Semantic Interoperability by means of Geoservices

Semantic problems in three use cases and approaches for potential solutions

L. Bernard[2], S. Haubrock[1], S. Hübner[3], R. Lessing[1], M.Lutz[2], U.Visser[3]

[1]Delphi InformationsMusterManagement (DELPHI IMM), Potsdam
{soeren.haubrock|rolf.lessing}@delphi-imm.de

[2]Institute for Geoinformatics (IfGI), Münster
{bernard|lutzm}@ifgi.uni-muenster.de

[3]Center for Computing Technologies (TZI), Bremen
{huebner|visser}@tzi.de

Introduction

This paper describes the main goals and the first results of the research project *Semantic Interoperability by means of Geoservices* (*meanInGS*). The project started in October 2002 as part of the research and development programme *Geotechnologies*. The three partners working on this project are Delphi InformationsMusterManagement (DELPHI IMM) Potsdam, Center for Computing Technologies (TZI) Bremen and Institute for Geoinformatics (IfGI) Münster.

The Problem – Schematic and Semantic Heterogeneity

The results of geoscientific research projects supply large, valuable datasets and powerful tools for the accomplishment of existing research tasks. A continuing use of these results, especially by institutions that were not involved in the original projects, often proves to be difficult. In order to overcome these problems syntactic, schematic and semantic interoperability have to be achieved.

The problem of syntactic heterogeneity emerged as a result of mostly native data formats and the development of monolithic or proprietary systems. The World Wide Web (WWW) supplies the basic infrastructure for the distributed use and multiple exploitation of data and systems (systems interoperability), while approved geoinformation technology standards developed by the OpenGIS-Consortium (OGC) and the International Organisation for Standardisation (ISO) provide the essential basis for syntactic interoperability and cataloging of geoservices and –data. Aborning geodata infrastructures like GDI-NRW (Kuhn et al. 2001) give examples of what can be accomplished by this approach and point out which challenging interoperability issues (e.g. GI service chaining) remain.

Although the basis for syntactic interoperability exists in many cases the usability of information that results from geoscientific research projects for institutions from different information communities (ICs)[1] will remain limited, because they are confronted with schematic and semantic heterogeneity.

These kinds of heterogeneity are a basic characteristic of all information sources. The use of individual parameters during the process of data collection, modifications and complements in the nomenclature, other or improved data collection methods and last but not least another sense (another viewpoint) of the 'world' are causes for this heterogeneity.

In order to overcome the problems resulting from schematic and semantic heterogenities in typical geodata infrastructures, the *meanInGS* project will develop and implement a concept to use existing and newly created technologies. Building on an existing data infrastructure intelligent geoservices will be designed and developed for achieving schematic and semantic interoperability in the geodata processing. A great importance is attached to the practical usability of the results. The project aims at the following advantages for geoscientific aspects:

- exchange of results between different geoscientific projects,
- utilisation of legacy geoscientific databases,
- utilisation of the databases for projects outside of the geoscientific context,
- avoidance of not updatable secondary data repositories.

Starting Situation

The basis for the developments within the project is a model for geodata infrastuctures that currently emerges in the national and international context within the framework of OpenGIS and ISO. First implementations exist in the form of so called “geoportals” and catalog services. In the following illustration the state of the art model is described shortly.

[1] „An Information Community is a collection of people (a government agency or group of agencies, a profession, a group of researchers in the same discipline, corporate partners cooperating on a project, etc.) who, at least part of the time, share a common digital geographic information language and share common spatial feature definitions. This implies a common world view as well as common abstractions, feature representations, and metadata.“ (OGC 1999)

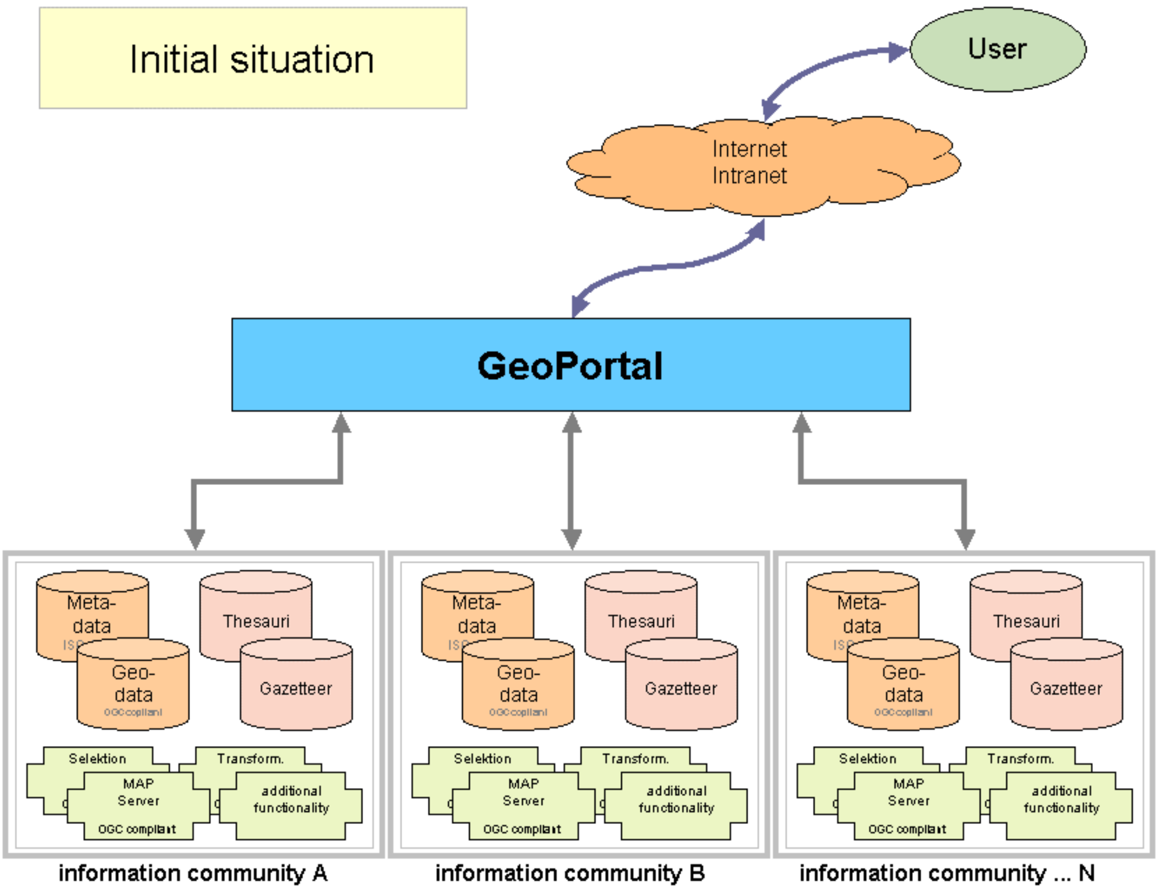

Figure 1: Information flow in current implementations of a geoportal

A number of ICs has gathered sets of basic and domain-specific data for their own use. Furthermore, each IC has implemented a system to offer their data to other ICs using multiple components (see Figure 1).

