

# Effective and Intuitive Search of Spatial and Spatially-Related Data

Susette Haegi, Martin Brändli

## Abstract

The database of the “Data Center for Nature and Landscape” stores data of endangered biotopes in Switzerland. These biotopes usually consist of biotope objects known by name, spatial and temporal extent. In addition, text documents and images illustrating the objects, and historical information documenting their origins are included in the database. The amount of these spatial, temporal and spatially-related data increased substantially during the last years. In order to enable comprehensive exploration of the database, a user interface for an effective and intuitive search is required.

We present an approach which is based on the Google search and response interface. In addition to the usually provided result consisting of a list of hyperlinked resources, results of spatially-related data are correspondingly labeled and simultaneously presented by their location on a map served by ArcIMS. Hyperlinks for further database exploration may be either selected from the resulting list or map.

## Introduction

Efficient and effective data retrieval is a crucial part of comprehensive data management, in particular for complete data exploration and analysis. Since the data repositories of today’s geographic information systems (GIS) and databases continually increase, particular emphasis has to be put on data retrieval techniques which ensure that all essential information may be experienced. Skilled GIS users normally have a profound knowledge of their data. They can apply the processing functionality offered by a local system in order to access and analyze their data. However, non-GIS experts with only limited information of the content of accessible data repositories need reliable support for data exploration. In addition, more demanding requirements of users of geographical information as well as technical progress in the Internet technology contribute to a high dissemination of spatial data on the Web. Providing techniques for finding essential and relevant information in this case is a major challenge in order to fully satisfy users of geographic data.

In this paper we address this problem from the point of view of the user interface design. We present a new approach for comprehensive data exploration using an effective and intuitive data search technique. The technique brings together the traditional mapping metaphor provided by map server applications (ArcIMS) and the very popular and familiar Google interface for information searching on the Internet. The approach in particular extends map server-based web applications which are normally based on predefined tools and navigation capabilities with the “meandering” way of information traversal provided by more general web tools. The discussion, design and implementation of methods which guarantee completeness of data retrieval for a given theme are not part of this paper. Corresponding approaches are discussed in Cartwright (2004) for instance. The work described here originates in the long term project “Data Center for Nature and Landscape“ (DNL) which is conducted in collaboration with the Swiss Agency for the Environment, Forests and Landscape. The project aims at the design and implementation of a database for the storage and management of data of environmental protection areas such as mires, flood plains and semi-arid meadows, with a particular focus on the storage

of the history or lineage of the data. Currently, the database has been developed and is in a mature state. Spatial data is accessible by a web application based on ArcIMS which is extensively used by a broad range of users of different GIS skills. Feedback of the user community revealed the need for a comprehensive and more intuitive way of exploring the data contained in the database.

The paper first describes the features of the database and the current web application. Second, the concept, design and implementation of the new search technique are presented and illustrated with some examples. We conclude with suggestions for further work.

## **Data center for nature and landscape (DNL)**

### **Data model**

In order to show the richness of available data of the DNL and the need for techniques outlined in the introductory section we dive into the description of the requirements and features of the DNL-database. The goal of database design of the DNL was to provide a recording and archiving tool for actions and decisions of environmental protection measures particularly originating from the Swiss Agency for the Environment, Forests and Landscape. The database should essentially enable to follow the complete genesis of existing data records.

Endangered biotopes, such as mires, flood plains and semi-arid meadows are usually managed by inventories including biotope objects known by name, number, spatial extent and additional characteristics such as type of vegetation, etc. Many of these biotopes are subject to a specific set of protection goals and measures. The selection of environmental protection areas normally consists of several steps, triggered for instance by a governmental decision which initiates the selection process. In a next step, scientific experts determine criteria for choosing appropriate areas. Subsequently, base documents (images, maps) are used for digitizing potential protection areas. During succeeding field work, potential areas and their extents are verified, followed by a refinement of digital objects. Finally, a legal document published by the government determines the exact position and extent of the protection areas as well as protection measures.

