

# Enhancing Spatial Data Infrastructures with Semantic Web Technologies

Thomas Vögele, Rainer Spittel  
Center for Computing Technologies (TZI)  
University of Bremen, Germany  
{vogele|rasp}@tzi.de

## SUMMARY

*Methods that were developed in support of applications for the Semantic Web are used to enhance information discovery services in a Spatial Data Infrastructure. A standard Catalog Service is thus enabled to resolve queries of the type “concept @ location in time” through the automatic evaluation of metadata formalizing the conceptual, spatial, and temporal semantics of geodata. A prototype of the system is implemented within the framework of the Interreg IIIB funded GeoShare project.*

**KEYWORDS:** *Semantic Web, Spatial Data Infrastructures, Catalog Service, Enhanced Query Broker*

## INTRODUCTION

The efforts of European government policies of modernization, partnership building, integrated service provision, and democratic accountability involve the usage of information systems with the ability to share information across the boundaries of organizations, as well as across regional and national boundaries. Although there is a growing number of information systems dedicated to this task, the full potential of modern communication and information technologies is not used very much. In particular this is true when it comes to utilizing geospatial data held within the public sector. Experts estimate that there is a growing market potential for the use of these data, while at the same time, adequate information about their content and availability is still missing (Greve, 2002). In order to improve the access to and the availability of geodata, Spatial Data Infrastructures (SDIs) are being created on a regional, national, and international level. Examples are the GDI-NRW (Bernard, 2002) on a regional, the evolving German geodata infrastructure GDI-DE (IMAGI, 2000) on a national, and the INSPIRE initiative [<http://inspire.jrc.it/>] on a European level.

On the other hand, a number of initiatives strive to extend the WWW into what is generally known as the “Semantic Web” (Berners-Lee, et al, 2001). The idea is to convert “machine-readable” into “machine-understandable” information by providing a well-defined meaning for the content distributed within the WWW. One approach to achieve this goal is the use of ontologies and logics to encode and reason about formal semantics (Wache, et al, 2001).

This paper outlines how ontology-based methods for semantic information integration can be used to enhance the capabilities of an SDI. The approach is tested in the GeoShare Network, an SDI that is currently developed on a regional and transnational level within the framework of the GeoShare project (Vögele, T., et al, 2003). GeoShare is a project co-funded by the EU Interreg IIIB North Sea Region Programme. The overall goal of GeoShare is to support regional and transnational activities in the thematic areas of spatial management/environmental protection, employment/education, and tourism/cultural heritage, mainly through a better and more efficient use of geospatial (i.e., digital maps, satellite images, etc.) and geo-referenced (i.e., text documents, images etc. with an explicit reference to geographic space) data.

The GeoShare Network will provide the technical infrastructure to improve the exchange and use of geodata within the participating European regions in Norway, Scotland, the Netherlands and Germany, as

well as within the whole North Sea Region. The Network is built to be compatible with the SDIs mentioned above in that it uses a state-of-the-art architecture based on standards and methods developed by the OpenGIS Consortium (OGC) (<http://www.opengis.org>).

This paper will give a brief overview of the GeoShare Network and its basic components. Special emphasis is given to the Catalog Service, one of the central services in any SDI. In the GeoShare Network, this service is extended by the introduction of Semantic Web technologies for the enhanced management and retrieval of data and services.

## **THE GEOSHARE APPROACH**

### **Spatial Data Infrastructures**

Geospatial data as well as geo-referenced data are available in most public and private organizations. To exchange such data between organizations, an adequate technical and organizational framework is necessary. Such a framework can be a Spatial Data Infrastructure (SDI) which integrates (geo)data and (geo)services into one distributed, but interoperable framework.