Apart from the data itself, a metadata catalogue is necessary as well as a thesaurus and a gazetteer. In order to enable users to access the data from "outside" the system, the functionalities of selecting, transforming and map-serving have to be realised.

One major problem is the fact, that not all ICs do have all of these components. If they do so, the entries in the metadata catalogue will often not be matchable among different ICs.

Offering metadata and data as described in figure 1 is already possible. However, in order to link multiple implementations to each other, intensive coordination efforts are necessary. In order to make it possible to compare and evaluate results from different ICs, some major semantic problems have to be solved.

Structure

In the following section the aims of the three project partners, Delphi InformationsMusterManagement (DELPHI IMM) Potsdam, Institute for Geoinformatics (IfGI) Münster and Center for Computing Technologies (TZI) Bremen will be described. The main part of the paper will then outline three use cases that have been developed. These will serve as a source for identifying and classifying practical problems that are caused by schematic and semantic heterogeneity and provide the framework for the development and implementation of methods and technologies to overcome these problems.

Aims of the Project Partners

Delphi IMM

It is the aim of Delphi IMM to extend their technology MSPIN (Software Tools for the **m**ediation of **sp**atial **in**formation) with semantic functionality for facilitating the searching in catalogues and the rendering of geodata. IMM focuses on the procedures of mapping geodata as well as the definition of user-specific viewpoints. The latter is a special goal of IMM. By implementing this, a customer could phrase his queries for geodata so that data providers can carry out a direct mapping of their data to that query.

IMM emphasizes on the integration of remotely sensed data into a service chain. The classification of this data has to be implemented as an automatic service.

Through an iterative approach in the following use cases issues of semantic interoperability are addressed step by step. The use cases described below serve as a guideline with respect to the investigation of the available data sources and the query formulations. An intensive information exchange with the two project partners is intended.

IfGI

The goal of the Institute for Geoinformatics in the *meanInGS* project is to specify services for semantic translation and to test them within geodata infrastructures. The methods and techniques will be developed and tested in the context of well defined use cases from the domain of geosciences. This will ensure that the application will be pragmatic and that the results will be useful for the domain of geosciences.

The first aim of the ifgi project work is to identify, analyse and classify problems in the use cases caused by semantic heterogeneity. These problems will form the basis for the development of services for semantic translation between catalogues, and between user requirements (or questions) and the services registered in these catalogues. This will require the extension of existing metadata models and catalogue services to include information on services (rather than only on data). This extension will provide the information that is necessary for an algebraic model of the semantics of user data. Finally, for testing the functionality of existing Web Map Services (WMS), Web Feature Services (WFS) and Web Catalogue Services (WCS) might have to be extended.

TZI

The goal of the Center for Computing Technologies (TZI) is to apply methods and technologies that were gathered in a number of recent research- and PhD projects to a practical use-case that is rooted in a functional geodata infrastructure. Thematically, these methods and technologies are focussed on intelligent information retrieval and semantic data integration of geospatial data. Technically, they are centered around logic reasoning based on qualitative conceptual, spatial and temporal models.

Within the meanings project, the TZI will analyse the chosen use-case regarding the need of semantic data integration and intelligent information retrieval. The methods and tools mentioned above will then be customised to fit within the chosen geodata infrastructure to fulfil the relevant tasks identified in the use case. This includes the implementation of conceptual, spatial and temporal ontologies specific to the use case. All components will be integrated in the geodata infrastructure, tested and, if necessary, modified and improved to meet the requirements of the use case.

Use Case Descriptions

Use Case I – Detecting hazard areas during flooding events

This use case is about visualising and simulating the water levels in a river catchment as well as detecting and visualising potential hazard areas due to flooding. Information systems in the context of civil protection have to use comprehensive and up-to-date data sets covering the whole catchment area. Instead, current applications in this field are at most very specific and static approaches. Furthermore, these systems are established after the incident has happened due to the problems in quickly gathering the appropriate data. Remote sensing components have not been integrated in such systems in an operational manner so far.

A fundamental difference between this approach and established ways is the implementation of the following aspects:

- integration of heterogeneous data sets from different information communities,
- use of nearly real-time data, derived from
  - quasi-continuous measurements and
  - remotely sensed data,
- integration of services processing remotely sensed data,
- dynamic selection and coupling of multiple models for simulation and evaluation.

The aim of this use case is to evaluate how a service can be used as an interface for gathering spatially distributed and heterogeneous data and providing this data to a user-selected modeler. However, it is not the goal of the following scenarios to create a complex model that could lead to realistic forecasts in its simulations.

In the context of this project it is rather important to work on the integration of heterogeneous data sets, to implement new automated services, to establish a generic service chain and to overcome further semantic problems, e.g. the issue of mapping between different nomenclatures. Hereby, the results of other research projects can be integrated. An intensive information exchange between the partners is intended.

In the following scenarios, an infrastructure will be proposed for the catchment of the Elbe river. This stream covers multiple regions resulting in different competences for the data. Currently, a great effort is put into several research projects targeting an integrated information system on the Elbe river (Bundesanstalt für Gewässerkunde 2000). Due to the severe flooding events back in summer 2002, the benefit of an effective and fast-working information system became clear (Ministerium des Innern des Landes Brandenburg 2001).

Depending on the stage in the implementation process two different user groups exist. In the first and second scenario, a basic description of the current situation in the river catchment area has to be visualised. Apart from the general public using the technology of the Internet, the potential users include experts in the field of river management. These experts have a certain interest in rapidly gathering the most up-to-date datasets to get an overview of the current situation in the catchment.

