Year: 2009

The Data Centre Nature and Landscape (DNL): Service Oriented Architecture, Metadata Standards and Semantic Technologies in an Environmental Information System

Bauer-Messmer, Bettina ; Wotruba, Lukas ; Müller, Kalin ; Bischof, Sandro ; Grütter, Rolf ; Scharrenbach, Thomas ; Meile, Rolf ; Hägeli, Martin ; Schenker, Jürg

The Data Centre Nature and Landscape (DNL): Metadata Standards and Semantic Technologies in an Modular Environmental Information System

Bettina Bauer-Messmer, Kalin Müller*, Lukas Wotruba*, Sandro Bischof*, Rolf Grütter*, Thomas Scharrenbach*, Rolf Meile*, Martin Hägeli*, Jürg Schenker[ǂ]*

[*] *Swiss Federal Institute for Forest, Snow and Landscape Research WSL*
Zürcherstrasse 111, CH-8903 Birmensdorf, Switzerland
[ǂ] *Federal Office for the Environment FOEN*
CH-3003 Berne. Switzerland

Abstract

The Data Centre Nature and Landscape, intended as a database for biotope information, constantly evolved over the last 11 years to become a virtual data centre connecting several databases located all over Switzerland. It was enriched with a user-friendly Web-interface, partially modularized into a Service Oriented Architecture and finally offers open and intuitive search facilities based on semantic technologies.
This paper shall give an overview of the hurdles that had to be taken and the many lessons to be learned during the last decade sharing our experience developing and maintaining environmental information systems.

1. Introduction

Data of protected biotopes of national importance in Switzerland is collected in thematic inventories being part of a national database, as required by the Swiss nature and cultural heritage protection act (Natur- und Heimatschutzgesetz) [Schenker 2008]. For most species in Switzerland there exist central databases, whereas faunistic and floristic data was and still is collected at the Centre Suisse de Cartographie de la Faune CSCF [CSCF] and the Centre du Réseau Suisse de Floristique CRSF [CRSF]. For protected biotopes, however, no central database existed, but data was collected decentralized in different locations in heterogeneous formats. In 1997, the Federal Office for the Environment [FOEN] decided to realize a centralized database for protected biotopes in cooperation with the Swiss Federal Institute for Forest, Snow and Landscape Research WSL [WSL], being the offspring for the project data centre nature and landscape (DNL).

The DNL manages all data about protected biotopes of national importance and further scientific inventories. In the first phase (1998-2001), a process oriented database was designed and implemented. Soon, the benefits of combining DNL data with other data from national databases were realized. In the second phase (2002-2009), the DNL database was extended to a Virtual Data Centre (DNL/VDC) integrating data from external data sources. Currently, a web interface provides easy search in the data centre as well as GIS functionality, The use of a semantic layer enables search even for untrained users. For the third stage (2010-2013) a service-oriented architecture is planned, and first steps in that direction have already been taken. In addition, the search is intended to adapt to semantic information from users [Bauer-Messmer et al. 2008][Scharrenbach 2008] .

This work is structured as follows: In the second chapter an overview over the data sources is given. Chapter 3 demonstrates concepts for standardized and non-standardized metadatalike lineage information. Chapter 4 addresses the connection of heterogeneous data sources in the Virtual Data Centre (VDC), whereas Chapter 5 shows how a service oriented architecture (SOA) ideally fulfils the requirements of the VDC. In Chapter 6 a retroactively added semantic search facility is sketched. Chapter 7 concludes this paper with an overview of the lessons learned in the past 11 years of development and gives an outlook on coming tasks.

2. Data and inventories

In the DNL, there exist three types of data providers: DNL database itself, other research units at WSL, and independent data providers all over Switzerland.

Data in the DNL database

The major inventories comprise the following biotope types: bogs, fens and amphibian spawning grounds, alluvial zones. Inventories comprise a few hundred (e.g. bog inventory) objects up to a few thousand objects (e.g. fen inventory). The DNL data centre comprises roughly 300’000 data items.

The DNL database not only holds thematic data in a relational database, but it is also connected to a GIS system holding spatial data, perimeters and point data of the protected objects.

Temporal information in the DNL is given either explicitly as dates for certain events or implicitly as relative time within processing chains of the lineage (see chapter 3.X or [Bauer-Messer & Grütter 2009]).

