EnviroInfo 2009 (Berlin)
Environmental Informatics and Industrial Environmental Protection: Concepts, Methods and Tools
Copyright © Shaker Verlag 2009. ISBN: 978-3-8322-8397-1

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

Bettina Bauer-Messmer, Lukas Wotruba, Kalin Müller, Sandro Bischof, Rolf Grütter, Thomas Scharrenbach, Rolf Meile, Martin Hägeli, Jürg Schenker[1]

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

Abstract

The Data centre Nature and Landscape (DNL), originally planned as a long-term database for biotope information, has constantly evolved over the last 11 years. It is now a virtual data centre connecting several databases located all over Switzerland, enriched with a user-friendly web interface, which has been partially redesigned into a service oriented architecture. It can now offer open and intuitive search facilities based on semantic technologies. Many lessons have been learned, particularly in the field of heterogeneous environmental information systems. The purpose of this paper is to share our experience with other experts developing and maintaining environmental information systems.

Keywords: *Enviromental Database, Environmental Information System, GIS, Semantic Technologies, Service Oriented Architecture*

1. Introduction

Data on protected areas of national importance in Switzerland have been, and continue to be collected in thematic inventories which are part of a national database. This database was established according to the requirements specified by the Swiss Nature and Cultural Heritage Protection Act (NHG) (Natur- und Heimatschutzgesetz) [Schenker 2008]. There are central databases for most species in Switzerland: faunistic and floristic data are collected in two centres in Switzerland: Centre Suisse de Cartographie de la Faune CSCF [CSCF] and the Centre du Réseau Suisse de Floristique CRSF [CRSF]; and data on fungi, lichen and bryophytes are stored in central databases. For protected biotopes and sites, however, no central database has been established so far. Data has been collected in different decentralized locations in heterogeneous formats. In 1997 the Federal Office for the Environment [FOEN] decided to set up a centralized database for protected areas in cooperation with the Swiss Federal Institute for Forest, Snow and Landscape Research WSL [WSL]. This is how the project Data centre Nature and Landscape (DNL) started.

In the first phase, 1998-2001, a process-oriented database was set up. It became clear early on that it would be useful to combine the DNL data with other data from national databases. This was done in a second phase, 2002-2009, by extending the DNL and creating a virtual data centre (DNL/VDC). A web interface allows to search in the metadata and data of the DNL easily and use GIS functionality. Furthermore, it makes use of semantic technologies, which will be further developed such as the adaption to the conceptualizations of the users [Bauer-Messmer et al. 2008][Scharrenbach 2008] . For the third stage, 2010-2013, a service oriented architecture is planned, and first steps in that direction have already been taken.

The purpose of the DNL is to manage all data regarding protected areas of national importance and to further expand scientific inventories. The three major goals include:

1. long-term archiving and, in particular, storing context and processing information in metadata (section 2).
2. accessing and aggregating data from heterogeneous data sources from national databases of FOEN in a virtual data centre (see section 3) using a service oriented architecture (section 4).
3. opening up the data collection for intuitive search using semantic technologies (section 5)

Section 6 concludes this paper with an overview of the lessons learned in the past 11 years of development and provides an outlook.

2. Long-term use of data: Metadata in DNL

The database should essentially allow the reconstruction of the complete genesis of the data records [Lanz et al. 2007]. Data must be available for administrative and scientific purposes, as well as for issues in the public domain. As a consequence, the traceability of data must be assured for coming generations and for use in different contexts. These two objectives are achieved by describing the data using non-standardized and standardized metadata, such as FGDC [FGDC] or ISO 19115 [ISO 19115]

2.1 Non-standardized metadata for the process lineage in the DNL

The core of the DNL database is a table in which some minimum information is stored about the data format, author, date, measuring devices, data processing tools, etc. for each data record. This non-standardized metadata at record level helps ensure data quality and further allows the reconstruction of the processing steps in temporal order, in so called ‘processing chains’ or ‘lineages’.

The processing steps (e.g. publishing a press release or surveying a raised bog object) are described by an abstract process type.

Many documents, such as PDFs describing a flood plain object, are also stored in the database. Each of these data items belongs to a specific inventory as well as to a process type that’s part of a processing chain.