In the third and fourth scenario, generic components are used in the service chains. This enables scientists to apply their own model with the most up-to-date and realistic datasets. Furthermore, different models can be compared to each other by dynamically choosing the appropriate model at runtime.

In the following the single components and interactions of a dynamic approach are sketched in four scenarios. Starting with a rather static approach, the generic aspect is going to increase in the second and especially in the third and fourth scenarios, leading to new semantic difficulties.

Scenario 1: Visualising the hot spots

In this first step, the objective is to visualise the water levels and potential hazard areas due to flooding in the catchment. The necessary geodata comprise the river network itself (as a GML feature collection (OGC 2002b)) and the current water level measurements (as spatially anchored datasets or features). The major task in this approach is to collect the distributed datasets and assign them to certain segments in the river network (spatial join).

The hydrological model in this scenario is static. Thus, only the spatial extent selected by the user and the current time are dynamic parameters in the process chain. To receive the user requirements an input form is provided by the client service (step one). The user can select a certain area of interest and choose between different periods of time (e.g. current situation or several past situations). With this approach the syntactic frame of the request can be standardised. Since the user can only choose between given options, no semantic problems occur when the query builder compiles the OGC-conformant query files (step 4). Only one specific hydrological model is used, so the datasets needed for its simulations can be statically implemented in the system.

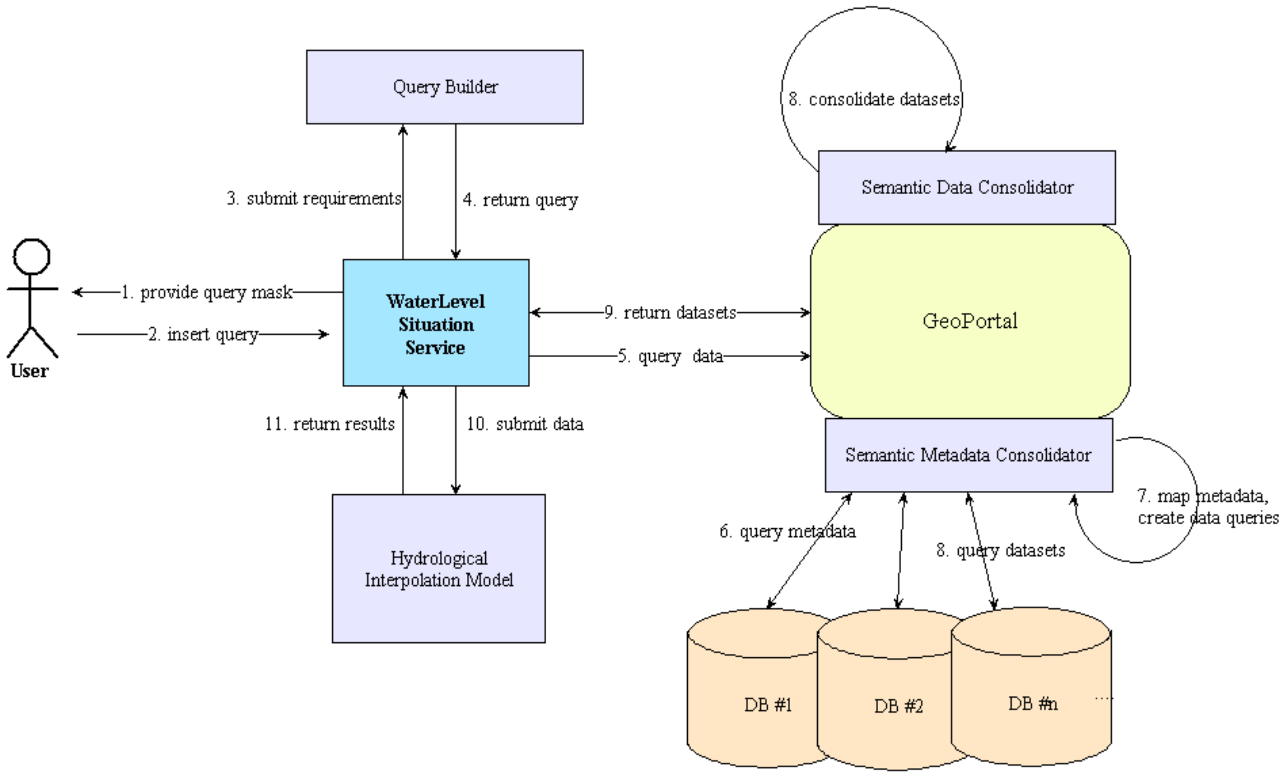

Figure 2: programme Flow of information in the first scenario

In step five the query is passed to the geoportal which serves as a hub to the single information communities providing their metadata and data repositories. Due to the fact that several institutions are responsible for the datasets of a large river network, they must be compiled from several databases and synchronised with each other. In this scenario the query for up-to-date water levels must be expanded to the public authorities of at least four federal states and at the national level. Only the union of these datasets provides an optimum coverage of the whole area.

It can not be expected that the datasets are compatible with each other: in two catalogs some attributes can describe the same content while being labeled differently and vice versa. Therefore, the syntactically conformant query has to be mapped to the metadata of the catalogs. This mapping can take place on the client side (i.e. geoportal) or on each of the server sides (metadata catalog system). The technology to perform the mapping has to be developed in the scope of this project.

In the next step, the different databases can be queried using the specific metadata descriptions that have been mapped to the base query in the step before.

Once the data has been collected in the geoportal, another service (the semantic data consolidator) will have to compile it to a single homogenous dataset. Especially the mapping of relevant attributes to each other (in this case the measurements) is an issue to be solved. Problems can occur in the case that the metric units are different from each other or the temporal context of the measurements is not specified equally. Some institutions might hold data about the same gauge, so duplicates have to be filtered out.

In the next step, the modeling service has to inter- and extrapolate the discrete water level measurements over the whole stream section which has to be evaluated by the hydrological model. Finally, the resulting estimation will be composed to a map indicating the user which areas are potentially at high risk. The assessment whether or not a river segment is at high risk will be based on a very simple reasoning provided by the model in this scenario.

The scenario described above is rather static, but already faces some semantic problems. In the next scenario, this concept is extended by another service leading to new aspects of semantic interoperability to be solved.