Inhouse databases at WSL

In addition to the data directly hosted in the DNL database, two national inventories are located within the WSL. For these databases, the lichen database and the fungi database the data can directly be accessed by a direct database connection.

Other datasources

The flora data of CRSF, fauna data of CSCF and bryophyte data is hosted in external institutions: the City of Geneva, the University of Neuchatel and the University of Zurich respectively.

External data sources are located in corporate networks applying heterogeneous security settings. Several restrictions apply in transferring the designated information to the central repository. There are three different security setups of the external data providers that need to be considered. Restrictions and access respectively can be characterized as follows:

1. No external access from outside to the corporate network
2. Restricted access with proprietary VPN-client
3. SSH-Access including a tunnel for the database network traffic

In the first case the external data provider delivers the data itself to a database table proxied from the internal WSL network. For the second security setup a manual procedure had to be chosen: an administrator connects to the corporate network using the provided VPN-client. A SQL-client software finally gathers the data and stores it in the central repository database. The method used for the third setup consists of an

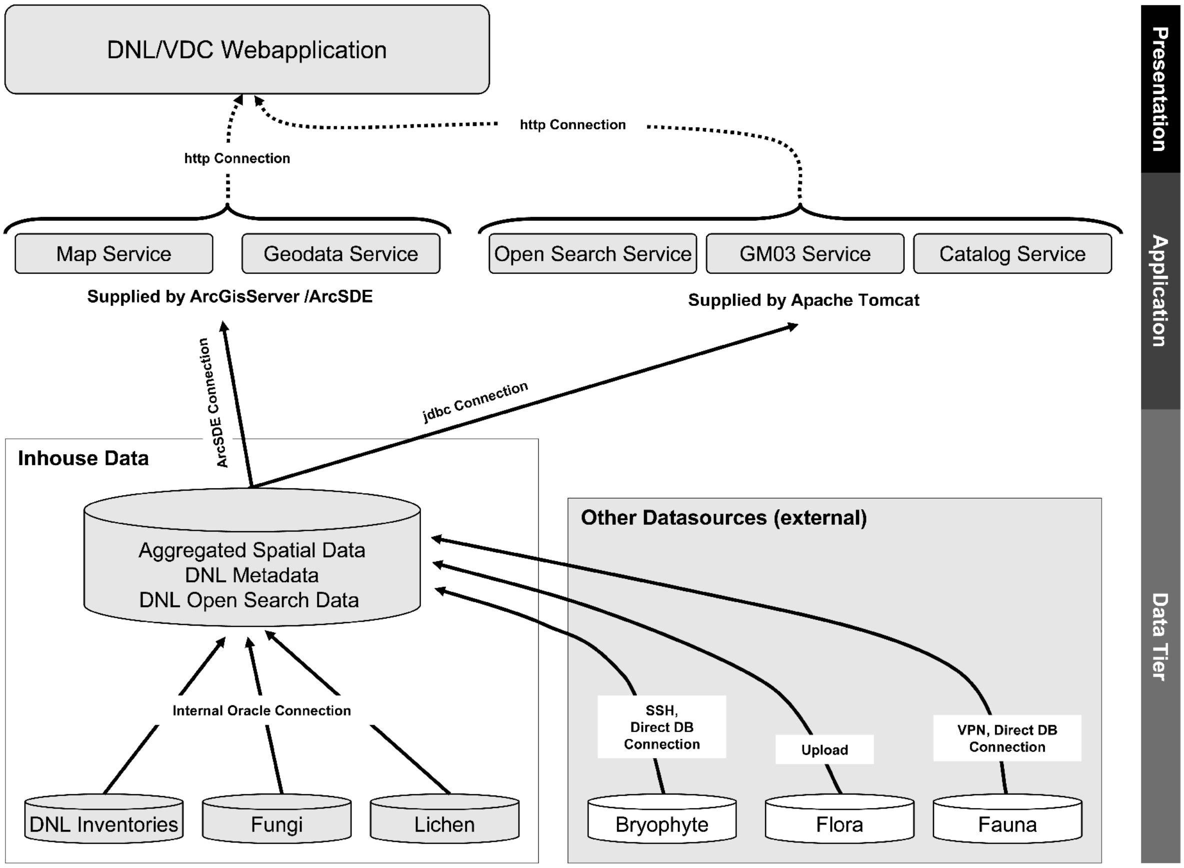

Figure 1: Overview of the main components of the Data Centre Nature and Landscape (DNL) and the virtual data centre (VDC)

automatically opened SSH connection with a tunnel to the data providers' database. The central repository database then collects the necessary datasets from the remote server over a database link.
Considering the time and effort needed to adjust the processes to any changes in security settings a more service oriented approach should be selected (see chapter 5).