The inventories are split into several inventory-snapshots. For each inventory-snapshot there is an individual processing chain. A comprehensive overview of the process-oriented database design can be found in Lanz et al. [2007]

2.2 Standardized Metadata: the Swiss metadata standard GM03

The Swiss metadata standard GM03, which is fully derived from the ISO 19115 [ISO 19115] was implemented in the DNL database. In contrast to the ISO standard the Swiss standard supports multilingual data descriptions, such as title, abstract or keywords, and allows the delivery of metadata to the national geodata portal (geocat.ch) [KOGIS 2005]. The role of the KOGIS in Switzerland is roughly comparable to the one of INSPIRE [INSPIRE] in Europe or the USGS’s Spatial Data Transfer Standardization [USGS] in the United States. The GM03 metadata is provided via web services and can be reused in different applications such as the VDC.

Metadata has relationships with three other elements in an environmental information system: applications, processes and data. We focus on the three relationships illustrated with bold lines in figure 1 on the left.

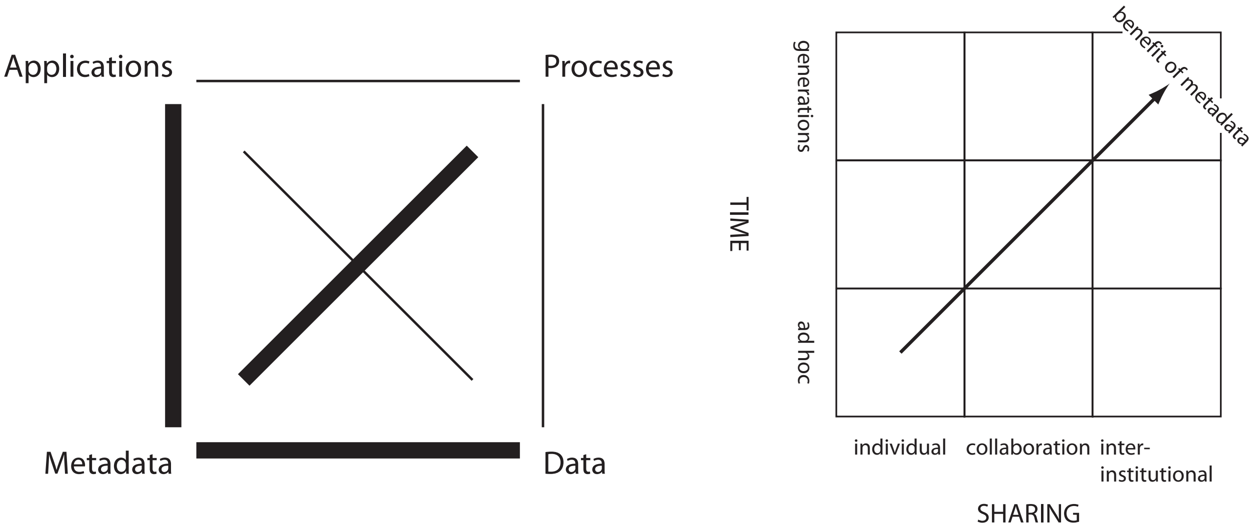

Figure 1 Relationships between elements of an environmental information system (left), and increasing the benefit of metadata (right).

Metadata are essential for sustainable data management. The value of metadata increases when data are persistent and many contributors are involved in collecting and processing the data (fig. 1, right). A further key factor in metadata collection is the direct benefit for users. Thus a catalogue should not be the final purpose of standardized metadata, indeed, further use of metadata is encouraged, including collaboration or collective intelligence applications. Software-engineers must integrate the metadata in applications and in user's everyday work and thus further increase the value of collecting and managing metadata. Benefits to users include an enhanced metadata standard that extends the usefulness of metadata by opening the metadata semantically for non-experts [Bischof & Bauer 2008].

3. Heterogeneous data sources: data from national databases

There are three types of data providers: data in the DNL database itself, data from other research units at WSL and data from other data providers all over Switzerland (see Table 1).

3.1 Data in the DNL database

The major inventories comprise the following biotope types: bogs, fens and amphibian spawning grounds and alluvial zones. Inventories range from a few hundred objects (e.g. in the bog inventory) to a few thousand objects (e.g. in the fen inventory). The DNL data comprises roughly 300,000 data items, including several thousand spatial objects and thematic entries and several hundred orthophotos and PDF documents.

