for a given problem, and (ii) to to apply the tool to a given data set over the Internet without having to port the software to a local machine. Our own MMM project [13] is one example of a software architecture that supports this paradigm.

Acknowledgements

Thanks to Pamela Drew, Ralf Kramer, Ralf Nikolai, Helmut Lessing, Jan Röttgers, Thomas Schütz, Amit Sheth, Walter Swoboda, and Jerry Ying for helpful comments. Thanks also to Jean-Marc Saglio, who hosted the first author's stay at the École Nationale Supérieure des Télécommunications in Paris, where parts of this work have been completed.

REFERENCES

[1] C. Bardinet, J. E. Dubois, J. P. Caliste, J. J. Royer, and J. C. Oppeneau. Data processing for the environment analysis: a multiscale approach. In *Space and Time in Environmental Information Systems*, Marburg, 1995. Metropolis-Verlag.

[2] W. J. Campbell and R. F. Cromp. Evolution of an intelligent information fusion system. *Photogrammetric Engineering and Remote Sensing*, 56(6):867–870, 1990.

[3] P. Drew and J. Ying. Metadata management for geographic information discovery and exchange. In W. Klas and A. Sheth, editors, *Managing Multimedia Data: Using Metadata to Integrate and Apply Digital Data*, 1997. McGraw Hill.

[4] Council of the European Communities. Council Directive (90/313/EEC) of 7 June 1990 on the freedom of access to information on the environment. *Official Journal of the European Communities*, L158:56–58, 1990.

[5] ESRI, editor. *Proc. 11th ESRI User Conference.* ESRI Inc., Redlands, CA, 1991.

[6] ESRI, editor. *Proc. 15th ESRI User Conference.* ESRI Inc., Redlands, CA, 1995.

[7] ESRI. Metadata management in GIS. Technical report, ESRI Inc., Redlands, CA, 1995. URL `http://www.esri.com/resources/papers/papers.html`.

[8] ESRI. SDTS - Supporting the Spatial Data Transfer Standard in ARC/INFO. Technical report, ESRI Inc., Redlands, CA, 1995. URL `http://www.esri.com/resources/papers/papers.html`.

[9] European Environmental Agency. Catalogue of datasources for the environment. Kopenhagen, August 1993. Version 930831.

[10] F. G. Fegeas, J. L. Cascio, and R. A. Lazar. An overview of FIPS 173, the Spatial Data Transfer Standard. *Cartography and Geographic Information Systems*, 19(5), 1992.

[11] FGDC. *Content Standards for Digital Geospatial Metadata.* U.S. Government, Federal Geographic Data Committee, Washington, DC, 1994. URL `ftp://fgdc.er.usgs.gov`.

[12] Geographic Designs Inc. Online documentation of Geolineus 3.0. Santa Barbara, CA, 1995. URL `http://www.geodesigns.com/gl_desc.html`.

[13] O. Günther, R. Koerstein, R. Müller, and P. Schmidt. The MMM project: Access to algorithms via WWW. In *Proc. Third International World Wide Web Conference*, 1995. URL `http://www.igd.fhg.de/www95.html`.

[14] O. Günther, H. Lessing, and W. Swoboda. UDK: A European environmental data catalogue. In *Proc. Third International Conf. on Integrating GIS and Environmental Modeling*, 1996. URL `http://www.ncgia.ucsb.edu/conf/santa_fe.html`.

[15] O. Günther and H.-J. Schek, editors. *Advances in Spatial Databases.* Number 525 in LNCS. Springer–Verlag, Berlin/Heidelberg/New York, 1991.

[16] G. Hallmark. The Oracle warehouse. In *Proc. 21st Int. Conf. on Very Large Data Bases*, 1995.

[17] D. Johnson, P. Shelley, M. Taylor, and S. Callahan. The FINDAR directory system: a meta-model for metadata. In D. Medykyj-Scott, I. Newman, C. Ruggles, and D. Walker, editors, *Metadata in the geosciences*, pages 123–137, Loughborough, Great Britain, 1991.

[18] V. Kashyap, K. Shat, and A. Sheth. Metadata for building the MultiMedia Patch Quilt. In W. Klas and A. Sheth, editors, *Managing Multimedia Data: Using Metadata to Integrate and Apply Digital Data*, 1997. McGraw Hill.

[19] H. Keune, A. B. Murray, and H. Benking. Harmonization of environmental measurement. *GeoJournal*, 23(3):249–255, 1991.