Scenario 2: Integration of remotely sensed data

For estimating the situation in a the river catchment more realistic, the datasets received by some tens of discrete gauge measurements unequally distributed along the river are insufficient. In order to get the “overall picture”, remote sensing data can make a crucial contribution.

The methods used in remote sensing provide a very fast and objective assessment of the situation in a large area of interest (in this case in the whole river catchment). Thus, the integration of remotely sensed data can show the flooded areas as well. On the other hand the measurement data is still necessary since remotely sensed data only provide a view in two dimensions (in the case of optical sensors), i.e. the areas covered by water can be extracted, but not the water level.

For this reason, a new service has to be established in the chain. The raw image data cannot help the normal user to find the regions that have been flooded. To interpret the data correctly classification experience is required. Therefore a classification service has to fill the blank between data collection and interpretation of the data by the user. This classification is intended to be fully automatic in order to provide a service at runtime. As a result, the classified flooded areas can be visualised together with the water levels in a map.

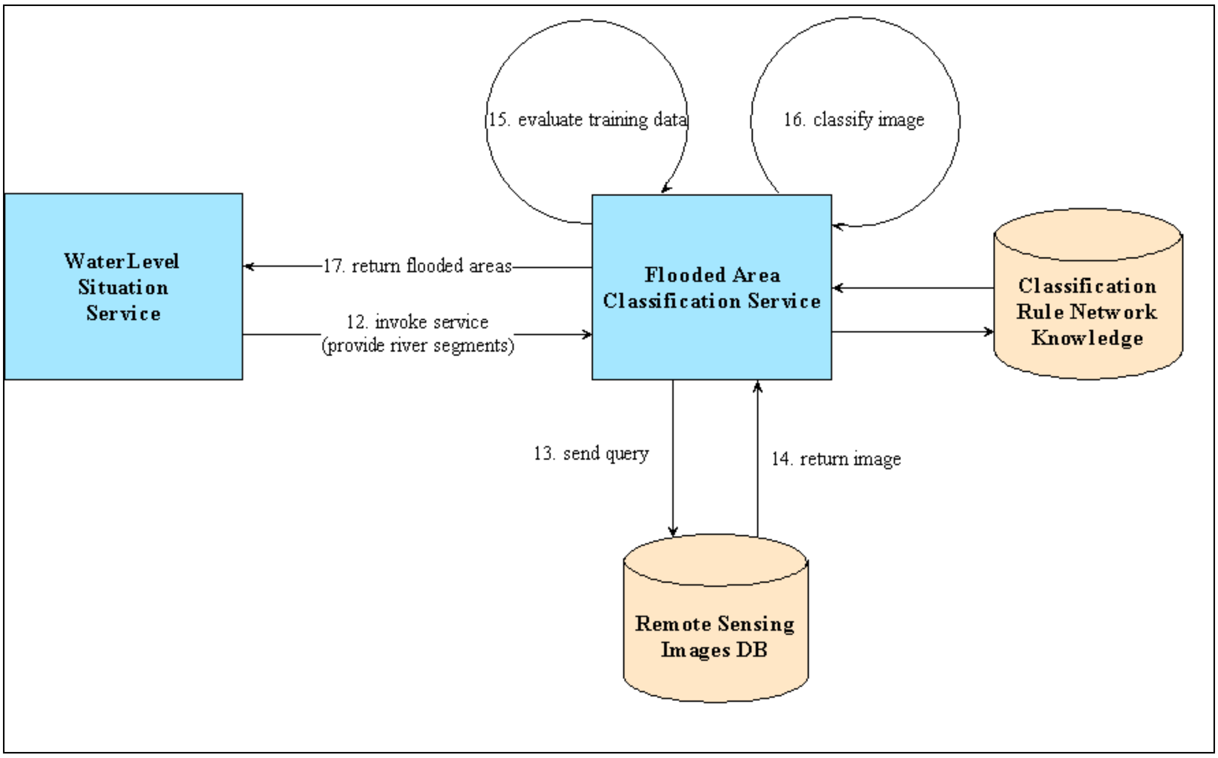

Figure 3: Integrating the classification service

After the steps 1 to 11 from the first scenario have been passed through, the extracted river segments can be used as input for the classification service. In step 13 the image data is requested for the area of interest. Those parts of the image that are overlapping the river segments can then be used as training data for calibrating the classification rule network. Additional information on how to construct the classification rules can be derived from a knowledge base. In a last step, the classification itself takes place and creates a feature collection of the areas which are flooded (see figure 3).

Scenario 3: Dynamic modeling and prediction

In this scenario the hydrological model is exchangeable. The user is intended to select the model of choice from a list. More sophisticated models need many data inputs, especially if they aim to make forecasts. The information about the datasets needed for the simulations is therefore depending on the model which will be used and has to be acquired at run-time. The hydrological model needs to compose formalised requirements in terms of the data needed for the simulation. The syntax of these compilations can be standardised using a certain XML-scheme, but the meaning of the attributes needed remains vague. They have to be mapped to the available datasets in the geoportal in the next step.

In this scenario a simple model has to be built which uses additional datasets, e.g. soil types with different transmissibility rates for estimating the runoff in an area. These additional datasets lead to more semantic problems: either the classification of derived attributes is already existing in the different datasets and has to be mapped to each other, or it has to be classified by another model using empirical data from a knowledge repository.

By choosing the model at runtime, the service chain has to be established dynamically. The appropriate workflow is similar to the one described in the road blockage estimation use case (“Ad-hoc Service Chaining”, see below).

Scenario 4: Coupling multiple models

The latter scenarios comprise the hydrological modeling of the river network. In the next step the advantages of a standardised generic service chain structure are exploited by adding a further modeling component, e.g. a system for the estimation of chemical exposure rates in the aquatic environment.