The DNL database not only holds thematic data but also spatial data within a GIS system, as well as temporal data, either as explicit dates or implicitly within processing chains of the lineage [Bauer-Messer et al. 2009].

3.2 In-house databases at WSL

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

1. Other data sources

The flora data of CRSF, fauna data of CSCF and bryophyte data are hosted in external institutions. External data sources are located in corporate networks that apply heterogeneous security settings. Several restrictions apply in transferring the designated information to the central repository at WSL. In particular three different security setups of the external data providers need to be considered, that restrict access:

1. No external access from outside to the corporate network
2. Restricted access via proprietary VPN-clients
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 involving a VPN-client must be chosen. A SQL-client software then gathers the data and stores it in the repository database. In the third case, an SSH connection is automatically opened with a tunnel to the data providers' database. The central repository database then collects the necessary datasets from the remote server via a database link. Considering the time and effort needed to adjust the processes to any changes in security settings, a more service oriented approach is planned (see section 4.2).

Data repository and institution	Characteristics of data and storage system
DNL, Swiss Federal Institute WSL, Birmensdorf	Vector data representing boundaries of biotopes and other areas, 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 animals. Oracle database management system with coordinates stored in regular columns (not spatially enabled)
Flora database, CRSF, City of Geneva	X/Y-coordinates representing observations of plants. Coordinates are stored in regular columns (not spatially enabled).
Bryophyte database, University of Zurich	Vector data representing approximate location (polygon) or exact X/Y coordinates of bryophyte samples. Geometric data stored in ESRI-Shapefiles, with corresponding attribute data in Oracle database tables.
Fungi database (swissfungi), WSL, Birmensdorf	X/Y coordinates representing fungi occurrences. Oracle database management system, with coordinates stored in regular columns (not spatially enabled). Coordinates specified with positional accuracy.
Lichen database (SwissLichens), WSL, Birmensdorf	X/Y coordinates representing lichen occurrences. Oracle database management system, with coordinates stored in regular columns (not spatially enabled), specified with positional accuracy.

Table 1: Summary of the data repositories in the DNL/VDC (adapted from Frehner et al. [2006])