Storing and communicating these facts shall deliver comprehensive knowledge about the realization procedure. The result was a so-called process-oriented data model which structures the above mentioned selection procedure for protection areas into clearly identifiable and distinguishable steps (Baumberger and Hägeli, 2000, Lanz et al. 2005). Every step is considered an individual process which is related to a certain type of data and corresponding metadata. Data may consist of geometric objects which represent the extent of protection areas, base documents such as images, documents containing the criteria for selecting protection areas and governmental documents which build the legal base for the areas. Metadata is provided according to the ISO International Standard 19115 for metadata concerning geographic information (International Organization for Standardization 2003). To describe the complete history or lineage of a dataset relationships between processes must be defined (Brändli, 2000). Currently, processes may be associated by defining predecessor and successor processes. As a result, processes are connected as directed graphs which allow tracking of related processes. Figure 1 gives an example of a simple graph that shows the origin, processes and involved datasets of a particular protection area.

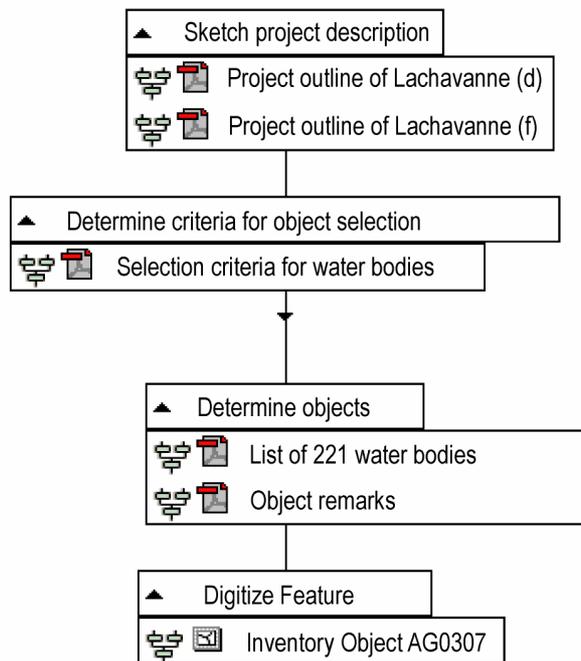


Figure 1: An example of a simple directed graph which represents the history of creating the inventory object AG0307. The graph shows the relations between the different process classes (for example “Determine objects”) and the corresponding processes. These processes consist of several pdf-documents such as selection criteria and further remarks for object delineation determining the extent of the inventory object.

### Web-based access using mapping metaphor

Visual access of data and metadata is provided by a web application which is described in detail in Baltensweiler and Brändli (2004). The web application is based on out-of-the-box components of ESRI’s Internet mapping software ArcIMS and additional software components of the Java 2 Platform including JavaServer Pages (JSP) and Java Servlets. The application makes use of the traditional mapping metaphor since primary access of data is based on data selection from provided maps. Web-content is prepared and rendered by either using JSP tags from the ArcIMS JSP Tag library or self-developed custom JSP tags and servlet actions. The client interface (see figure 2) consists of a toolbar for interactive GIS functions, a map area and a table of content which organizes the inventories as map layers in a folder structure. Metadata about a specific inventory according to the ISO metadata standard can be retrieved by clicking on an icon which is associated to each layer. This act invokes a JSP custom action which executes the necessary SQL statements on the database and generates a resulting new web page. To obtain information for individual inventory objects, the user can perform a spatial selection on the active layer. Again a JSP custom action is invoked which provides an attribute table containing attribute values and hyperlinks to related base documents or images. The hyperlinks point to subsequent custom actions which retrieve the associated documents from the database. For extracting the lineage or history of a data set the hyperlink *Lineage* may be selected and a web page according to figure 1 is presented.

