Meta Information Concepts for Environmental Information Systems

Ralf Denzer, Reiner Güttler
Hochschule für Technik und Wirtschaft des Saarlandes
Goebenstr. 40
66117 Saarbrücken
Germany
{denzer,guettler}@htw.uni-sb.de

Presented at the second annual conference of GeoComputation '97 & SIRC '97, University of Otago, New Zealand, 26-29 August 1997

Abstract

The paper gives an overview of a number of aspects of the meta information discussion for Environmental Information Systems (EIS) over the past 7 years. While meta information has mostly been mentioned in the context of Environmental Data Catalogues (EDCs) and/or Catalogues of Environmental Data Sources (CDSs), our group uses meta information for the integration of environmental data into environmental networks. From this viewpoint, we also need EDCs and network navigation components, but our goal was one step further than the above mentioned projects: they usually stop in front of the data source and do not offer integration concepts to connect the data source into a network (Denzer,1995).

In this paper, we will discuss a number of applications of different meta information models, which can be described by a general model to represent meta information. The generic idea of this model has been published(Denzer 1996). The first chapter is a modified extract of this publication in order to make clear the different implementation presented in later chapters.

1. A Generic Meta Information Model

In order to describe general meta data categories, we destinguish between *semantics*, *syntax*, *structure*, *navigation, history and summaries*. We will describe these categories with an example from the bottom up.

1.1 Semantics

By *semantical meta information* we denote additional information (additional to the raw data) which is used to describe the meaning of information. Semantical meta information is therefore the information which is needed to describe a data item such that it is interpretable by a user (from the same application field) who has not sampled the data himself.

As a less abstract term we can also use the term *data description* as a synonym for *semantical meta information.*

As you can see in Fig. 1, we append a set of meta information items to the raw data. The meaning of the meta information items can be general knowledge (like *address of data provider*) and therefore be understood by the general public, or it can have domain specific meaning which is only understood by an expert in the specific application area (like *field method*). This means that the set of meta information items may be different for different user groups.

1.2 Syntax

By *syntactical meta information*, we denote information which is used to describe the way the raw data is stored and/or can be accessed. Syntactical meta information is unimportant for end users and is only used by software systems to access the data. Syntactical meta information usually consists of information about the data type of the raw data and an access method.

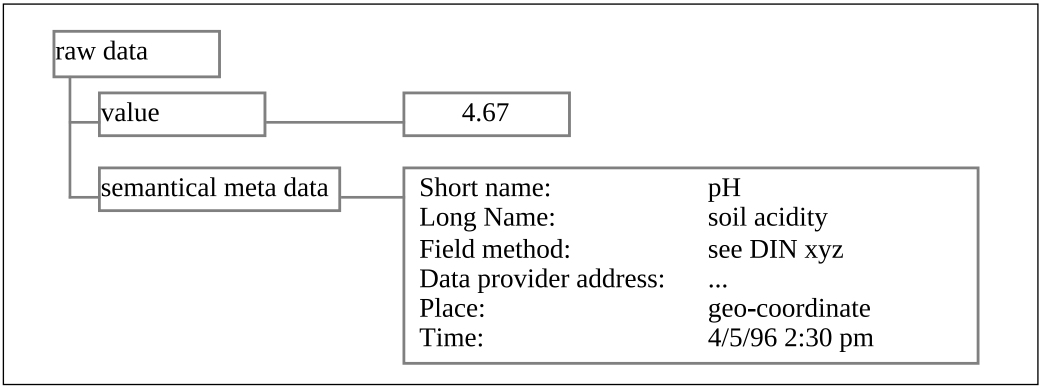

Fig 1. Semantical meta information

Fig. 2 shows how syntax information is appended to the existing set of information describing the raw data. A minimum of information regarding the data type and access methods may be given depending on the way the data is stored (in this example a relational database).

1.3 Structure

Up to this point, we have shown single data items and how their semantics and syntax can be described. In reality, data objects can_nott be described as single items very often. Commonly, aggregates of data items form an *environmental object*, and there is meta data which applies to the whole object as well as to the single raw data item. Therefore it is necessary to describe the *structure* of data objects as well, we denote this description as *structural meta information*.