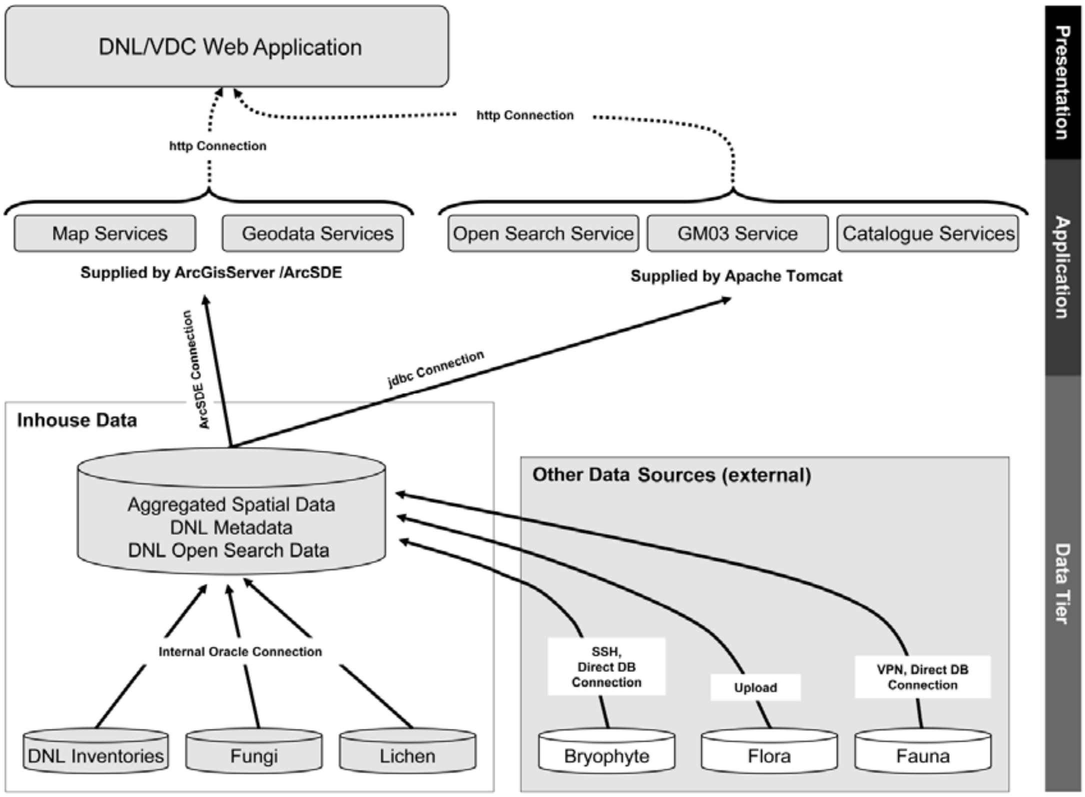

Figure 2: Overview of the main components of the DNL/VDC

4. Virtual data centre (DNL/VDC) and its service oriented architecture

4.1 Virtual Data Centre (DNL/VDC)

The main purpose of the VDC is to act as a central point of access to distributed environmental data sources. The VDC is intended to be a computing platform which integrates and shares distributed data for combined retrieval, query, analysis, and display. To accomplish these goals, the VDC has to ensure that:

- the autonomy of the involved institutions and databases is guaranteed, including control over the accuracy of the data provided;
- the database schemas, maintenance procedures and applications involved must remain unaffected;
- the architecture and implementation of the VDC can be extended and is easy to maintain.

This is where a modular architecture is very effective. [Frehner and Brändli 2006]

We used a modular software design, with service oriented software principles. The architecture of the VDC is based on web service technologies, open standards, open interfaces and the reuse of open source software components. We built an architecture known as three-tier architecture into the VDC using the Java EE platform (see fig. 2).

The data tier consists of distributed and integrated data repositories. These data are aggregated, stored and updated regularly in a data cache located at WSL. In the near future, this process should take place happen via web services, allowing us to remain independent of remote database schemas. The autonomy of the institutions involved would be guaranteed (see chapter 3). Furthermore, a change in the external data structures would not affect the structure of the aggregated spatial data, as the interface of the web services would have been defined.

The application tier contains the business logic encapsulated in services performed on the web server or delegated to an application server. The delegated functionalities are aimed at data retrieval and GIS functionalities for querying and analysing spatial data. The application Server is based on ArcGis Server 9.3 [ESRI].

The presentation tier is implemented as a Java Web application running inside a Tomcat application server (tomcat.apache.org). We used the JavaServer Faces framework to build the user interface according to our application logic. With this architecture we have been able to achieve the desired requirement of providing flexibility and easy maintenance.

The functionalities in the VDC cover both GIS functionalities, e.g. the mapping and query functionalities, and on functionalities for data retrieval. The mapping and query functionalities are based on ArcGIS Server functionality. Initially, the VDC only covered the most basic map navigation tools. We have since added further methods to the application, in particularly VDC-specific query tasks, which we plan to enhance with extended spatial analyses tools according to FOEN’s requirements. The mapping and query functionalities can be accessed by separate tabs in the user interface.

The first module includes the “Catalogue Services” and "GM03 service", and allows the user to query the metadata of VDC’s spatial data via strings. The second module, called the “Open Search Service”, uses a semantic approach to find data, datasets and other documents (see chapter 5).

4.2 Web Services and Service Oriented Architecture (SOA) in the VDC

As mentioned above, the VDC aims at integrating spatially distributed data and disseminating aggregated data and results. It is appropriate for working with web services and developing a SOA.

SOAs take web services as their 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 embedding calls to each other in their source code. This is one of SOAs’ main characteristics. It implies that the design of the applications must be well structured and clear decisions about the integration and communication of the single software components implemented as services must be made. Meeting these conditions requires enabling services to be open, interoperable, extensible, reusable and autonomous. This allows flexibility in the composition of new applications [Melzer 2007].

The VDC combines spatially distributed data. In addition, it uses software components as well as the expert knowledge of independent work units, such as GIS specialists, metadata specialists or ontology researchers, which makes a modular setup highly desirable. As the VDC is intended to integrate the components for an evolving system and not for a frozen application, it also requires well-defined interfaces. These factors provide the basis for a SOA. Figure 2 gives an overview of the service-based architecture of the VDC, described in the next paragraph.