Data repository and institution	Characteristics of data and storage system
DNL, Swiss Federal Institute WSL, Birmensdorf	Vector data representing boundaries of endangered biotopes, thematic data, process-oriented lineage, GM03 metadata. Oracle database management system, spatially enabled by Spatial Database Engine (SDE) of ESRI.
Fauna database, CSCF, University of Neuchâtel	X/Y-coordinates representing observations of endangered animals. Oracle database management system, coordinates are stored

	in regular columns (not spatially enabled)
Flora database, CRSF, City of Geneva	X/Y-coordinates representing observations of endangered plants. Coordinates are stored in regular columns (not spatially enabled)
Bryophyte database, Institute of Systematic Botany and Botanical Garden, University of Zurich	Vector data representing approximate location (polygon) or exact X/Y-coordinates of bryophyte samples. Geometric data are stored in ESRI-Shapefiles, corresponding attribute data in Oracle database tables.
Fungi database (swissfungi), WSL, Birmensdorf	X/Y coordinates representing fungi occurrences. Oracle database management system, coordinates are stored in regular columns (not spatially enabled). Coordinates are specified with positional accuracy.
Lichen database (SwissLichens), WSL, Birmensdorf	X/Y coordinates representing lichen occurrences. Oracle database management system, coordinates are stored in regular columns (not spatially enabled). Coordinates are specified with positional accuracy.

Table 1: Summary of the data repositories contributing to the DNL/VDC (adapted from [Frehner & Brändli 2006])

3. Metadata in DNL

3.1 The Need for Metadata

Archiving and sharing data are two of the main objectives DNL. "The goal of the database design is to provide a recording and archiving tool for actions and decisions of environmental protections measures. (…) The database should essentially enable to follow the complete genesis of existing data records" (Lanz et al.). As a consequence, data must be accessible over a very long period in time, whose duration is not known in advance. Not only must the traceability of data be assured for coming generations but also its use in different contexts. The data is required to be available for supporting public authorities, scientific purposes as well for public affairs. These two objectives are met using metadata descriptions (FDGC, ISO 19115 or others), guaranteeing the sustainability of data.

3.2 Implementation of GM03

The Swiss metadata standard GM03, fully derived from ISO 19115 , was implemented in the DNL database. In contrast to the ISO standard, GM03 supports multilingual data descriptions for title, abstract or keywords and supports the national geo-data portal (geocat.ch). For reutilization, metadata is demanded to be accessible via web services from different applications e.g. VDC. For each inventory, a basic metadata set is recorded in GM03 XML dialect and stored in a relational database as a CLOB.

3.3 Metadata in Use

Metadata has independency mainly with three other elements of a geo information system. In addition to the metadata management and catalog software, metadata can supply and enrich various applications with additional information or can control application. Furthermore we focus on the three bold lines (see Fig) metadata-applications, metadata-processes and metadata-data.

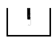

Fig. Interdependencies between elements of an information system.

The history of data as well as its projection into future is described by processes, and processes are described by metadata. A process description is a dynamic view onto data and is thus always a subset of data description. But this part of the metadata standard is more generic as other concrete geo-metadata with precise field description. The ISO standard does not contain a full model to describe process with predecessor and successor.
No possibility to describe metadata as the further purpose of use or the quality of metadata itself.
The metadata are used normally in a catalog today. User is often not willing to collect (recording) additionally to data of intersect. Data collection often cost months or years in money millions of Euro but users are not willing to spend more "ten minutes" to record metadata. The value of metadata must and could be increased.