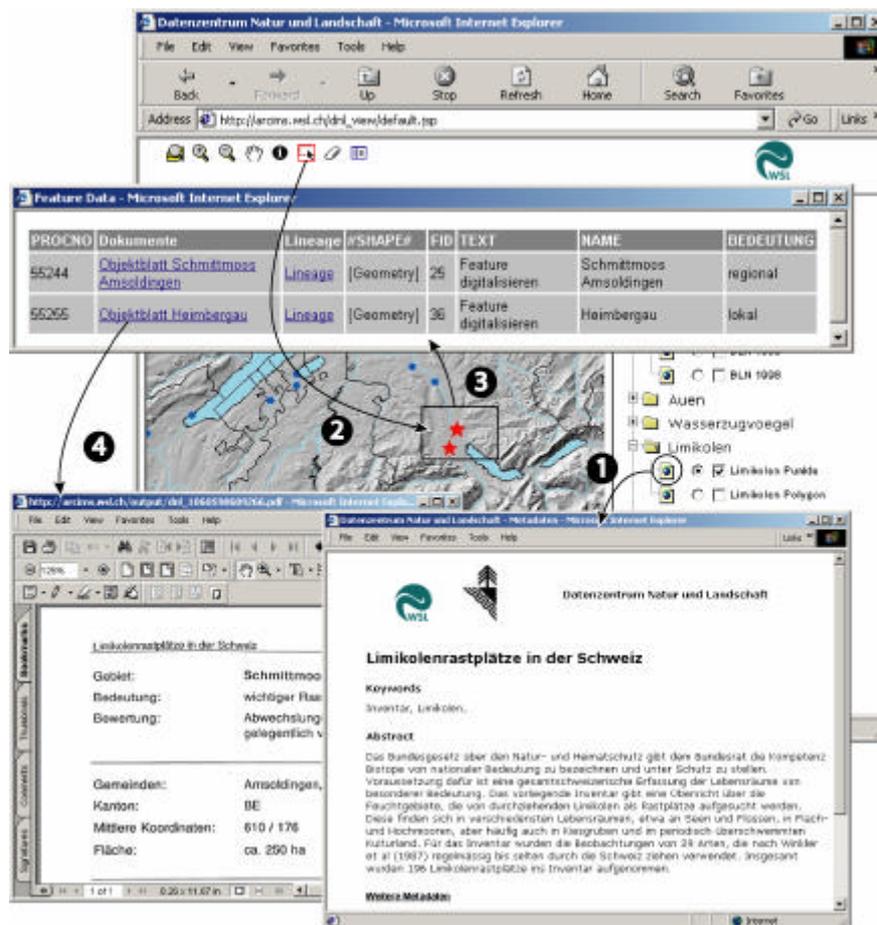


Figure 2: Web application interface. Visualizing metadata: (1) Retrieves metadata of a specific layer (2) performs a spatial selection for retrieving (3) tabular data and metadata (4) link invokes custom servlet to retrieve base documents of a specific object (from Baltensweiler and Brändli 2004).

## Intuitive search

### Concept

Data access presented in the last section uses the mapping metaphor. The user selects one or more particular objects from a data layer and browses data along predefined navigation paths (hyperlinks, menus, etc.). The current application obviously offers no search engine which would allow exploring data based on keywords or other advanced searching possibilities. Extending the existing web application with a comprehensive and intuitive search tool enables a more direct access to stored data. *Intuitively* in our case means that users may apply search or work techniques which they are familiar with by their daily work. This should help that users of different GIS and database skills may query the database without knowing the particular process-oriented data model behind the scenes. The very popular and probably most used tools for accessing information from the Internet are the search portals offered by Google or others. The familiarity of users with these tools leads to the idea to show the results of queries both as their location on a map, next to hyperlinks for further information. This technique is based on the idea of Google Local (<http://local.google.com>). Google describes the concept as following: “Sometimes

the information you're looking for is related to a particular place (...). Google Local locates neighborhood stores and services by searching billions of pages across the Web, then cross-checking those results with Yellow Pages data to pinpoint the local resources you want to find" ([http://local.google.com/help/faq\\_local.html](http://local.google.com/help/faq_local.html)).