The Catalogue Services have been developed mainly for the VDC, but it should also be possible to use them in further applications. They allow searches within simple Metadata, e.g. to find the names of species or their status of protection. The GM03 Service is an implementation of the Swiss Metadata Standard (GM03). It accordingly provides search options for spatial inventory data within a well-known information model. It was developed entirely independently and has been applied in the VDC. The Catalogue Services and the GM03 Service are implemented as web services, typically consisting of a WSDL document, which provides a machine-readable description of the operations and uses XML messages that follow the SOAP standard for communication [Melzer 2007].

The Open Search Service provides a sophisticated intuitive search tool that also permits users with little knowledge about the available data to access the database. Like the GM03 Service, it was developed independently and has been adopted 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 to the spatial data and allows them to be queried. They are implemented as Map Services in the ESRI's ArcGIS Server 9.2 [ESRI], and provide functionalities for Web mapping and Web cartography and provide a range of tools for detailed data analyses. They have an associated WSDL document and can be accessed via LAN or Internet. They have been specially designed so that they can be used in combination with the many ESRI software parts, including the ArcGIS Desktop or, in our case, ArcGIS Server applications. These services can also be published as Web Map Services. Downloading of the spatial and thematic data goes via ESRI's Geodata Services, which are defined similarly to the Map Services, with a WSDL document for describing the supported operations.

Currently, the integration of external data is achieved by spatially enabling these X/Y-data and their thematic attributes. The generated spatial data are generalised bounding boxes since some data are sensitive and cannot be offered to users with the exact locations. Once all the data have been spatially enabled, the database acts as a cache providing the spatial data to the application, via the Map Services (see above).

5. Semantic Search

In order to provide an open and intuitive access to the DNL datacenter, particularly for non-expert users, a semantic search was implemented [Grütter et al. 2008].

5.1 Structure of the Ontology

The bilingual ontology consists of both a German and a French ontology [Bauer-Messmer et al. 2007]. Although the two ontologies are largely independent, they are related to each other by means of terminological axioms expressed as equalities, called bridges, such as "Moorlandschaft ≡ paysages_marécageux". The ontology contains 1,155 bridged 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 the nominative singular or the nominative plural form. We only discriminate between proper names, common names and taxons (cf. Table 2).

Proper names are used to label individuals. They are usually not translated and the unique name assumption (UNA) holds.[1] An example of a proper name is "Plaun Segnas Sut", which indicates that names may include several words.

Common names are used to label non-taxonomic classes and properties. They usually are translated and the UNA does not hold. Examples of such common names are "Moorlandschaft" (mire landscapes) and "Geometrie" (geometry).

Taxons are used to label classes in taxonomies. Although 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.

Since labels not only store synonyms but also similar terms, a search term matching a label value of an item is subject to a moderate semantic expansion with the readout of the label values. For instance, the

---

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

terms “paysage marécageux” (mire landscape) and “marais” (mire), which are labels used with the item `paysage_marécageux` in the French ontology, do not share the same extension. Mire landscapes contain not only mires but also non-mire areas. Conversely, some mires are located outside a mire landscape. This is not what we mean by the term “semantic expansion”. Rather, we refer to the 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 2: Names and their implementation in the ontology

5.2 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 (1,2) and query evaluation (3,4).

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 the set of terms that serves as the source data structure for all further processing. The matches are case- and number-insensitive (singular or plural). There is no limit to the number of input terms.

2. Each term is semantically expanded. The expansion depends on the type of 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 the lowest subclasses are expanded into subclasses, the lowest subclasses are expanded into individuals, properties are expanded into domain individuals, and individuals are expanded into 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 used to query the database (cf. step 4) is a formula in the 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 can then be queried. For the prototype, the search space was defined as a comment field in one data table, a text field in another data table and an extension field in a third data table, the third data table was included in order to distinguish sets with data about objects from data sets containing documents. Considered together, these data fields are the target data structure.

According to the categorization introduced by Efthimiadis (1996), this 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.

6. Lessons Learned and Outlook

During the development of the DNL in recent years, many obstacles have had to be overcome. The main lessons we learned are summarised below:

Project management and organisational structures

- Expect the unexpected: project sponsors, users and data providers have to constantly adjust to a changing environment and the requirements for the DNL/VDC system therefore change accordingly.
- If there is no direct benefit, no one will take action. For example data owners tend to provide information and metadata only after repeated requests.