The architectural paradigm of most Spatial Data Infrastructures (SDIs) is to provide interoperable web-services to manage, retrieve and process geodata. To ensure the interoperability between the various services integrated in such a framework, the OpenGIS Consortium (OGC) defined a number standard interface and software specifications. These specifications play an important role in the design and implementation of spatial data infrastructures. Since the foundation of the OGC initiative in 1994, most of the technical and conceptual challenges with regard to interoperable geospatial data and applications have been addressed. Based on a general conceptual approach (e.g. OGC, 1999), the OGC has issued implementation specifications for some of the most important technical topics (see <http://www.opengis.org>), while a number of related topics are still in the stage of discussion papers.

The feasibility of the technical concepts underlying such services-based infrastructures have been tested and proven by a number of projects and initiatives (e.g., the implementation test-bed developed within the GDI-NRW (Bernard, 2002)). At the same time, OGC-conformant software is implemented by a growing number of GIS software providers, both in the commercial and the open source sector. As of November 2003, the OGC web pages (<http://www.opengis.org>) list about 200 products registered as “OGC-conformant” or “implementing OGC specifications and interfaces”.

The value of a spatial data infrastructure depends on the services that it is able to offer. Catalog Services are a key component of almost every SDI. Based on the respective metadata descriptions, these services provide detailed information about the data and services that are integrated in the infrastructure. From a user’s perspective, a Catalog Service provides a tool to discover which data sources and services are available within an SDI, and which of those data and services are best suited for a given task.

### **The GeoShare Network**

Following the general SDI paradigm, the *GeoShare Network* implements a set of distributed, web-based geoservices. Within the network, each service is accessible through a well-defined interface. Whenever possible, the implemented interfaces are compatible with the respective international standards and specifications issued by the OpenGIS Consortium (OGC) and the ISO. The overall technical design of the GeoShare Network is therefore similar to designs that are used by the regional, national and international geodata infrastructures described above. Because it is compatible with these initiatives, it will be able to eventually integrate the *GeoShare Network* into the emerging European SDIs.

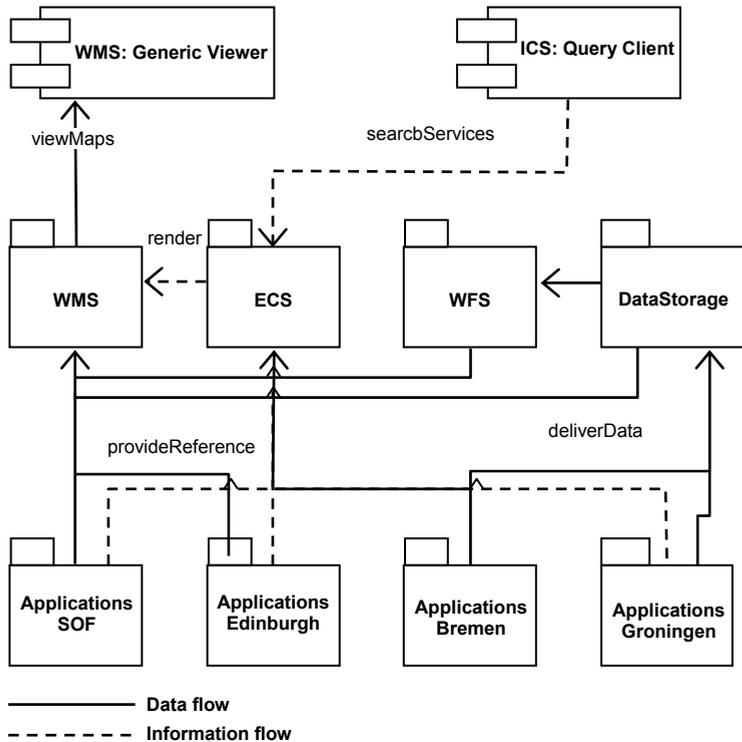


Figure 3: The GeoShare Network (schematic)

Within the GeoShare Network, we distinguish between *basic* and *advanced* services. A cluster of basic services forms the backbone of the *GeoShare Network* and is hosted on a dedicated hardware platform by one of the project partners in Bremen (the TZI, Universität Bremen) (Figure 1). This service cluster consists of:

- A storage service that allows the project partners to store geodata in a number of databases and data formats (the *GeoShare DataStorage*),
- an online data catalog that provides search functionality and facilitates the access to the applications, services and data in the Network (the *GeoShare Enhanced Catalog Service (ECS)*),
- a tool to visualize digital maps individually or in an integrated layered view (the *GeoShare Generic Viewer* in combination with cascading Web Map Services), and
- a service to provide full access to geodata stored in the *GeoShare DataStorage* or other data stores (the *GeoShare Web Feature Service (WFS)*).