In fig. 3, several data items with their meta data are composed to an object. Additionally, a semantical description of the overall object is given, which consists of the meta information applying to each of the items (e.g. *data provider address* would no longer be meta information of *pH*, it would be part of the semantical description of the object).

The semantical description of the overall object is again a list of meta data items, according to the description of a single data item. In this case, the description of structure is such that an *object* consists of a list of attributes. It is important to notice that an attribute can be of type *datatype* (single datatype, vector, time value, ...) or of type *object* itself. This also applies to each of the metadata items (an object of class *field method* is meta data item for an at-

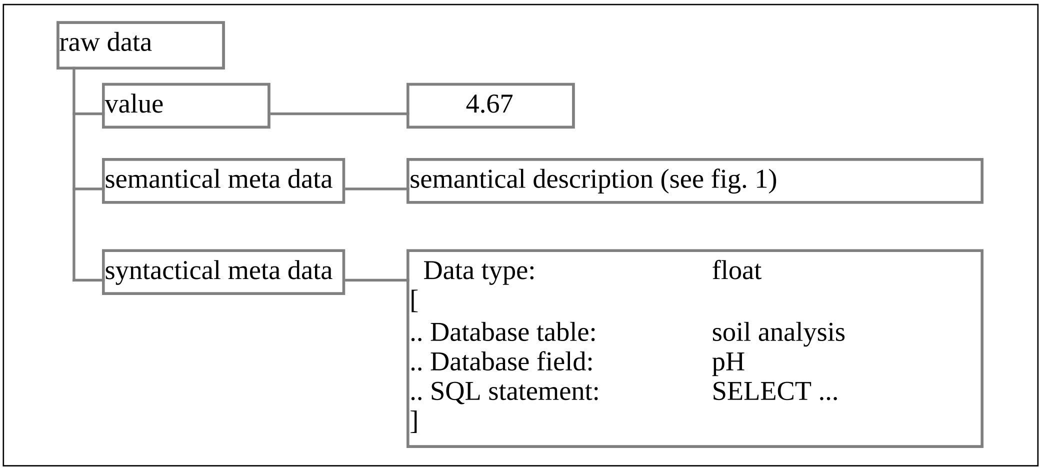

Fig. 2. Syntactical meta_information

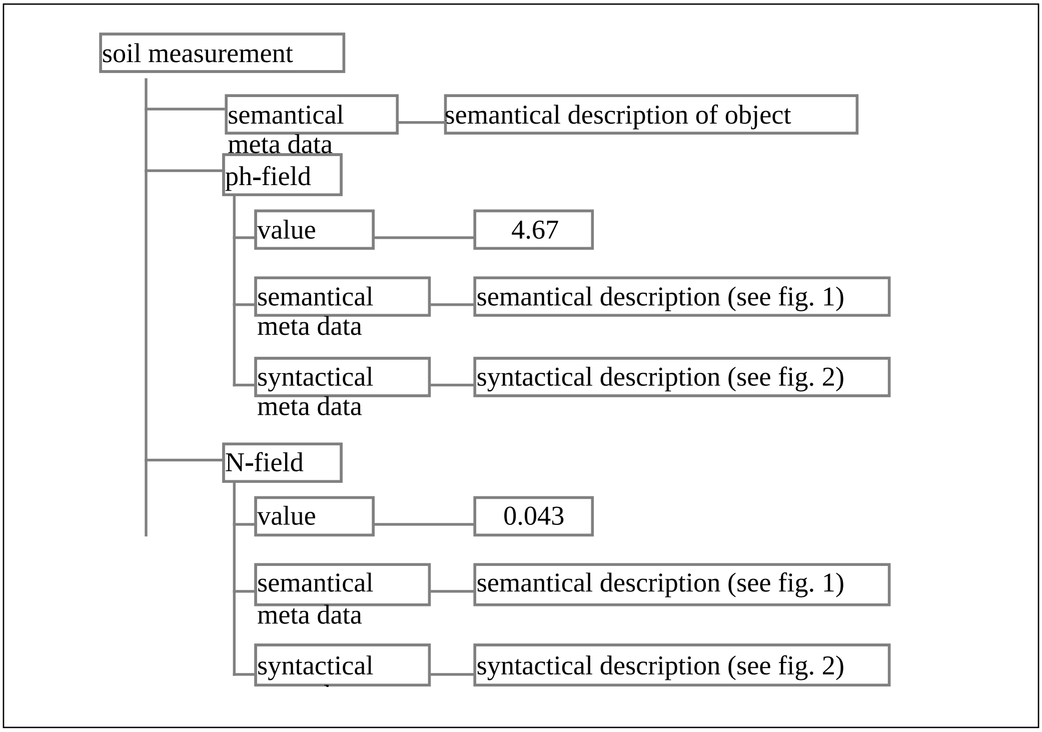

Fig. 3. Structural meta information

tribute *pH* of an object from class *soil measurement*). This meta data model is therefore inherently object oriented, but it would also be possible to describe the structure of the whole with other methods.

1.4 Navigation

Semantics, syntax and structure are entities used to *describe* environmental objects. Another important issue is to *locate* environmental objects. By *navigational metainformation* we denote such information which is used to locate objects and data sources of interest. Navigation occurs within systems (search masks, keyword lists, inventories, etc.), or among systems or even whole networks. Environmental *data catalogues* are one of the means to locate objects.

Fig. 4 gives an example of a data catalogue for one information system. The catalogue combines a list of object classes, a hierarchical tree (table of contents) and links from chapters to class descriptions. Such a catalogue may also include a list of keywords which can be inspected to find information of interest. A data catalogue on the level of an organization or of a whole network would look differently (see Fig. 5). The entries in the table of contents or keyword lists build links to information systems (e.g. data sources). We call such a catalogue *meta catalogue*.

Navigation is much more than that. It includes issues of search engines, statistical information and it raises issues about how to organize information sources over a whole network. Bad experiences with information searches on the Web illustrate these problems.

1.5 History

The problem of history of environmental measurements has widely been ignored over the past years. Why is this the case? First, history means that samples may be produced *by different measurement technology* over the course of time. This increases the design and maintenance efforts for an information system significantly . Second, history also means that the *data structures* change over time. This is even worse for an information system design. Third, history, or changing methods, produces problems in the comparability of data, which is a problem for the scientists, a problem they produced themselves by changing the method.

History in terms of meta data means that historical records about how data has been sampled are important for the description of the data. In practice this means that each of the meta data objects in Figures 1 to 3 must be recorded historically and may even change their structure.

