The Disappearance of Landscape Objects in Databases:
An Ontology Driven Approach to Object Representation.

B. Waldvogel

Swiss Federal Institute for Forest, Snow and Landscape Research (WSL)
Zürcherstrasse 111, CH-8903 Birmensdorf
Email: bettina.waldvogel@wsl.ch

1. Introduction

Data about landscape objects is typically stored in heterogeneous database systems. As a consequence it is often not trivial to keep track of all information and data items that refer to one particular landscape object. Landscape objects are spatial objects that share common characteristics. In this paper we focus on a special type of landscape objects: protected biotopes in the data centre nature and landscape (DNL), which are well defined as survey polygons (Bauer-Messmer 2009).
Objects are supposed to have an identity although they may undergo different kinds of changes (Hornsby and Egenhofer 2000). In our case the identity of a landscape object is given by an official number. Nevertheless there are many obstacles: data collections which refer only to a part of an object or to more than one object, objects which overlap other partitions such as community or cantonal borders, object mergers and splits.
The present work summarizes some general ideas on how to further develop the semantic enrichment of our data centre for nature and landscape (DNL) towards a more technical level, so that the connection between the landscape objects and the data objects will not get lost in the database.

2. The path of a landscape object into the database

Different data collections are being gathered about biotope objects: survey polygons, textual descriptions, tabular data about vegetation, geomorphology, and many more. All these data objects refer to landscape objects. In the database different data collections are stored according to their technical characteristics, e.g. survey polygons in a geographic information system, tabular data in a relational database. According to good database management practice all data objects carry a key that refers to the corresponding landscape object. Nevertheless an important piece of information along the path from landscape to database is still missing: we call it the technical semantics.

3. Semantics of an object's way from landscape to database

Ontologies are shared conceptualizations (Gruber 1983), which can be used either for semantically enhanced user-interfaces or for technical interoperability between computers.
Application ontologies describe concepts depending both on a particular domain and task (Guarino 1998). A domain ontology reflects the conceptualizations of a specific domain (or the application relevant aspects thereof). The task ontology holds information about the data representation on the logical level as well as mereo-

topological information. We suggest to design an application ontology with a further component: a technical ontology, which reflects the inherent semantics of the underlying storage architecture and database models (see Figure 1).

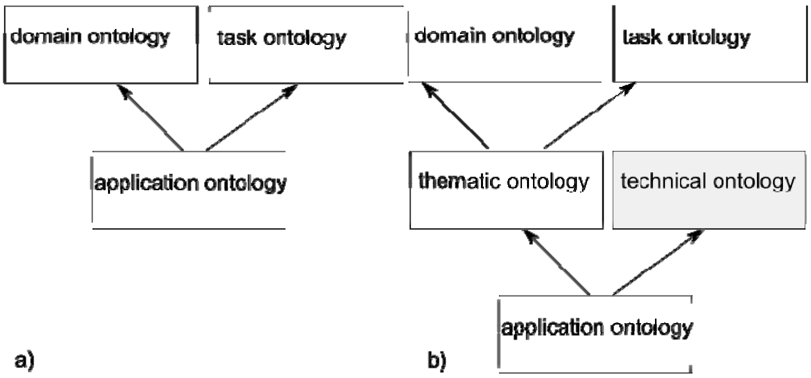

Figure 1. Application ontology as defined by Guarino (1998) a) and our approach: an application ontology enriched with a technical ontology (grey shaded) b).

The gap between human conceptualizations and the data conceptualizations is bridged by a thematic ontology (i.e. a task ontology and domain ontology). The gap between the data conceptualization and the database conceptualization is bridged by a technical ontology (see Figure 2).

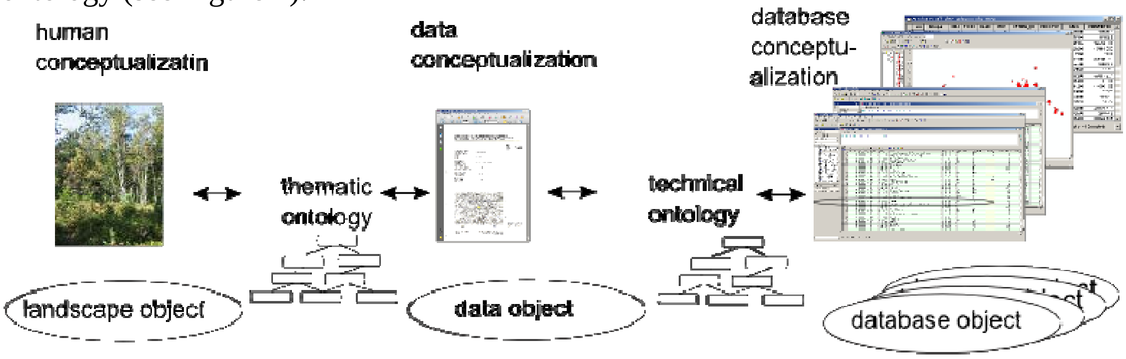

Figure 2. Semantics along the path from landscape to database: Data objects that refer to one landscape objects are split into many database objects and stored in a heterogeneous storage architecture.