## ENHANCED CATALOG SERVICES (ECS)

### OGC Catalog Services

As stated in the general principle of the European INSPIRE initiative, an SDI should make it “easy to discover which geographic information is available, fits the needs for a particular use and under which conditions it can be acquired and used” ([http://inspire.jrc.it/principles\\_en.html](http://inspire.jrc.it/principles_en.html)). For this reason, metadata

descriptions of data sources and services, as well as appropriate catalog services to manage these metadata play a key role in any SDI. They provide the user of an SDI with the functionality that is needed to discover, among a potentially large number of data sources and services, the ones that are best suited to fulfill a specific task. Besides expressive metadata descriptions, powerful search methods are the basis for an optimized information discovery process.

The OGC Catalog Service Specification (OGC, 2002) describes a catalog service as an integration of a discovery, an access, and a management service. While the catalog service software used in the GeoShare Network implements all three services, we will focus in this paper on the discovery service only.<sup>11</sup>

### **Complex Information Requests**

A standard OGC compatible Catalog Service retrieves information by syntax matching, i.e. the search terms specified in the information request are matched against the metadata descriptions of the resources registered with the Catalog Service. Typically, a resource is considered to be thematically relevant if any of the specified search terms match the topic keywords in the metadata. Likewise, if the information request specifies a geographic location (in terms of point coordinates or a bounding box), the match between this location and the resources spatial footprint given in the metadata is used to determine spatial relevance.

A more sophisticated search tool should be able to go beyond simple string matching and bounding-box reasoning. It should not only find resources that exactly match the specified query, but also “near hits”, i.e. resources that only approximately match a query. One way to do this is to retain the semantics of terms and spatial relations in formalized ontologies, and to apply automatic reasoning processes to these formalizations. This approach has been used in BUSTER, an information broker middleware developed at the Center for Computing Technologies (TZI) in Bremen (Visser, 1999; Hübner, 2003; Vögele and Schlieder, 2003; Visser et al, 2002). BUSTER uses thematic, spatial and temporal ontologies and respective reasoning engines to resolve requests of the type *concept @ location in time*.

### **The GeoShare Enhanced Catalog Service (ECS)**

To be able to resolve complex information requests, the GeoShare Network integrates two components, a standard OGC-compliant catalog service (developed by Delphi IMM, a Potsdam-based geo-consulting company) and BUSTER as a tool to specify complex, knowledge-based queries. Together they form the GeoShare Enhanced Catalog Service (ECS) (Figure 4). For integration into the ECS, the BUSTER system was transformed from a client-server application into a web service. In the first prototypical implementation, the interface to this web service is only partially OGC compliant: While the ECS Query Broker is able to send OGC compliant information requests to other services (i.e., to standard Catalog Services), the information needed to build complex information requests has to be delivered through an extended (i.e., non-standard) OGC interface (Figure 4). This limitation is necessary because for the specification of complex information requests, the ECS needs a set of parameters that are not (yet) supported by the OGC catalog services specification.

---

<sup>11</sup> As of fall 2003, the *OGC Catalog Service Specification* was not available in a finalized version. The catalog service currently used in the GeoShare Network follows an earlier draft version of this specification, but will be modified accordingly as soon as the final specification becomes available.