After the implementation and use of metadata and in reflection with the independency with other elements eight lacks of metadata could be identified:

- *(Integration) Not well integrated:* Metadata are rarely integrated in related applications which deal with geo-data. Otherwise proprietary GIS applications encapsulate metadata in an own way that sharing and access from outside is impossible or can only realized with a huge effort. A lot of structured and semi-structured contents of metadata should appear and reuse in application for user information or decision making.
- *(Catalogs) Manly used for catalogs:* The main application of metadata is doubtless the catalog. More or less a list which can search through and metadata records are displayed and metadata standards are drew mainly for this function to document the existence of a certain data set. It is certainly an important functionality of a metadata application and these catalogs must be managed for persistence geo data sets. But these narrow view avoid (a broader view) new functionality which increase the value of metadata.
- *(Collection) Additional effort to collect metadata:* Mostly the metadata are collected after the data, separated in an own task. Data collection ands processing could last sometimes years and cost millions, but the willingness to describe the data set by metadata is surprisingly small. The process to collect metadata is something apart from the data collecting process. The result is no or a too short description of data sets therefore data collection should support metadata collection; these two process should fit to one another. And the outcome of these two processes should be the result and no more extra effort must be achieved.

- *(Imprecision) Imprecise gathering:* The result of metadata collection is normally a brief description which contains mandatory fields or fields in the core of the metadata standard. Fields in the comprehensive standard are rarely populated. User some times overstrained with the huge richness of those standards. This will result in imprecise collection; misunderstood fields are not filled or wrong filled. Imprecise metadata records results as well from the forgotten maintenance.
- *(Storage) Metadata stored in different and apart systems:* Every application has its own repository; standards are stored in proprietary format without an open defined access. Metadata often stored centralized, these is a direct consequences of the dominant catalog function of metadata and further more data repository founded for a specific purpose handicap again, the integration of metadata in different applications.
- *(Benefit) Users do not see benefit:* The every day use of a user concern data. For the user the additional effort for metadata is not directly evident. Dominant catalog function is under some point of views not evident. Since there is not more functionality than documentation, categorization and metadata do not support every day work, the user can not recognize benefit and more effort must be well legitimate.
- *(Enhancement) No enhancements:* The metadata standards allow enhancement concerning in structure and content. But the standards are not originally drawn to support recent concepts of collaboration or collective intelligence; the spirit of metadata standard came from older more static paradigm, where the data collections are the center and not the function even less collaboration and growing as well collecting additional information of metadata records. These illustrates the example of the purpose field in metadata standards, these filed is optional and could be filled up once per metadata record, but every user who downloads the corresponding data has no chance to add a further purpose to complete the data description and to inform other potential user or the originator for what the data are good. As a second example, the quality of metadata could be specified by the metadata but there is no possibility to specify the quality of metadata itself (self reflexive).
- *(Processes) Not process-orientated:* Metadata standards have intended describing processes in more generic manner. Other information which concerns the process of data collection and processing are placed in other section of huge metadata standard. Even geo data especially is more a process sometimes a long process geo metadata should not describe and support a state but rigorously the genesis of data.

Our assumption concerning metadata is they are essential for a sustainable data management in nature conservation etc. Under this condition, we suggest to improve the value of metadata that user has a direct profit.

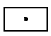

Fig. Increasing the level of benefit of metadata

But not in every circumstance metadata has the same importance. We have isolated two dimension of metadata which influences the interest of metadata. Time and Sharing: Users are not interested to record metadata in its every day use not each data set has to be described. All so called "ad hoc" solutions do not need documentation otherwise geo-data which has broader time range, which has a longer persistency, must be documented. The other dimension is the complexity of sharing. Geo-data individually used must rarely be documented. Collaboration in projects between different organizations (institutions) and inner exchange call for metadata.

3.4 Metadata for the Process Lineage

The core of the DNL database is a flat table that stores a minimum set of information for each data record like data formats, author, and date. Obviously bein non-standardized, this metadata at record level nevertheless ensures data quality and allows to reconstruct the processing steps in temporal order, that are called processing-chains or lineage. A processing step (e.g. publishing a press release, surveying a raised bog object, etc.) is described by an abstract process type.

Since each data item belongs to a specific inventory (cf Chapter 2), it belongs to a process type that is assigned to a processing chain. For example, a document describing a flood-plain will be part of the flood plain inventory, DESCRIPTION OF THE PROCESS OR DELETE THE EXAMPLE. The inventories are split into several inventory-snapshots. For each inventory-snapshot there exists an individual processing chain. A comprehensive overview of the process-oriented database design can be found in [Lanz et al. 2007]

4. Virtual data centre (DNL/VDC)

The VDC is intended to be a computing platform integrating and sharing distributed data for combined retrieval, query, analysis, and display. Therefore, the VDC has to fulfil the following requirements [Frehner and Brändli 2006]:

1. The autonomy of the involved institutions and databases must be guaranteed, including the control over the accuracy of the provided data (Chapter 2).
2. The involved database schemata, maintenance procedures and applications must remain unaffected.
3. The architecture and implementation of the VDC must be extensible and easy to maintain.