4. Requirements for the technical ontology

The following requirements are identified:

- For each database object a reference must be given to the corresponding data object and landscape object.
- The semantics of all database tables and their attributes must be reflected in the technical ontology.
- The semantics of all cross references such as foreign-key relationships must be reflected in the technical ontology.
- The semantics of the storage architecture of the temporal model (e.g. snapshots, object lifecycles, processing chains) and spatial model must be included.

Both the integration of ontologies in relational database systems (Auer and Ives 2007) and the automatic generation of ontologies from database schemas (Hert 2009) are currently being investigated. Our approach suggests to include additional semantics in the technical ontology so that landscape objects will not get lost in the database.

5. Examples

5.1 Broken Lifecycle – Lost Identity

Before 2002 there were two separate objects in the bog inventory: object 494 (Tanzplatz) and object 252 (Madils). After 2002 there is only one bog object left, object 252. Object 494 does no longer exist as an individual object. It has become a part of object number 252.

In a traditional database approach the objects are stored in one table and the partial objects in another table. There are only mereological references to the objects to which the partial object belong. However there is no data structure to keep track of the history, if objects become partial objects and vice versa. The same problem occurs, when objects from one inventory are moved to another inventory, e.g. a bog deteriorates and becomes a fen. Database systems typically do not support "multiple identities". A technical ontology can help to establish lifecycle-relationships across inventories and partition hierarchies.

5.2 Individual Data – Lost Data Items

For different object types different data is available. For instance for hunting ban areas there are lists of protected animals. For bogs and fens there are vegetation charts. And for some objects there are several levels of hierarchically arranged partitions of geometries. A technical ontology reflects the relationship between the object types and individual objects and their storage locations (e.g. database, tables, columns).

5.3 Processing Chains – Lost History

There are long processing chains from collecting raw data to the final data products. These processing steps are documented in the database. Their semantics, however, needs further documentation. Keeping this information in a technical ontology allows seamless reasoning over all technical aspects of the data processing and the storage architecture.

6. Discussion

Landscape objects disappear in the database, because they are not stored as entities but in a heterogeneous storage architecture. Depending on the type of landscape objects different data is available. In order to bring together the pieces (i.e. database objects) a technical ontology is suggested. The detailed structure of this technical ontology is subject to further research, however a list of basic requirements for the technical ontology is presented.

There are parallels of our approach to the five-universe-paradigm (see Table 1) introduced by Fonseca et al. (2002). However we put more emphasis on the semantic enrichment of the technical levels.

Table 1. Comparison of the five-universe-paradigm and our approach.

Five-universe-paradigm	Ontology driven object representation
Physical universe	Landscape objects and human conceptualization
Cognitive universe	Thematic ontology reflecting the human conceptualization
Logical universe	Thematic ontology reflecting the data conceptualization
Representation universe	Data objects
Implementation universe	Technical ontology reflecting the database conceptualization and database objects

Standardized metadata can also be used to bridge the gap between the data objects and the database objects. Our choice fell on ontologies because they use well defined formats (e.g. RDF or OWL) and many tools are available to automatically process the information.

Criteria to analyze object representations are proposed by Marr (1982): (1) accessibility of the information (2) scope of a wide range of objects (3) unique representation of objects (4) stability in the model and (5) sensitivity, i.e. availability of detailed features. Our ontology driven approach to object representation is particularly strong in point 2, 4 and 5. It is open to all kinds of landscape objects carrying a wide variety of additional data and it allows to seamlessly represent the semantics from landscape to database object.

Acknowledgement

Many thanks to my colleagues at WSL for the fruitful discussions. I would also like to thank the anonymous reviewers for their helpful comments. This work has been funded by the Swiss Federal Office for the Environment (FOEN).

References

Auer S, Ives Z: Integrating Ontologies and Relational Data. Technical Report MS-CIS-07-24, Department of Computer and Information Science, University of Pennsylvania (2007)

Bauer-Messmer B, Wotruba L, Müller K, Bischof S, Grütter R, Scharrenbach T, Meile R, Hägeli M, and Schenker J, 2009, The Data Centre Nature and Landscape (DNL): Service Oriented Architecture, Metadata Standards and Semantic Technologies in an Environmental Information System. *Environmental Informatics and Industrial Environmental Protection: Concepts, Methods and Tools. Proceedings of the 23rd Int. Conf. EnviroInfo*, Berlin, Germany, 101–112.

Fonseca F, Egenhofer M, Davis C and Câmara G, 2002, Semantic Granularity in Ontology-Driven Geographic Information Systems. *Annals of Mathematics and Artificial Intelligence*, 36(1-2):121-151.

Gruber T R, 1993, A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199-220.

Guarino N, 1998, Formal Ontology and Information Systems. *Frontiers in Artificial Intelligence and Applications*. 3-18.

Hert M, 2009, Relational Databases as Semantic Web Endpoints. Proceedings *of the 6th European Semantic Web Conference (ESWC)*, Springer.

Hornsby K and Egenhofer M, 2000, Identity-Based change: a foundation for spatio-temporal knowledge representation. *Int. J. Geographical Information Science*, 14(3), 207-224.

Marr, D., 1982, *Vision*, publisher W.H. Freeman and Co., San Francisco