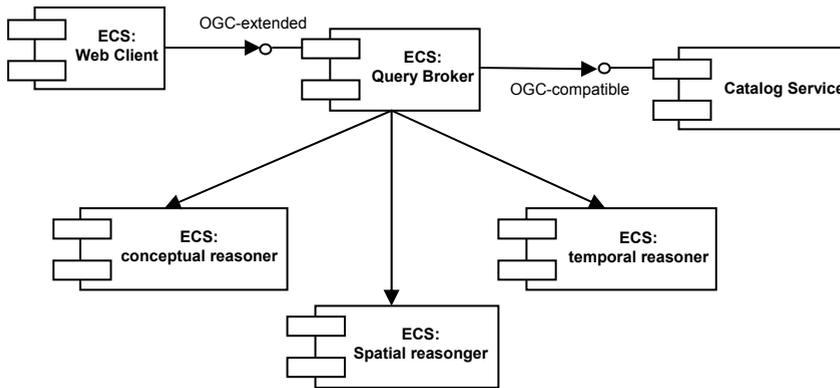


Figure 4: The Enhanced Catalog Service (ECS)

The first step to specify an information request through the ECS Web Client is to select the thematic, spatial, and temporal domains of interest (e.g., the domains “Spatial Planning”, “The North-Sea Region”, and “Events since 1995”). This choice results in a listing of all domain-specific application ontologies in the thematic, spatial, and temporal search dimensions that are registered with the system. Within each search dimension, all application ontologies have in common that they are based on the same vocabulary which is given through a domain ontology. Based on this common vocabulary, the system performs an automatic on-the-fly ontology integration.

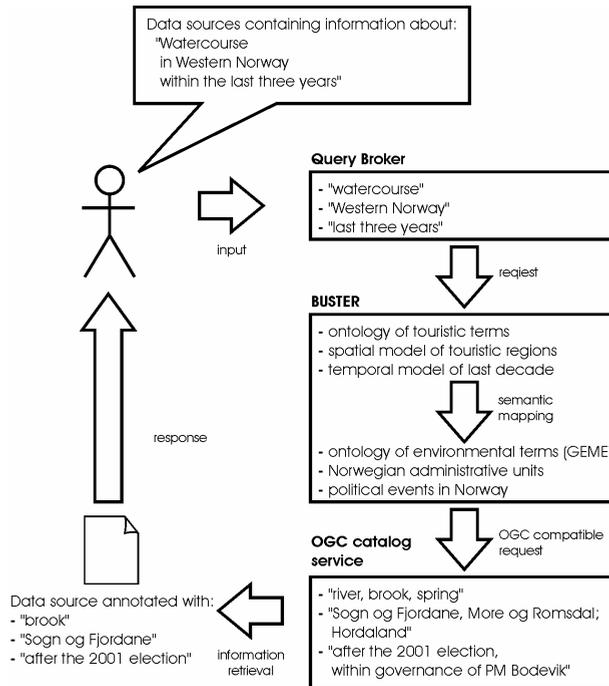


Figure 5: Expansion of an information request in the ECS

From the application ontology that is best suited for a specific information request, the user may select suitable search terms for each search dimension. The ECS expands these terms by selecting all sub-

concepts (for the thematic search term), spatially related place names (for the spatial search term), and relevant time periods (for the temporal search term) that can be found in the integrated application ontology.

For example, through the ECS user interface, a request like “*Watercourse in Western Norway within the last 3 years*” can be specified. For each part of the request, terms from specialized application ontologies are selected, i.e., an ontology of touristic terms for the conceptual part, a spatial model of touristic areas for the spatial part, and a generic temporal model for the temporal part. The ECS conceptual, spatial, and temporal reasoners would then expand this request into a query like:

```
"(themes.keywords LIKE '%river%' OR themes.keywords
LIKE '%brook%' OR themes.keywords LIKE '%spring%')
AND (spatials.keywords LIKE '%sogn of fjordane%' OR
spatials.keywords LIKE '%more og romsdal%' OR spatials.keywords
LIKE '%Hordaland%')
AND (temporals.keywords LIKE '%after the 2001 election%' OR
temporals.keywords LIKE '%within the governance of PM Bodevik%')"
```

To do so, the system would integrate additional application ontologies, like the GEMET environmental thesaurus (Nax and Lethen 1999) (for the conceptual part), a spatial model of the administrative units of Norway, and a temporal model of major political events in Norway.