The goal of the response interface is to combine the results of a typical Google-search with the results of a spatial selection provided by a GIS. Results are both presented as a list of hyperlinks pointing for instance to documents or images and corresponding objects drawn on the map.

The above quotation of Google Local shows the general potential of this approach but also reveals the fundamental problem: In order to make a web page accessible and comparable to other web pages within a spatial framework, the geographical scope of the web resource must be known. Different approaches for assigning localization information are proposed in literature such as manual classification, indexing based on area or zip codes extracted from the web document, the use of gazetteers based on identified names, etc. For a thorough discussion of these techniques see for instance Jones et al. (2004), Ding et al. (2000), or Buyukkokten (1999).

While Google depends on the Yellow Pages for calculating the geographic scope of a web page, many of the data stored in the DNL and retrieved as results implicitly include a geographical reference, such as the polygons of inventory objects. These data records do not need any further handling. The remaining data, such as documents, documenting images and others are linked by the process-oriented data model presented in the last section. Assigning a geographical scope to these resources is achieved by traversing the lineage chain until a data record with a spatial extent is found.

### **Implementation**

The implementation of this approach extends the web application presented in the previous section (see figure 2). ArcIMS serves as the base for handling spatially-related requests. Java classes and custom JSP tags are used to implement the handling of keyword-based queries and to generate the combined presentation of results on the map and the list of hyperlinks.

The search input field is placed next to the other tools of the web application. At the moment there is only one query field implemented for querying the database. In the future there may be extensions for filtering the data by biotope, by location or by other constraints. The realised search tool allows multiple search terms. Added words are treated with the AND-operator. The possibility to avoid a term can be realized by adding a minus sign ('-') in front of the word. An exact phrase is expressed by putting quotation marks around the search term. Words, if written in upper or lower case, are found in the database, only the exact phrases with the quotation marks need the correct writing of the words.

The results of a query are shown in figure 3. The results are simultaneously presented and labeled objects by their location and extension served by ArcIMS and with hyperlinks for further information. Figure 3 shows the map after a search with the keyword 'Alpen', displaying four objects, labeled by their object number.