Data modelling

- The process-oriented data model forces the parties involved to discuss and clarify what is required and can be used as documentation for each data collection (e.g. inventory).
- Grouping the data into a framework (such as inventories) is often more complicated than it seems, particularly because a very complex reality is mapped onto a rigid database schema.
- There are several obstacles in temporal modelling: The time shifts involved in processing large data sets tend to be considerable, the temporal granularities (days, months, years) tends to mixed and often only vague data is available.
- A process-oriented data model allows the information about the workflow (which might otherwise be easily lost) to be archived. Nevertheless, the reusable processing chains turned out to be not as powerful as we had hoped, mainly because most of the inventories represent some sort of special case.

Metadata

- Metadata collection and management has no impact if this rich source is not used for enhanced functionality. The metadata must be tailored for specific use, bridging the gap between data and metadata. Meeting the standards is a secondary goal. The information contained in the metadata should be rigorously applied to bridge the gap between data and metadata.
- *Mainly used for catalogues:* The main application of metadata is undoubtedly the catalogue. This leads to unwanted effects, since the metadata are rarely integrated into related applications but tend to be kept in separate systems.
- *No support for collective intelligence and collaborations:* The metadata standards allow enhancements concerning the structure and content. However the standards were not originally drawn up to support more recent developments in of collaboration and collective intelligence in computer science [Segaran 2007].
- *Not process-oriented:* Metadata standards were originally intended to describe processes in a generic manner. Information about the processing chains (also called lineage) of data collections are spread over different section of the standards. The ISO standard does not contain a full model to describe processing chains with predecessors and successors. The creation of environmental data, however, is a relatively complex processing chain, which means information about lineage shoud be an integral part of the data archive.

Heterogeneous data sources

- FOEN and other similar institutions are very interested in finding ways of sharing information and data about the environment, even though they tend to be very heterogeneous, since current research questions related to environmental protection cannot be tackled through analysing and modelling isolated data. The autonomy of the institutions and databases involved must, however, be guaranteed and their individual database schemes, maintenance procedures and applications must remain unaffected. A service oriented architecture supports this philosophy.
- The SOA has not yet been fully implemented in the VDC. It was originally intended to provide access to the spatially distributed data repositories via Web Feature Services (WFS)

[Frehner and Brändli 2006] but this has had to be postponed in favour of pragmatic solutions, especially for transferring external data (see fig. 2). WFS have a variety of advantages in terms of workflow: They are explicitly designed to visualise, analyse and model spatial data. As they are a kind of web service, the control over the data to be retrieved and over their spatial accuracy remains with the providers, i.e. with external institutions. At the moment the control mechanisms are based on clear arrangements about which tasks WSL performs on behalf of the data providers, such as spatially enabling and generalising the point data.

- The technical implementation usually requires much less time than the negotiations with the data providers to reach an agreement on the technologies and data formats to be used.
- System administrators, network security managers, application programmers, data providers and users typically have different goals, which makes it often quite difficult to find reasonable solutions.

System architecture

- Even if the system architecture has not yet been made fully service oriented, using web services instead of a monolithic application has clear benefits, for example, being able to services and ensure the independence of the providers' platforms.
- The transition from a traditional web application to a service oriented architecture proved to be more time consuming than originally expected, as it involves far more than just wrapping modules into web services. The whole architecture has to be redesigned from scratch.

Semantic search

- The use of an application ontology composed of two component ontologies in German and French has proved to be the right design decision in a bilingual context. The advantage is that the dialogue language at the user interface is “automatically” specified by the language in which the search terms are entered. Only for the database are query terms from both component ontologies used.
- Currently, search terms that are not represented in the ontology are excluded from searches. In order increase the probability that users will use terms as conceptualized in the DNL domain, the ontology must be as comprehensive as possible. It should also be possible to bypass the ontology when querying the database in instance where the user enters a term that has not been represented but still matches data sets. Finally, mismatching terms should be captured and frequent mismatches incorporated in the ontology as it evolves. This latter point has been addressed by a recent project at WSL [Bauer-Messmer et al. 2008, Scharrenbach 2008].
- The ontology has been enriched with a hierarchy of spatial relations [Grütter & Bauer-Messmer 2007] as introduced by the Region Connection Calculus (RCC) [Randell et al. 1992]. It is planned to add temporal relations taken from Allan's interval algebra [Allen 1983], which should provide a logical basis for processing combined, spatio-temporal-thematic queries.