To meet these demands, the VDC was implemented as a modular system, mostly realized as a Service Oriented Architecture. The Java EE platform and corresponding web technologies were used to build the so called three-tier architecture (Fig. 1).

The *data tier* integrates distributed data repositories in a central database at WSL where the data is aggregated, stored and updated regularly. While currently proprietary solutions have to be used , this process will be realized using webservices in the coming project phase. Hence this process becomes independent from the remote database schemas while preserving the autonomy of the involved institutions. Furthermore, a change in the external data structures will have no effects on the structure of the aggregated spatial data.

Encapsulated in webservices, the *application tier* contains the business logic. The remote functions include data retrieval as well as GIS functionalities for querying and analysing spatial data.

The *presentation tier* is implemented as a Java web-application running inside an Apache Tomcat application server (http://tomcat.apache.org). Relying upon the JavaServer Faces Framework (REFERENZ), the web-application follows the Model-View-Controller paradigm (REFERENZ).

The VDC provides GIS functionalities, e.g. the mapping and spatial queries, Furthermore, it provides-methods for data retrieval that are realized as two independent modules. The first module includes the "Catalogue Service" and "GM03 Service" and allows querying the metadata of spatial data by strings. The second module, called "Open Search Service" uses a semantic approach to find data, datasets and further documents (Chapter 6). Each module can be accessed via its own User Interface.

The mapping and query functionalities are provided by an installation of ArcGis Server 9.3[ESRI]. While at the start, the VDC only covered the most basic map navigation tools, further methods were adde to the application, in particular VDC-specific query tasks. Enhancements are planned like extended spatial analyses tools according to the requirements of the FOEN.

5. Service Oriented Architecture

The intention of the VDC integrating spatially distributed data and disseminating aggregated data and results (s. above) strongly advises the development towards a Service oriented architecture.
A SOA consist of webservices as the basic constituents (Erl, 2005; Newcomer and Lomow, 2005). They use defined protocols which describe how one or more services can "talk to" each other, instead of services embedding calls to each other in their source code. [Melzer 2007] This main characteristics of an SOA demands a well-structured design of the applications with thorough decisions about the integration and communication of the single software components implemented as services. Meeting these conditions, an SOA supports openness, interoperability, extensibility, reusability and autonomy of these services. Furthermore, the services can be flexibly combined allowing an easy creation of new applications.

The VDC combines spatially distributed data and uses software components and expert knowledge of independent work units, such as GIS specialists, metadata specialists or ontology researchers. Thus a modular setup is highly demanded for. As the VDC is intended to join the components for not only one specific application, but rather an evolving application, it furthermore demands for well defined interfaces. With these facts provide the basis of an SOA. Figure xx gives an overview of the service based architecture of the VDC:

The Catalog Services have been developed mainly with regard to the VDC, though taken into account to be reusable in further applications. It allows to search within some simple Metadata, such as the names of species or its status of protection. The GM03 Service is an implementation of the Swiss Metadata Standard (GM03), thus provides search options for spatial inventory data within a well known information model / standard. It has been developed entirely independent and are being reused in the VDC. The Catalog-Services as well as the Metadata-Service are implemented as web services, typically consisting of a WSDL document in order to provide a machine-readable description of the operations and using XML messages that follow the SOAP standard [Melzer 2007] for communication.

The Ontology-service will provide a sophisticated intuitive search tool that as well addresses users with little knowledge about the available data. It is currently being developed independently and is being reused in the VDC. It is also implemented as a web service with a WSDL document and the use of SOAP messages.

The Map Services provide access and queries of the spatial data. They are implemented as Map Services from ArcGIS Server 9.2 from ESRI (Literaturangabe). They comprise functionalities for Web mapping and Web cartography, but also many further associated methods for detailed data analyses. They have an associated WSDL document and can be accessed via LAN or Internet. They are particularly designed to use with the many ESRI software parts, e.g. ArcGIS Desktop or, as in our case, with ArcGIS Server applications. These services can as well be published as Web Map Services. The download of the spatial and thematical data happens via the Geodata services from ESRI defined similarly as the Map Services, with a WSDL document for description purposes of the supported operations.