[20] A. Koschel, R. Kramer, R. Nikolai, W. Hagg, and J. Wiesel. A federation architecture for an environmental information system incorporating GIS, the World Wide Web, and CORBA. In *Proc. Third International Conf. on Integrating GIS and Environmental Modeling*, 1996. URL `http://www.ncgia.ucsb.edu/conf/santa_fe.html`.

[21] R. Kramer and T. Quellenberg. Global access to environmental information. In R. Denzer, D. Russel, and G. Schimak, editors, *Proc. 1995 International Symposium on Environmental Software Systems*, International Federation for Information Processing (IFIP), pages 209 – 218, London, 1996. Chapman and Hall.

[22] R. Kuggeleijn. Managing data about data. *GIS Europe*, 4(3):32–33, 1995.

[23] R. Lenz, M. Knorrenschild, C. Herderich, O. Springstobbe, E. Forster, J. Benz, W. Assoff, and W. Windhorst. An information system of ecological models. Technical Report 27/94, GSF, Oberschleissheim, Germany, 1994. URL `http://www.gsf.de/UFIS/ufis`.

[24] H. Lessing. Umweltinformationssysteme - Anforderungen und Möglichkeiten am Beispiel Niedersachsens. In A. Jaeschke, W. Geiger, and B. Page, editors, *Informatik im Umweltschutz*, Berlin/Heidelberg/New York, 1989. Springer–Verlag.

[25] H. Lessing and T. Schütz. Der Umwelt-Datenkatalog als Instrument zur Steuerung von Informationsflüssen. In L. Hilty, A. Jaeschke, B. Page, and A. Schwabl, editors, *Informatik für den Umweltschutz*, Marburg, 1994. Metropolis-Verlag.

[26] J. McCarthy. Metadata management for large statistical databases. In *Proc. 8th Int. Conf. on Very Large Data Bases*, 1982.

[27] D. Medykyj-Scott, I. Newman, C. Ruggles, and D. Walker, editors. *Metadata in the geosciences.* Loughborough, Great Britain, 1991.

[28] Ronald B. Melton, D. M. DeVaney, and J. C. French, editors. *The Role of metadata in managing large environmental science datasets - Proc. SDM-92*, Richland, WA, 1995. Pacific Northwest Laboratory. Technical Report No. PNL-SA-26092.

[29] K. Mikkonen and A. Rainio. Towards a societal GIS in Finland - ArcView application queries data from published geographical databases. In ESRI, editor, *Proc. 15th ESRI User Conference*, Redlands, CA, 1995. ESRI Inc.

[30] D. Miller and K. Bullock. Metadata for land and geographic information - an Australia-wide framework. In *Proc. AURISA'94*, pages 391–398, Sydney, Australia, 1994.

[31] D. Miller and B. Forner. Experience in developing a natural resource data directory for New South Wales. In *Proc. AURISA'94*, pages 391–398, Sydney, Australia, 1994.

[32] F. Ostyn. The EDRA - fueling GIS applications with required geographical information. In ESRI, editor, *Proc. 15th ESRI User Conference*, Redlands, CA, 1995. ESRI Inc.

[33] W. Pillmann and D. J. Kahn. Distributed environmental data compendia. In *Proc. 12th IFIP World Computer Congress.* Elsevier, 1992.

[34] F. J. Radermacher. The importance of metaknowledge for environmental information systems. In O. Günther and H.-J. Schek, editors, *Advances in Spatial Databases*, number 525 in LNCS, pages 35–44, Berlin/Heidelberg/New York, 1991. Springer–Verlag.

[35] D. Seaborn. Database management in GIS: Is your system a poor relation? *GIS Europe*, 4(5):34–38, 1995.

[36] M. Siegel and S. Madnick. A metadata approach to resolving semantic conflicts. In *Proc. 17th Int. Conf. on Very Large Data Bases*, 1991.

[37] U.S. Geological Survey. Spatial Data Transfer Standard (SDTS). Reston, VA, 1992. URL `http://sdts.er.usgs.gov`.

[38] U.S. Government. Coordinating geographic data acquisition and access: The national spatial data infrastructure. Washington, DC, April 1994. U.S. Executive Order 12906.

[39] D. R. F. Walker. Introduction to metadata in the geosciences. In D. Medykyj-Scott, I. Newman, C. Ruggles, and D. Walker, editors, *Metadata in the geosciences*, Loughborough, Great Britain, 1991.