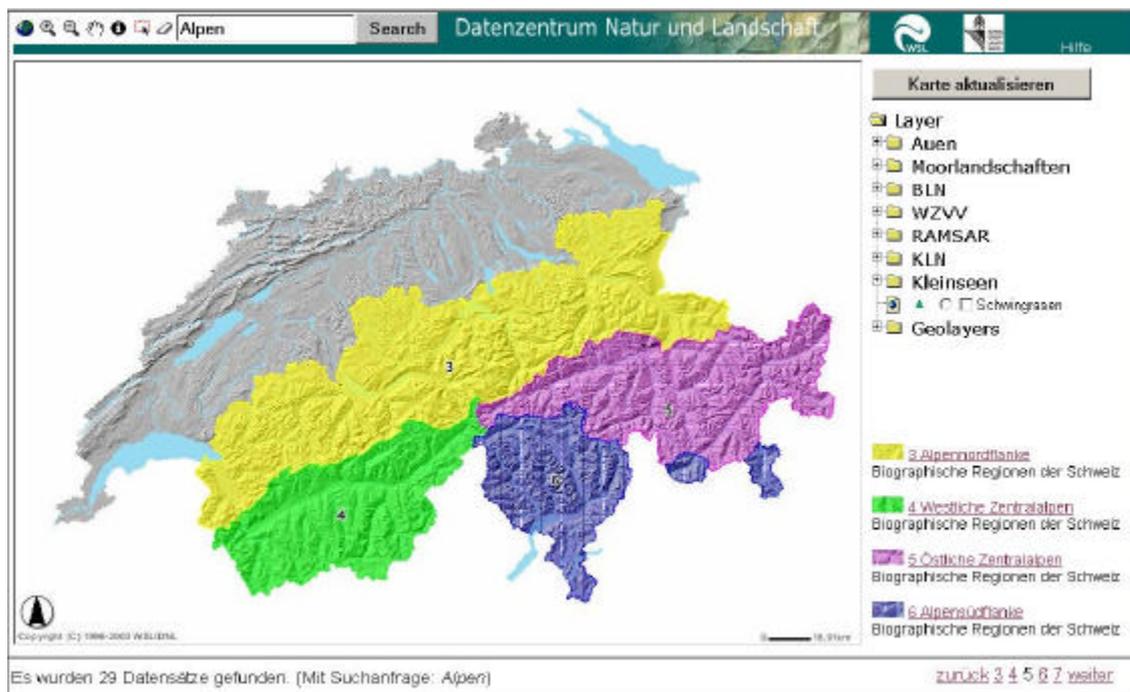


Figure 3: Results of a query with the keyword 'Alpen'. Four polygons of inventory objects are placed and labeled (by their object number) on the map. Clicking the hyperlinks on the right side or the polygons (labels) in the map offers further information about the chosen object (see figure 4). The numbered hyperlinks at the bottom allow the navigation between the results.

Results may be further explored by either clicking the hyperlinks or the objects (labels) in the map. This further information is collected in a table which may contain other hyperlinks for more details. Figure 4 shows two rows of different search results. Like in Google, clicking the numbered hyperlinks at the bottom allows to go to the next page of results. Again, these next results are presented as labeled objects on the map and as a list of traversable links.

OBJNO	PROCNO	PROCTYP	INVENTAR	KOMMENTAR	DOKUMENTE	LINEAGE
1706	60732	804	Bundesinventar der Landschaften und Naturdenkmäler von nationaler Bedeutung, 2. Serie, 1983	Objektblatt Berner Hochalpen und Aletsch-Bietschhorn-Gebiet (südlicher Teil)	<a href="#">Objektblatt</a>	<a href="#">Lineage</a>
3	91927	78	Biogeographische Regionen der Schweiz	Alpennordflanke	---	<a href="#">Lineage</a>

Figure 4: After clicking an object (labels) in the map or a hyperlink one of this two table rows may appear. Beside the text information other hyperlinks enable more exploration of the data. The hyperlink 'Objektblatt' opens a document with data about the specific object. The hyperlink 'Lineage' shows the process chain (lineage) of the preceding processes which directed to this process (according to figure 1).

The implementation has some major problems to solve:

- Combination and compiling of resulting spatial objects: The DNL database stores object geometries of different types of biotopes and corresponding inventories in separate SDE layers. Therefore, query result sets may consist of geometries from one or more layers which must be rendered in a combined way. The chosen approach to this problem is the creation of dynamic layers during map rendering which include the selected object geometries. The choice of this approach was also indicated since not all geometric data are provided in the ArcIMS map service and the corresponding AXL-file. Using dynamic layers increases the flexibility of the application in order to include any available geometry stored in the database.
- Filtering of attribute data: Using dynamic layers comes with a serious drawback, however. The interface should allow to select query results in order to show attribute data from the map using an identify button or by drawing a rectangle. While ArcIMS offers the possibility to filter dynamic data for display purpose, the matching attribute table is not accessible. As a workaround, this spatial filter is implemented with separate code using the ArcSDE Java API. With this API, a spatial select such as an identify action takes the coordinates, connects to the database and filters the geometric objects resulting from the initial query.
- Labeling of resulting objects: The labeling of points and lines is expanded with priority position rules. But for a polygon there is no such rule. The label will be always placed in the center of the polygon. The problem is not completely solved at the moment, since resulting polygons and corresponding labels may overlap. The implementation of a more sophisticated approach for label placement is one of the topics with a high priority.