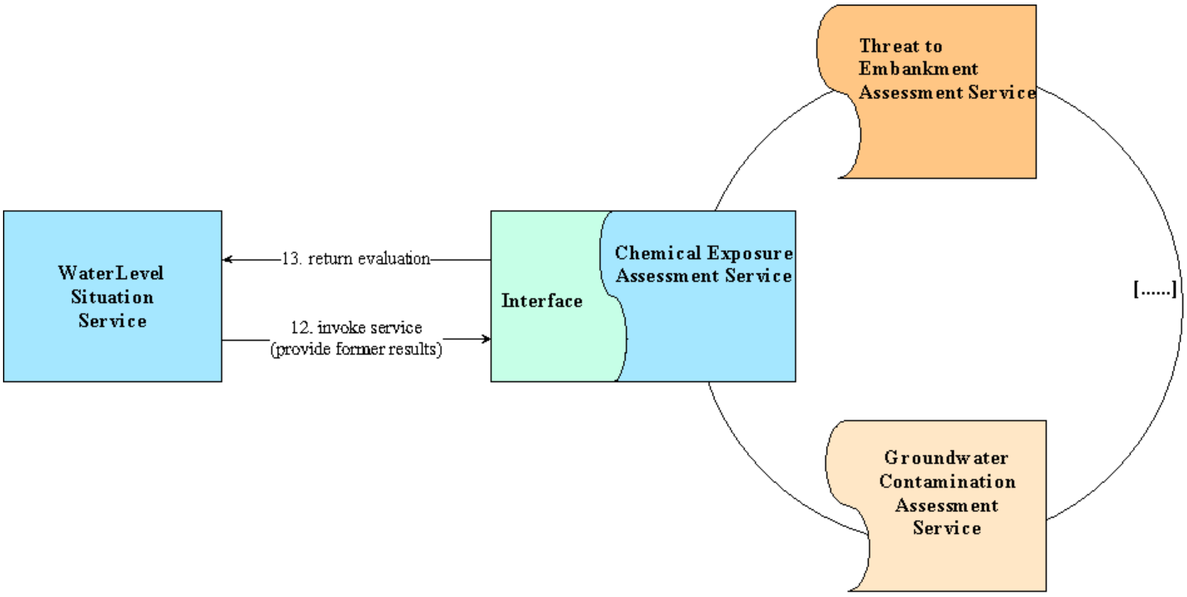

Figure 4: Workflow with pluggable proceeding models

For this scenario an existing model will be used and adapted to the required interface format.

One option is to use and adapt the *Geography Referenced Regional Exposure Assessment Tool for European Rivers (GREAT-ER)* (Matthies et al. 2001).

Problems caused by semantic heterogeneity

The data to be processed in this use case originate from multiple information communities resulting in heterogenous formats, different spatial and temporal resolutions as well as other types of semantic heterogeneity (e.g. naming conflicts). Solving the issues that are connected with these differences is the major task of this project.

Thus, the focus is on semantic metadata interpretation (e.g. do the data sets representing water level measurements belong to the same variable?), the semantic data interpretation (mapping between different nomenclatures, e.g. porosity of soil or vegetation classes) and data fusion (pre-processing the data for further utilisation). Furthermore, the integration of remotely sensed data by implementing a new service for automatic classification is associated with tremendous semantic problems.

In order to establish a flexible, generic system for time-critical disaster management some semantic problems occur and have to be solved. Additional services are required and have to be implemented. One new service is the automatic extraction of the flooded areas out of the raw image data.

Establishing dynamic service chains as well as the integration of data from several different information communities brings up new mapping problems that are intended to be solved in the scope of this project.

Use Case II – Estimating Road Blockage after Storms

The use case is set in the area of disaster management and mitigation. Storm events may cause severe road blockage by windfall timber, especially on roads that run through forest areas. In December 1999 the winter storm “Lothar” caused an accumulation of about 30 million solid cubic metres in windfall timber in Baden-Wuerttemberg. For a few days some villages were cut off from the rest of the world because of road blockage and it remained extremely dangerous to stay in the affected areas because of the devastated state of the forest (Gemeindetag Baden-Württemberg 2000).

The main actor in the use case is a person in the agency responsible for ensuring road safety. In Baden-Wuerttemberg this is the Forestry Directorate. After a heavy storm they coordinate the assignment of the Governmental Disaster Relief Organisation (Technisches Hilfswerk, THW) and of the Federal Armed Forces in the disaster area. They have to keep track of where and how much help is needed by the local authorities to clear the road blockages as quickly as possible. In order to coordinate the clearing operations effectively a rough estimation of the roads that are most likely to be affected by fallen timber is required.

GIS support

The estimation of roads likely to be affected by the storm can be supported through GIS analysis. The analysis is based on the assumption that those forest stands with overaged trees be most strongly affected by storms. At what age a tree can be called overaged depends on its species.

At first a selection of forest areas that are potentially at risk from windthrow based on the information of species and age is made. Secondly the result is intersected with road data in order to identify the roads that run through stands with overaged trees. In order to perform this analysis the user needs to have access to data of the local street network as well as data of forests that contain detailed information about tree species and age. If this information is not available, the analysis can not be accomplished.

GI web service support

The application of GI web services providing and displaying the information could help to achieve a more flexible flow of information instead. GI web services can provide access to up-to-date data as well as the flexibility to extract the required information out of several data resources by combining them.

In the next section the state-of-the-art scenario is described where generic Web Feature Services (WFS) and Web Map Services supporting Styled Layer Descriptor (WMS/SLD) are statically chained to form a complex service for estimating road blockage after storms. This scenario was realized in the just finished GDI NRW Testbed II, the second testbed of the Geodata Infrastructure Initiative of North Rhine-Westphalia (Bernard 2002).

Starting from this scenario, two future scenarios are developed describing how (1) remote sensing data could be incorporated to gain additional information and (2) a more flexible way of service chaining can be achieved by approaching the problem of semantic interoperability.

Scenario 1: Chaining Distributed Web Feature and Web Map Services

In the context of the GDI NRW Testbed II a "road blockage" service for the use case described above was implemented. This section describes the existing implementation and sketches how it could be extended by including remote sensing data.

Scenario 1.1: Current implementation

The “road blockage service” (*http://xtra.interactive-instruments.de/demo/demo-wfs.html*) lets the user select a lower age limit for trees, a certain tree species on which to constrain the query and a maximum number of forest features to be returned. The service returns a map of the road network against a topographic map with those roads highlighted for which a blockage is most likely.