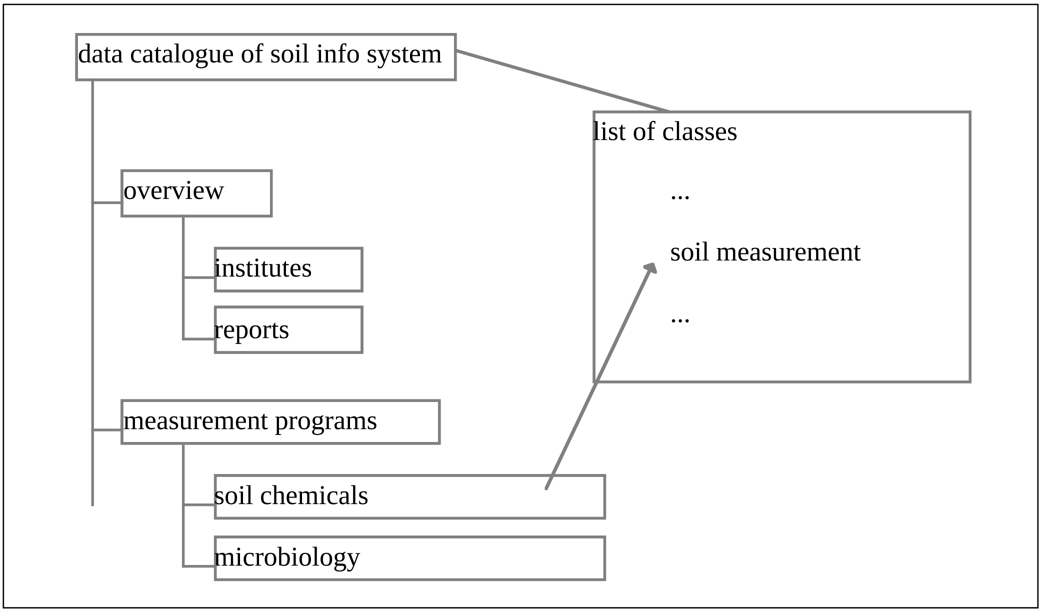

Fig. 4. Environmental data catalogue for an information system

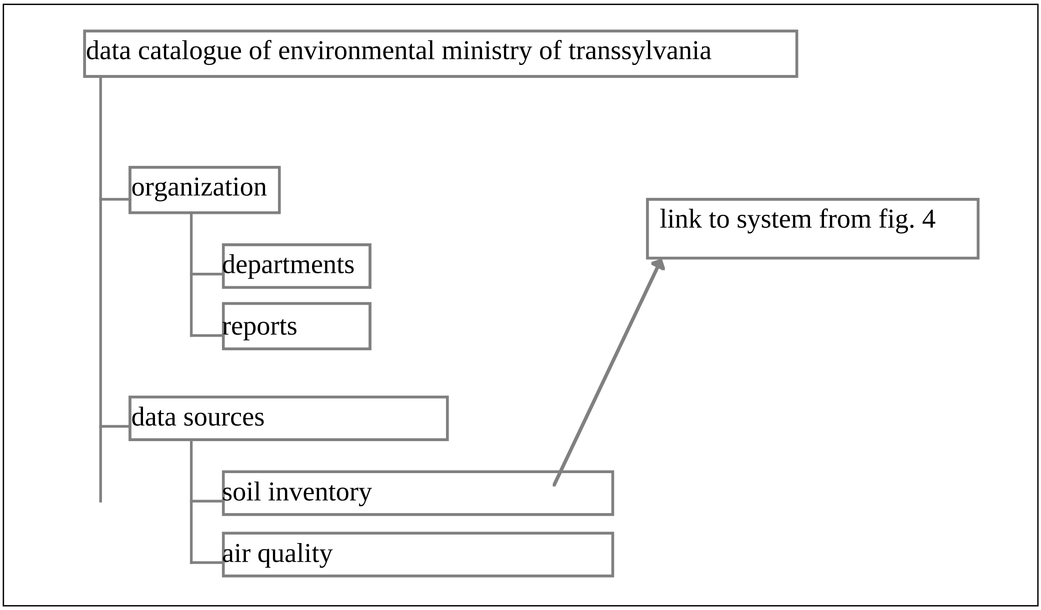

Fig. 5. Environmental data catalogue (meta catalogue) of an organization or network

1.6 Summaries

Summaries are used to give an overview at each level of meta information (e.g. number of classes in a catalogue, number of objects in a class, percentage of attributes used, overall summary of time or geographical scale, etc). Summaries are implemented to help users during navigation and although these are very simple mechanisms, they are not used frequently.

2. Application Examples

In this chapter, we give three examples of the implementation of the concepts mentioned above. The examples are very different in nature.

2.1 SIRIUS Meta Information Model

The SIRIUS (Saarbrücken Information Retrieval and Interchange Utility Set) system was our first implementation of a meta information concept. The goal of SIRIUS is to provide an integration architecture for open EIS. This architecture has been documented in various publications (Denzer, 1995). Meta information in the context of SIRIUS is mainly used for the following purposes:

- to provide a data catalogue for an existing information system,
- to document the information classes of this information system in terms of class syntax, structure and semantics,
- to use the class documentation for the access of the information system and, to provide networked catalogues (meta catalogues) for the organization of a SIRIUS network.

The meta information used in SIRIUS is a very simple model, where

- catalogues are hierarchical trees,
- classes are described by a set of attributes,
- classes are linked into nodes of the catalogue,
- the class structure is given by a list of primitive attributes,
- the attribute syntax is given by its data type (plus an optional list attribute),
- each class can have a different description, and
- the semantical meta information for each class and each attribute is just a free text.

In terms of distribution of data, SIRIUS is completely network transparent.

2.2 FAM Meta Information Model

FAM (Forschungsverbund Agrarökosysteme München) is a big agricultural research project funded by the German government. In this project the operation of a farm is monitored on a long term time scale. A large number of institutes (at the time of our involvement in the project around 60) collect all possible data associated with the operation of this farm and use this information for ecological assessments.

In 1994 and 1995, our group developed a meta information model for the database of the FAM project. This model is, to the best of our knowledge, the most detailed and flexible meta information model implemented at this time. The differences between the FAM model and SIRIUS are twofold:

- no network component (which was not needed), and
- the description of classes is much more detailed.