However, the SOA in the VDC has not yet been fully implemented. The originally intended access of the spatially distributed data repositories via Web Feature Services (WFS) had to be postponed in favour of pragmatic solutions, especially for transferring external data. (see Figure 1): As mentioned above (see Chapter 2 and Figur 1), these pragmatic solutions range from uploading data to our repository to SSH-access and VPN-clients / access?. Momentarily, the integration of the distributed data goes with spatial enabling of these X/Y-Data and its thematic attributes. The generated spatial data are generalised bounding boxes, of e.g. 2x2km as some data are sensitive and can not be offered with the exact locations the users. Once, all the data are spatially enabled, the database acts as a cache, which provides the spatial data to the application, via map services (see above). The originally proposed WFS according to the OGC standards are still aimed at in the future because they have a variety of advantages in terms of the workflow: They are explicitly designed to visualise, analyse and model spatial data. As it is a special case of a web service the control about which data can be retrieved, would stay with the provider (external institutions). Furthermore, these are spatially enabled data and, therefore, a further control about the spatial accuracy would stay completely with the providers. As for now, the control mechanisms are based on clear arrangements about the parts that we perform for the data providers, such as spatially enabling and generalising the point data.

6. Semantic Search

In order to provide an open and intuitive access to the DNL datacenter, particularly for non-expert users, a semantic search has been implemented.

6.1 Structure of the Ontology

The bilingual ontology is composed of two independent ontologies in German and French. They are related to each other bv means of terminological axioms in terms of equalities, called bridges, such as Moorlandschaft ≡ paysages_marécageux. The bridged ontology holds 1,155 items. These items refer to classes, properties and individuals. The names of the items together with synonyms and similar terms are represented as values of label properties. All label values are nouns in nominative singular and nominative plural form. We only discriminate between proper names, common names and taxons (cf. table 1).

Proper names are used to label individuals. Usually, they are not translated and the unique name assumption (UNA) holds.[1] An example for a proper name is “Plaun Segnas Sut”. As can be seen from this example, names may include several words.

Common names are used to label non-taxonomic classes and properties. Usually, they are translated and the UNA does not hold. Examples for common names are “Moorlandschaft” (moorlands) and “Geometrie” (geometry).

Taxons are used to label classes in taxonomies. While they are usually translated, the UNA still holds, because the translated terms are considered as similar terms and not as synonyms. For example, “Castor” is the latin taxon for the beaver genus.

[1] As they are not translated, it is sufficient to assert individuals in a single ontology. We assert them in the German ontology.

Since labels not only store synonyms but also similar terms, a search term matching a label value of an item experiences a moderate semantic expansion with the readout of the label values. For instance, the terms “paysage marécageux” (moorland) and “marais” (mire), which the item paysage_marécageux in the French ontology is labeled with, do not share the same extension. Moorlands contain mires but also non-mire areas. Conversely, some mires are located outside a moorland. We do not refer to this when using the term “semantic expansion”. We rather refer to logical inferences drawn by a reasoner operating on the bridged ontology (cf. next section).

	Item	Translation	UNA	Example
Proper name	individual	no	yes	“Plaun Segnas Sut”
Common name	non-taxonomic class, property	yes	no	“Moorlandschaft”
Taxon	class in taxonomy	yes	yes	“Castor”

Table xx: Names and their implementation in the ontology

6.2 Ontology-Based Query Processing

Ontology-based query processing involves a sequence of (pre-) processing steps which are described in this section. Consider the situation where a user enters one or more search terms into the search form and submits the query. The search consists of two actions: query expansion and query evaluation.
1. The input is analyzed. The input character string is cut into coherent substrings and each of the substrings is compared with the vocabulary of the ontology. The vocables that match any of the substrings are added to a set of terms which is the source data structure for all further processing. The matches are case-insensitive and insensitive w.r.t. number (singular or plural). The number of input terms is not limited.
2. Each term is semantically expanded. The expansion depends on the type of the term (class, property or individual) and on the conceptual structure specified by the ontology. Based on the assumption that in most cases the user is looking for individuals we apply the following rules: Classes that are not undermost subclasses are expanded to subclasses, undermost subclasses are expanded to individuals, properties are expanded to domain individuals, and individuals are expanded to types (i.e. classes). Applying these rules, a list of expanded terms together with synonyms and similar terms is retrieved from the bridged ontology.
3. The terms are logically connected to each other. Before querying the database, the terms in the source data structure and, optionally, the terms selected from the list of expanded items are connected to each other through the logical connectives AND and OR.
As a result the WHERE clause of the SQL statement which the database is queried with (cf. step 4) is a formula in conjunctive normal form (CNF). The literals of this formula are the vocables which match the search terms, the synonyms of these vocables, similar terms and – except for the individuals – their translations. This formula can be seen as the intermediate data structure of the process.
4. The database is queried. Using the SQL statement generated in step 3 the database is queried. For the prototype the search space was defined as a comment field in one data table, a text field in another data ta-