According to (ISO/TC-211 & OGC 2002) the "road blockage" service represents an aggregate service (opaque chaining). In a black box manner it accesses a number of OGC-conformant WFS (OGC 2002b) and WMS/SLD (OGC 2002c;2002a) provided by different members of the GDI NRW:

- a WFS hosted by the Institute for Geoinformatics in Münster serving forest features from a database provided by the North Rhine-Westphalian Department for Forestry,
- a WFS hosted by interactive instruments (http://www.interactive-instruments.de/) serving road features from a database containing the North Rhine-Westphalian road network,
- a WMS hosted by the North Rhine-Westphalian agency for data processing and statistics (LDS, http://www.lds.nrw.de/) providing topographic maps, and
- a WMS hosted by interactive instruments for displaying the above-mentioned features in a map.

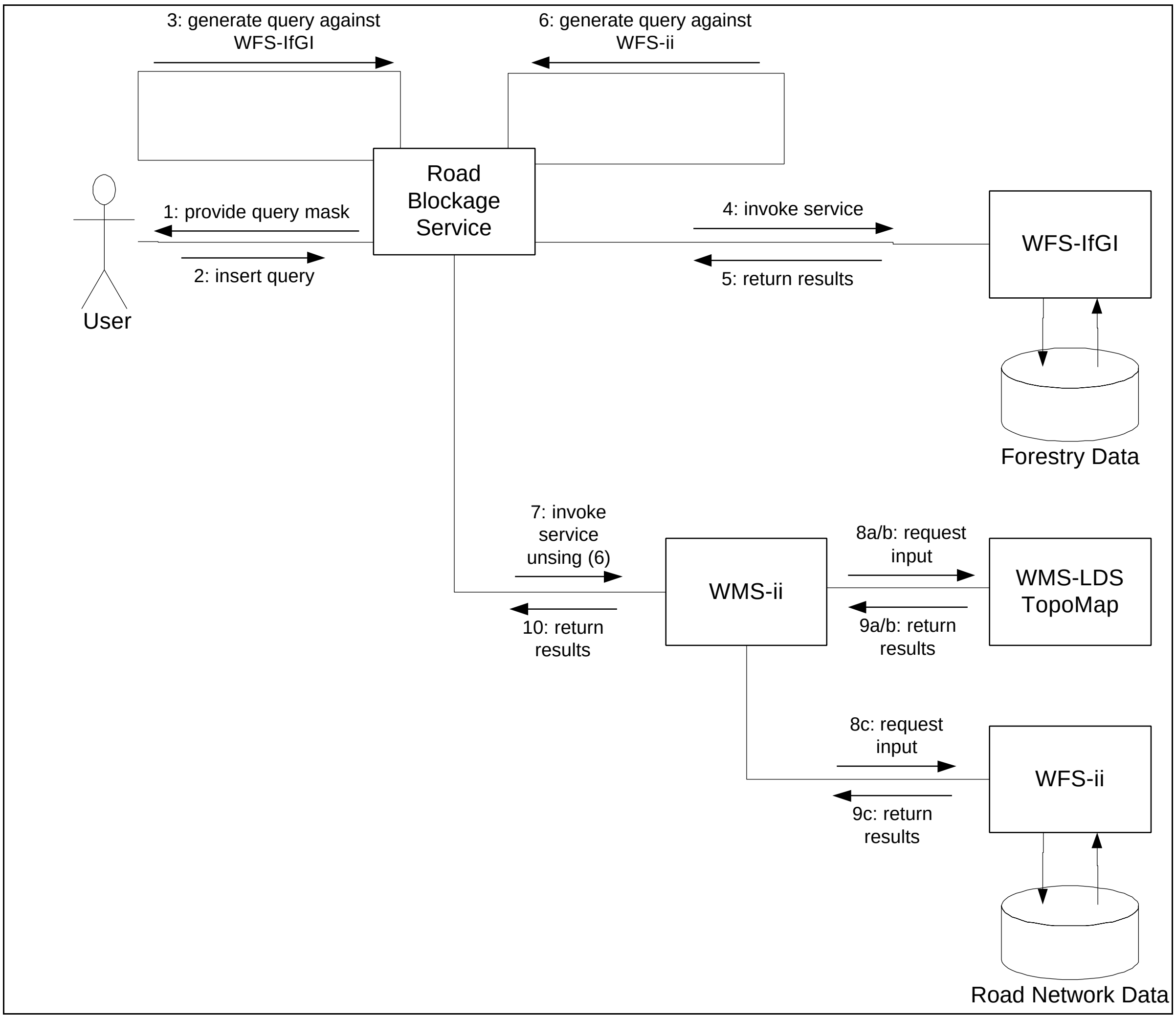

Figure 5: Flow of information in the state-of-art scenario

The data and services are annotated according to the current International Standard for Geographic Information Metadata Services (ISO/TC-211 2001) resp. for Geographic Information Services (ISO/TC-211 & OGC 2002). The implemented data model is conform to the XML schemas of the OpenGIS Geographic Markup Language Implementation Specification (GML 2.1.1).

Scenario 1.2: Including Remote Sensing Data

The first query in the chain described in the previous section looks for forest areas that are potentially at risk from wind throw, i.e. that contain trees of a specified species and age. The database on forests of the Forestry Department offers this information, but only for forests that are state property. It does not, however, contain any information on privately owned forest areas. This is a fundamental problem with geographic data because the content (i.e. attributes) are intimately tied to the application context or discipline for which the data was collected. In order to get a fully satisfactory response for a request (i.e. one containing privately and publicly owned forest stands) it is desirable to have access to additional data sources that contain information about forest areas. Thus, it will eventually be possible to dynamically choose the most suitable source of information or to combine several sources to answer the given question.

In the “road blockage” use case the dilemma that precise information is only available for publicly owned forests could be solved with the help of remote sensing data. In this extended scenario at least three additional services are required:

- a service providing the remote sensing data for the given area,
- a service for identifying forest areas in remote sensing data, and
- a service for estimating areas with overaged trees from remote sensing data

The information on age and tree species that is available for the forest areas that are state property (provided through the Forestry Department’s WFS) could be used as training data for the classification and the estimation services. However, even with this training data an automatic classification for identifying areas with overaged trees from remote sensing data is a very difficult task. Therefore the estimation service might require a “human in the loop” for controlling the classification.