Compared to SIRIUS, the FAM model describes classes as follows:

- a class has again a list of attributes, but these attributes can be of any type, including other objects, therefore the data model is recursive
- each class and each attribute has a set of meta information attached to it, and this set is not only a free text but a list of meta information attributes which can be of any type (primitive types and objects); as objects can contain objects, also the meta information model attached to every class and/or attribute can be recursive
- the meta information contents of any class or attribute can change over time, reflecting change in the reality (i.e. the model can store a history of e.g. measurement instruments)
- the class structure and the meta information struc-

ture (of a class or any attribute) can change over time (i.e. even if the reality changes its structure, the information model can reflect this and can even remember the structure at a certain date in the past).

It is easy to imagine, that this meta information model is neither trivial to understand nor trivial to implement. But our investigations showed clearly, that this is the set of information needed to document a long scale research program such that the information can still be used after a longer time period.

2.3 TEMSIS Meta Information Model

TEMSIS (Transnational Environmental Management Support Information System) is a project funded by the EU under the Environmental Telematics program (Schimak 1996). The goal of the system is the support of environmental information and planning in the area around the French-German border near Saarbrücken and Saargemuines. We are part of a consortium of 8 partners developing this system.

Our colleagues at the Austrian Research Center Seibersdorf are developing the meta information server. Our group is responsible for the information services between the server and data sources as well as between server and client applications on both sides of the border (for this purpose, a port of SIRIUS is used). As our tasks in the overall project were related, we have worked closely together in the modeling of the meta information. The TEMSIS meta information model is located between the two models mentioned above in the following areas:

- meta information is a list of primitive data type objects, not a text as in SIRIUS, but not recursive as in FAM
- the TEMSIS model does not distinguish between objects, attributes and classes. What this really means compared to SIRIUS or FAM will come out in the future. It seems depending on the way the meta information is organized in the catalogue, this can be completely irrelevant to the end user and will only be noticed by the system designer.
- the TEMSIS model does not store any history, nor does it directly reflect the distributed nature of EIS (TEMSIS uses one centralized server)
- the TEMSIS model introduces a new very powerful idea from our friends in Seibersdorf. Links between information objects, which are used to model relationships between objects and can be used extensively for navigation.

3. Discussion

The three models are very different in nature and purpose. They also reflect the different reasons of how and why to use meta information in an EIS. SIRIUS uses a very simple model for the interconnection of EIS and therefore describes objects only on a very abstract level. FAM, in comparison, is an extremely detailed and sophisticated model, which is able to model anything, but it is not easy to use. We have not been able, due to limited funds in this particular project, to implement the user interface components which handle the complexity of the model, especially for the persons who have to maintain the meta information system. The TEMSIS model appears to be a good compromise for a public information system, which does not have the same detailed need for documentation as is found in a research program. However, we do not have any experiences yet with the model, as the demonstrator system will be installed this summer. Also, it is limited to one central server, although the information services are capable for link up to a network.

4. Conclusion

The comparison of the three projects shows that there is not THE meta information model for the world or for EIS. In every case and under different circumstances, a different way to use meta information will be useful. But we strongly believe that there is a generic way of thinking about meta information, which may be reflected by the first chapter of this article and which may have been implemented in a most generic way in the FAM meta information model.

If we look back to the past 7 years, since the strange word "meta information" became common (and not many people know what it is), we can also see that a convergence towards usable approaches in EIS took place.

References

Denzer (1995) R. Denzer, R. Güttler, SIRIUS - Saarbrücken Information Retrieval and Interchange Utility Set, in: R. Denzer, D. Russell, G. Schimak, International Symposium on Environmental Software Systems 1995 (ISESS 1995), Malvern, PA, USA, June 1995, Chapman & Hall, 1995

Denzer (1996) R. Denzer, R. Güttler , Requirements of Meta Information Models for the Environmental Domain, Journal of Computing and Information, Vol. 2, No. 1, pp. 1112-1120, 1996

Schimak (1996) G. Schimak et. al. (TEMSIS consortium), TEMSIS - a first step towards transnational environmental information systems, ECO-Informa 96, Global Networks for Environmental Information, November 1996, Epcot Science and Technology, Lake Buena Vista, Florida, USA