ble and an extension field in a third data table, the last in order to distinguish sets with data about objects from data sets containing documents. Altogether, these data fields are the target data structure.
According to the categorization introduced by Efthimiadis (1996), the described query expansion is *interactive*, *based on knowledge structures* and *collection dependent*: The user defines the scope of expansion by selecting one or more terms from a list. The class and property hierarchies in the ontology provide a knowledge structure which is dependent on the kinds of data held in the database: The ontology specifies a conceptualization of the domain to which the data in the database belong.

7. Lessons Learned and Outlook

During the development of the DNL in the past years many obstacles had to be overcome. The major lessons learned were:

Project management and organisational structures

- Expect the unexpected: project sponsors, users and data providers constantly adjust to a changing environment and therefore the requirements for the DNL/VDC system change accordingly.
- If there is no direct benefit, no one will take action. E.g. data owners provide information and metadata only upon request.

Data modelling

- The process oriented data model has forced the involved parties to discuss and specify explicitly define the demands from each other and can be used as a documentation (i.e. non-standardized metadata) for each inventory.
- Process oriented data model has structured the temporal development of the inventories
- Grouping the data into inventories is often more complicated than it seems, particularly because of the fact, that a very complex reality is mapped in a rigid database schema.
- Considerable time shifts in the processing of large data sets make it quite difficult to model the temporal aspects, even when rough temporal intervals are used instead of exact times.
- Process-oriented data model allows to archive information about the workflow which otherwise easily sinks into oblivion. Nevertheless the idea of reusable processing chains was not as powerful as planned, mainly due to the fact, that almost every inventory represents some sort of special case.

Metadata

Heterogeneous Datasources

- Sharing information, in spite of the heterogeneity, is of great interest to FOEN and the related institutions since current research questions related to environmental protection may not be approached by isolated analysis and modeling. But the autonomy of involved institutions and databases must be guaranteed: their individual database schemes, maintenance procedures and applications must remain unaffected. A service oriented architecture supports this philosophy.
- WFS was originally the architecture of choice. The main obstacle was, that the data providers are technically not prepared to implement such a service. Furthermore, if large amounts of data are requested and exchanged by Geography Markup Language (GML, OGC, 2004b), as ot os isiaööy dpme om WFS (Literatur: Frehner…), latency can be a problem. For these reasons, currently, a local database serves as a persistent cache to store the data and provide them for the VDC application. The spatially enabled local database using SDE overcomes this shortcomings.
- The technical implementation takes usually much less time than the discussions with the data providers in order to find an agreement on the technologies and data formats.

System architecture

- Even if the system architecture is not fully service oriented yet, the benefits from using web-services instead of a monolithic application become evident: the reusability of services, independence of the providers' platforms.
- The transition from a traditional web application to a service oriented architecture proved to be more time consuming than expected. It takes far more than just wrapping modules into web services. The whole architecture has to be redesigned from scratch.

Semantic search

- The use of a domain ontology composed of two component ontologies in German and French proves to be the right design decision in a bilingual context.
- The ontology must be as comprehensive as possible in order to bridge the gap between the users conceptualizations and the one of the DNL data. Users search terms which are not found in the ontology must be added to the ontology, ideally by the users directly. There must be a possibility to bypass the ontology in case the user enters a term which is not represented but still matches data sets when querying the database.

For the future the further development into a fully service-oriented architecture is planned, particularly the connection to the external data sources over web services. Furthermore the use of the semantic search is planned to be expanded to cover external data sources. Data analysis tools, which will be implemented in the future, will further enhance attractiveness of the DNL/VDC.