We plan to develop the DNL further by implementing a fully service oriented architecture and particularly by improving the connection to the external data sources via Web Services. Once the external data will be fully accessed via Web Services, the DNL/VDC will be developed towards a GIS Mashup enhanced with thematic data and semantic search facilities.

Since 2008 test users work with the DNL/VDC and give feedback, which helps to improve the quality of the data centre. Currently only public data is accessible over the DNL/VDC. For the future however it is planned to add critical data to the DNL/VDC, which will be accessible only to certain user groups. Particularly, the cantonal administrations are supposed to see only data relevant for their cantons.

Better semantic search facilities, as well as new analytic tools, will further enhance the attractiveness of the DNL/VDC in near future for all user groups including public non-expert users.

Acknowledgements

This project was funded by the Federal Office for the Environment (FOEN). Many thanks to all our 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. Many thanks to Silvia Dingwall for proof-reading the manuscript and thanks to the three anonymous reviewers for their helpful suggestions.

References

Allen, J.F. (1983) Maintaining Knowledge about Temporal Intervals, Communications of the ACM, Vol. 26, Nr. 11, November 1983, 832-843.

Bauer-Messmer, B., Grütter, R. (2009) Semantic modelling of temporal information in a database for biotope inventories. Accepted for publication ITEE 2009, Information Technologes 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

Bischof, S. and Bauer-Messmer, B. (2008): Semantic Enhancement of Environmental Metadata. In: Möller, A.; Page, B.; Schreiber, M. (eds) Environmental Informatics and Industrial Ecology. Proceedings of the 22nd International Conference on Informatics for Environmental Protection. September 10-12, 2008, Leuphana University Lueneburg, Germany. Aachen, Shaker. 123-131.

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.

Erl, T. (2005). Service-oriented archtiecture. Concepts, technology, and design. Upper Saddler River, N.J.: Prentice Hall.

ESRI (Environmental Systems Research Institute). (2009). ArcGIS - Server GIS. Retrieved March 3, 2009, from **http://www.esri.com/>http://www.esri.com**

FGDC: Federal Geographic Data Committee, http://registry.gsdi.org/

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.

GM03: Schweizer Metadatenmodell, http://

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., Frehner M., (2008) First Experiences with an Ontology-Based Search for Environmental Data. In: Proceedings of the 11th AGILE International Conference on GI Science, 6-8 May 2008, Girona, Spain.

INSPIRE: Retrieved June 3rd, 2009 from **http://www.inspire-geoportal.eu/**

ISO 19115 (2005): Geographic information – Metadata.

KOGIS (2005): GM03 – Metadatenmodell. Ein Schweizer Metadatenmodell für Geodaten. Version 2.3 (Final Version).

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., Werner, S., Sauter, P., Hilliger von Thile, A., Flehmig, M., Zengler, B., et al. (2007). Service-orientierte Architekturen mit Web Services. Konzepte - Standadards - Praxis. 2. Auflage. München: Elsevier GmbH, Spektrum Akademischer Verlag.

Newcomer, E., & Lomow, G. (2005). Understanding SOA with web services. Boston, Mass.: Addison-Wesley. OGC: Open Geospatial Consortium, http://www.opengeospatial.org/

Randell, D.A., Cui, Z., Cohn, A.G. (1992) A Spatial Logic based on Regions and Connections. In: Nebel, B., Rich, C., Swartout W. (eds.) Principles of Knowledge Representation and Reasoning, pp. 165–176. Morgan Kaufmann, San Mateo, CA, 1992.

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 for the Environment (FOEN), 3003 Bern, Switzerland

Segaran, Toby (2007): Programming Collective Intelligence. Building Smart Web 2.0 Applications. O'Reilly.

USGS: U.S. Geological Survey, both retrieved June 3rd, 2009 from **http://www.usgs.gov/** and Spatial Data Transfer Standard http://mcmcweb.er.usgs.gov/sdts/

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