## Conclusions

We presented an approach for an intuitive user interface for comprehensive data exploration of a database containing a diversity of different data types such as documents, images and spatial objects. The approach combines the mapping metaphor for web applications based on map server software and the search techniques provided by Internet search portals. The particular work of implementing a Google-like search is still under development and is in the state of a working prototype. ArcIMS as the map server software behind the scenes extended with custom Java beans and custom JSP tags efficiently and effectively enabled the implementation of the tools shown in this paper. A thorough test phase with users of different GIS and database skills will be the next step of our work. From our experience, we expect that the users will be able to access the database content much faster and more appropriate to their needs. Although we think that the technique proposed here is very promising, the intuitive search capabilities must be further extended towards what we call an *open* search. Current searches of the database are based on an exact match of given keywords. An *open* search will also allow retrieving information which is similar or nearly similar to typed search expressions. Offering this capability requires to benefit from the semantics of the database content. Semantics may best be expressed by ontologies which will be included in the database in the near future.

## References

Baltensweiler, A. and M. Brändli (2004): Web-based Exploration of Environmental Data and Corresponding Metadata, in particular Lineage Information. In: Scharl, A. (ed.): Environmental Online Communication. Advanced Information and Knowledge Processing Series, Springer, London, pp.127-132

Baumberger N. and Hägeli M. (2000): Using metadata in multistep preprocessing and longterm monitoring. GIS Work Session UN/ECE, Conference of European Statisticians, Working Paper No. 15, April 10-12, Neuchâtel, Switzerland. 8 pp. (<http://www.unece.org/stats/documents/2000.04.gis.htm>).

Brändli, M. (2000): A process-oriented approach for representing lineage information of spatial data. *3rd AGILE conference, Helsinki/Espoo, Finland*. <http://www.fgi.fi/agile2000>

Buyukkokten, O., J. Cho, and H. Garcia-Molina (1999): Exploiting geographical location information of web pages. Proceedings of Workshop of Web Databases (WebDB'99) held in conjunction with ACM SIGMOD'99, ACM Press, pp. 91-96.

Cartwright, W. (2004): Engineered Serendipity: Thoughts on the Design of Conglomerate GIS and Geographical New Media Artifacts. *Transactions in GIS*, 8 (1), pp. 1-12.

Ding, J., L. Gravano, and N. Shivakumar (2000): Computing Geographical Scopes of Web Resources. Proceedings of the 26<sup>th</sup> VLDB Conference, Cairo, Egypt.

International Organization for Standardization 2003. ISO 19115 - Geographic information – Metadata, International Standard ISO 19115:2003.

Jones, C. B., A. I. Abdelmoty, D. Finch, G. Fu, and S. Vaid (2004): The SPIRIT spatial search engine: architecture, ontologies and spatial indexing. In: M. J. Egenhofer, C. Freksa, and H.J. Miller (eds.): *GIScience 2004*. LNCS 3234, Springer, Berlin Heidelberg, pp. 125-139.

Lanz, A., M. Brändli und A. Baltensweiler (2005, to appear): Collection and sharing of landscape data: A large-scale, long-term view. In: Kienast, F., Ghosh, S., Wildi, O. (eds.): *A changing world: challenges for landscape research*. Springer Landscape Series.

## **Author Information**

Susette Haegi  
Swiss Federal Institute for Forest, Snow and Landscape Research (WSL)  
Section Landscape Inventories  
Züricherstrasse 111  
CH-8903 Birmensdorf  
Switzerland  
Phone: +41 44 739 26 89  
Fax: +41 44 739 22 15  
Email: [susette.haegi@wsl.ch](mailto:susette.haegi@wsl.ch)

Martin Brändli  
Swiss Federal Institute for Forest, Snow and Landscape Research (WSL)  
Section Landscape Inventories  
Züricherstrasse 111  
CH-8903 Birmensdorf  
Switzerland  
Phone: +41 44 739 23 92  
Fax: +41 44 739 22 15  
Email: [martin.braendli@wsl.ch](mailto:martin.braendli@wsl.ch)