8. Acknowledgement

This project was funded by the Federal Office fort he Environment (FOEN). Many thanks to all the former DNL colleagues Andri Baltensweiler, Angéline Bedolla, Martin Brändli, Nick Baumberger, Laurence Cuche, Marcel Frehner, Christian Ginzler, Susette Haegi, Corinna Heye, Jörg Sparenborg and Charlotte Steinmeier.

9. References

Bauer-Messmer, B., Grütter, R. (2009) Semantic modelling of temporal information in a database for biotope inventories. Accepted for publication ITEE 2009, Information Technologies in Environmental Engineering, May 2009, Thessaloniki, Greece.

Bauer-Messmer, B., Scharrenbach T., Grütter, R.(2008) Improving an environmental ontology by incorporating user-input. In: Möller, A, Page, B. and Schreiber, M. (eds): Environmental Informatics and Industrial Ecology, EnviroInfo 2008, Lüneburg, Germany, pp 559-566.

Bauer-Messmer, B., Grütter, R. (2007) Designing a Bilingual Eco-Ontology for Open and Intuitive Search. In: Gómez, J.M., Sonnenschein, M., Müller, M., Welsch, H., Rautenstrauch, C. (eds.): Information Technologies in Environmental Engineering. Springer-Verlag, Berlin Heidelberg 143–152

CSCF Centre Suisse de Cartographie de la Faune, www.cscf.ch

CRSF Centre du Réseau Suisse de Floristique, www.crsf.ch

Efthimiadis, E.N. (1996) Query Expansion. In M.E. Williams (ed.). Annual Review of Information Systems and Technology (ARIST), 31, 121–187, 1996.

ESRI GIS and Mapping Software: www.esri.com

FOEN: Federal Office for the Environment, www.bafu.admin.ch

Frehner, M. and M. Brändli (2006) Virtual database: Spatial analysis in a Web-based data management system for distributed ecological data. Environmental Modelling & Software, 2006(21): p. 1544-1554.

GML: Geography Markup Language from the Open Geospatial Consortium (OGC): http://www.opengis.net/gml

Grütter, R., Bauer-Messmer, B.(2007) Combining OWL with RCC for Spatioterminological Reasoning on Environmental Data. In: OWL: Experiences and Directions (OWLED). CEUR Workshop Proceedings (2007)

Grütter, R., Bauer-Messmer, B., Hägeli, M. (2008a) Extending an Ontology-based Search with a Formalism for Spatial Reasoning. In: Proceedings of the 23rd Annual ACM Symposium on Applied Computing (ACM SAC 2008), March 16-20, 2008, Fortaleza, Brazil.

Grütter, R., Scharrenbach, T., Bauer-Messmer, B. (2008b) Improving an RCC-Derived Geospatial Approximation by OWL Axioms. In: A.P. Sheth, S. Staab, M. Dean, M. Paolucci, D. Maynard, T. Finin, K. Thirunarayan (Eds.), The Semantic Web – ISWC 2008. 7th International Semantic Web Conference, ISWC 2008, Karlsruhe, Germany, October 26–30, 2008, Proceedings. LNCS (Volume 5318, pp. 293–306). Berlin Heidelberg: Springer, 2008.

Lanz, A., Brändli, M., Baltensweiler, A. (2007) A large-scale, long-term view on collecting and sharing landscape data, In: A changing world. Challenges for landscape research, Kienast, F., Wildi, O., Ghosh, S., (Eds), Springer Landscape Series, Vol. 8, 2007, pp. 93-112.

Melzer, I. (2007) Service-orientierte Architekturen mit WebServices. Konzepte – Standards – Praxis. 2. Auflage, Elevier Spektrum Akademischer Verlag, pp. 356.

Scharrenbach, T. (2008). End-User Assisted Ontology Evolution in Uncertain Domains. Accepted for publication, 7th int. Semantic Web Conference ISWC, Karlsruhe, Germany.

Schenker, J. (2008) Datenzentrum Natur- und Landschaft (DNL) Virtuelles Datenzentrum (VDC), INSIDE 4/08, Federal Office fort he Environment (FOEN), 3003 Bern, Switzerland

Vogelwarte: Swiss Ornithological Institute, www.vogelwarte.ch

WSL: Swiss Federal Institute for Forest, Snow and Landscape Research, www.wsl.ch