In the ECS Query Broker, the extended query is transformed into an information request that is compatible with the OGC specifications for catalog services. This extended information request is forwarded to a catalog service (in this case, the ECS catalog service). In the catalog service, the query is matched against the resources metadata and the result of this matching process is returned as a metadata description of all matching resources (Figure 5). In the Query Broker, the result set is ranked according to thematic, spatial and temporal relevance before being forwarded to the search client. Through the search client, the user may then select, visualize and retrieve the most relevant data sources (e.g., map layers). The visualization of the query results can be done with the help of the *GeoShare Web Map Service* and the *GeoShare Generic Web Map Client*. Using cascading web map services, these tools allow to simultaneously access multiple digital maps on different map servers and display them in one integrated view.

## Summary and Conclusions

The enhanced information discovery technologies described above allow to retrieve geospatial and geo-reference data through an assessment of their relevance with respect to thematic, spatial and temporal search terms. This extends information retrieval provided by standard catalog services from a purely syntactic to a more semantic level. The next step will be the implementation of an enhanced catalog service (ECS) as a fully integrated component of an OGC service chain. This will allow to enhance any standard OGC catalog service with semantic web technologies without having to modify the respective service interface.

## BIBLIOGRAPHY

- Bernard, L., Experiences from an implementation testbed to set up a national SDI. in 5th AGILE Conference on Geographic Information Science. 2002. Palma de Mallorca.
- Berners-Lee, T., J. Hendler, and O. Lassila, The Semantic Web. Scientific American.com, 2001.
- Greve, K. and L. Plümmer, Abschlusskommunique des D21-Kongreß Geoinformationswirtschaft 2002, Bonn. 2002.
- Hübner, S., Qualitative Abstraktion von Zeit für Annotation und Retrieval im Semantic Web, masters thesis, Dept. of Computer Science (FB 3). 2003, Bremen University: Bremen.
- IMAGI, Konzeption eines effizienten Geodatenmanagements des Bundes. 2000, Interministerieller Ausschuss für Geoinformationswesen (IMAGI).
- Nax, K. and J. Lethen, ETC/CDS General Multilingual Environmental Thesaurus (GEMET) - General Information on GEMET. 1999, ETC/CDS.
- OGC, The OpenGIS Abstract Specification. 1999a, Open GIS Consortium.
- OGC, OpenGIS® Catalog Services Implementation Specification, Rev: 1.1.1. 2002c, OpenGIS Consortium.
- Visser, U., T. Vögele, and C. Schlieder, Spatio-Terminological Information Retrieval using the BUSTER System. in In: Proceedings of Environmental Informatics 2002. 2002. Vienna, Austria.
- Visser, U., H. Stuckenschmidt, G. Schuster, and T. Vögele, Ontologies for Geographic Information Processing. Computers & Geosciences, 2002. 28: p.103-117.
- Vögele, T. and C. Schlieder, Spatially-Aware Information Retrieval with Place Names. in In: Proceedings of the FLAIRS 2003. 2003: AAAI Press.
- Vögele, T., R. Spittel, U. Visser, and S. Hübner, GeoShare - Building a Transnational Geodata Infrastructure for the North Sea Region. in Geodaten und Geodienste Infrastrukturen - von der Forschung zur praktischen Anwendung. 2003. Münster: IfGI prints.
- Wache, H., T. Vögele, U. Visser, H. Stuckenschmidt, G. Schuster, H. Neumann, and S. Hübner, Ontology-Based Integration of Information - A Survey of Existing Approaches. in IJCAI 2001 Workshop on Ontologies and Information Sharing. 2001. Seattle, WA.