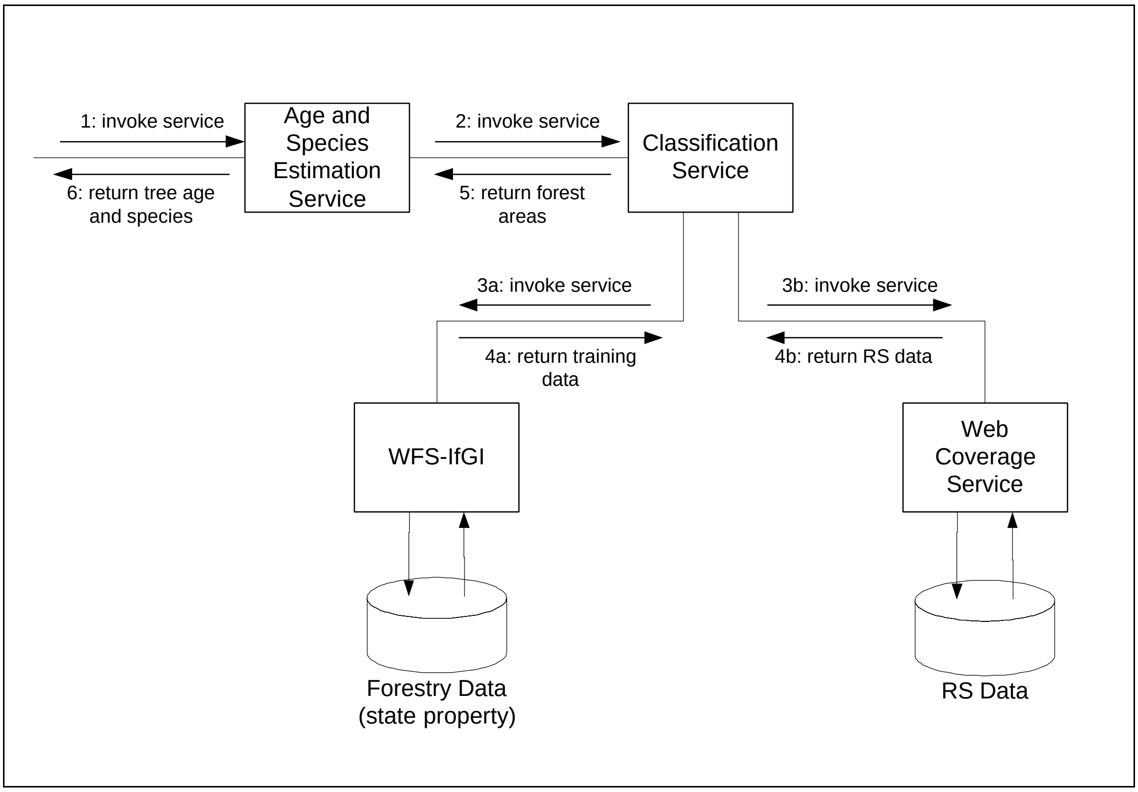

Figure 6: Flow of information in the “remote sensing” extension to the state-of-the-art scenario

Limitations

As the services used in the chain were not developed specifically for the service chain described, this scenario illustrates the benefits of interoperability standards, at the same time representing the state of the art in the area of GI service chaining. However, the scenario also has a number of limitations.

In order to be able to implement the client application the provider had to have knowledge of the services’ existence, their capabilities and their application schemas. The client application only works with these specific services and application schemas. Furthermore it cannot be reused for the execution of different chains as its workflow management is tied to this specific chain.

These limitations are addressed in a second, more flexible scenario, which is presented in the following section.

Scenario 2 – Ad-hoc Service Chaining

The second scenario describes a workflow-managed (translucent) service chain (ISO/TC-211 & OGC 2002). While the ISO/TC 211 standard only states that in such a scenario a pre-defined chain is selected by the user, we further assume that the chain definition does not exist at the time the user poses his question, but that it is assembled in an ad-hoc fashion by a workflow (composition) service. Accordingly, the user should be able to specify his question in a generic user interface, provided by appropriate human interaction services. Registry services should help the user to further specify his question by providing (semantically rich) information on available services. Workflow services will then be responsible for composing and executing a chain that answers the question thus, that the semantics of the answer match the semantics specified by the user. This includes the subtasks of translating the user's question into formal requirements, generating a solution strategy to solve a complex problem with several smaller tasks, matching the requirements to the capabilities of the available services and offering some quality measurements to evaluate the fitness of use of the service chain's results.

The flow of information is depicted in Figure 7. First, the user has to enter his query. In order to assist the user in this task, this will have to be a highly interactive and iterative rather than a linear process (1a-d). In order to do that, capability descriptions of available services are matched against the user requirements. That will require information on the available services and data from the registries as well as formalized domain knowledge. The query entered by the user is then translated into a workflow describing a service chain (2). Queries against the components of the chain have also to be formulated in appropriate query languages (3).

The execution of the actual service chain (as described in the first scenario) is controlled by the workflow management service.

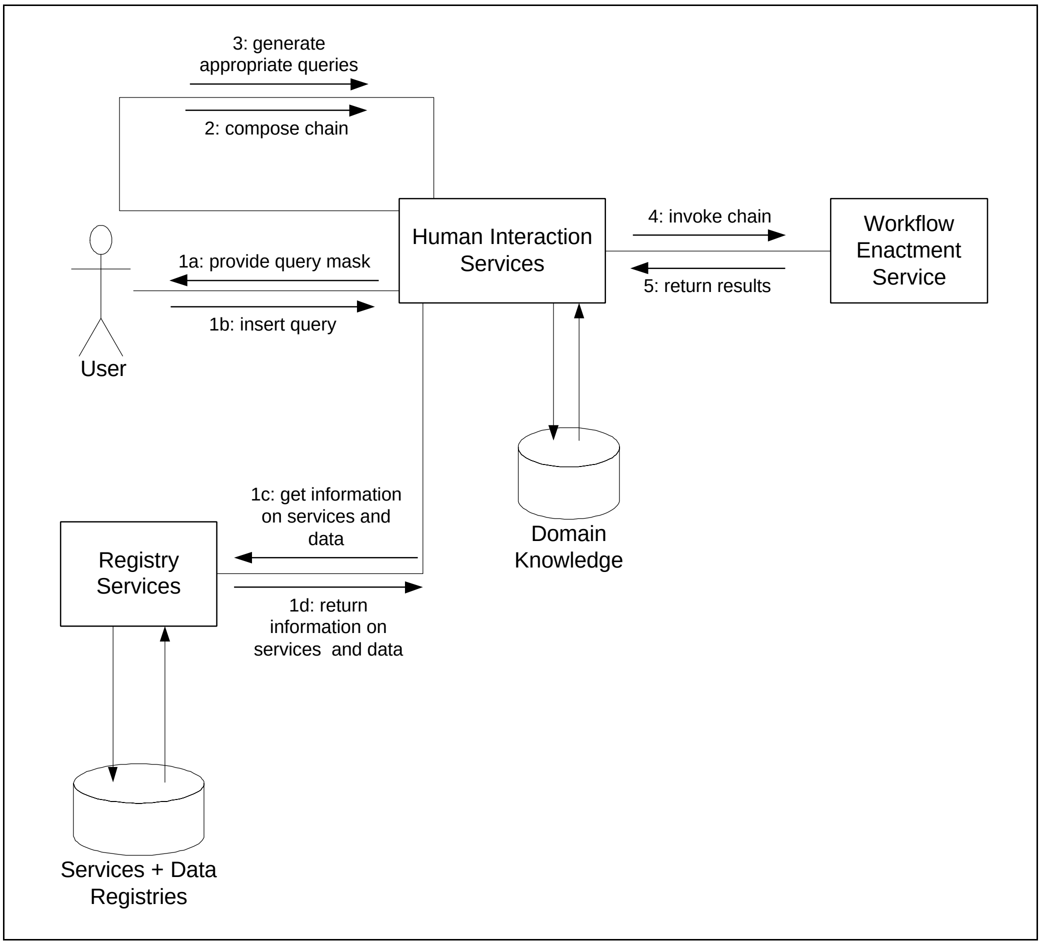

Figure 7: Flow of information using translucent service chaining (the actual execution of the chain is omitted)

Problems caused by semantic heterogeneity

The forestry data as well as the road data that is to be processed in this use case can originate in multiple information communities, resulting in heterogenous data models. The main focus in this scenario will be to integrate information from different sources (e.g. Forestry Directorate and remote sensing data) to derive new information (e.g. likelyhood of storm damage for roads) and to automate this process.

Thus, the focus is on semantic metadata interpretation (e.g. do the data sets representing forests refer to the same kinds of forests?), the semantic data interpretation (mapping between different nomenclatures, e.g. forest vs. woodland) and matching of service capabilities to human (for service discovery) and ultimately service requirements (for automatic service composition).

Use Case III: European Water Framework Directive

This use case deals with the semantic problems that will arise from the implementation of the Water Framework Directive (WFD) of the European Union (European Commission 2000) from December 22nd 2000. Within three years this directive has to be implemented into national law.

The WFD aims at bringing all surface water bodies and the ground water to at least a “good state” until the year 2015. The definition of contiguous River Basin Districts (RBD) should prohibit that administrative and political borders obstruct the water protection and encourages an integrated view on the rivers and their catchment areas.

The foundation for all decisions to be made and tasks to be carried out in the scope of the WFD is an extensive river basin management (Vogel 2002), which is based on the repetitive control of a variety of biotrophical and abiotrophical parameters. A combination of these criteria provide a possible rating between “bad” and “very good” for each river basin. However, the exact specifications of theses parameters are still in progress.

The survey of the required data is no single task but implies a permanent monitoring of the river state and results in the creation of regular reports for the European Commission. Additionally, the WFD attaches great importance to provide the collected information to all the interested citizens of the countries involved in the project.

The implementation of the WFD results in the necessity of governmental organisations at all levels (from local authorities, local environmental agencies and water federations, to federal environmental ministries, national environmental agencies, and the responsible institutions in the EU) to exchange and aggregate substantial data. Furthermore, it is reasonable to provide the collected data for scientific purposes and third parties. Additionally, it is important to be considerate to the demands of informing the public, since only by ensuring a simple access to the information, transparency can be guaranteed. Thus, the citizens of the countries involved form a user group with its own, specific characteristics.

Due to the necessity of agencies and organisations to co-operate through all administrative and political borders, manifold semantic problems on several levels occur. On the one hand the data are of a very heterogeneous nature, because they have been gathered by diverse institutions with varying methods and therefore exist in different formats, temporal resolutions and with different meanings. These data have been collected over years and should be made available for future needs. However, no exact specifications on the methods for collecting the data have been formulated until today, making a further usability of the data difficult. Solving these problems by providing innovative approaches is the goal of this project.

On the other hand there are considerable demands on the interoperability of metadata. Up to now all participating institutions use their own classifications and ratings in order to describe water bodies, their basins and protected areas, and to indicate the measured water quality.

In the same way the consequences of human influences (settlements, sources of pollution, etc.) have to be categorised into object classification schemes. Also on this level extensive datasets have been built up over the years, which should be made available for further use, too. Although national standards for metadata specifications will be established in the future, the need of a semantic translation of the describing data between the European countries will exist further on.

In the current stage of the implementation of the WFD, the responsible local and regional authorities are creating an inventory of the available natural resources. This phase is due to last until 2004. The activities related to the creation of an inventory include the cataloging of legacy data as well as intensive acquisition of new data. As the standardisation of a number of methods and criteria for key parameters is not yet completed, these activities cause the need for data integration in the future. One example is the classification of river types, for which several competing classification schemes (using fish populations, using geomorphological parameters, etc.) exist. Another example is the evaluation of water quality using biological parameters. Here, some indicator organisms are classified, but the respective analysis methods are not. The following use case scenarios will explore the specific needs of translating between different classification schemes, providing these tasks as services, aggregating several services into service chains and using remote sensing methods to support the monitoring process.

Scenario 1: Semantic translation of the river unit classification criteria

In the framework of the WFD each river basin has to be subdivided into several smaller river basin units, which will be classified independently from the other parts. This subdivision is oriented on meaningful biological or geomorphological factors including the prevalent fish population (Böcker 2002) or the sediment type of the river. This is done mainly because it is not useful to compare a region that is near to the source of the river to a region that is near to the estuary.

Since several classification schemes are already in use and will be used in the future in the participating countries, the translation between these schemes has to be carried out. Therefore this scenario will lead to a semantic translation service which is based on ontologies describing the used classification schemes. With the help of the service two tasks can be carried out:

- A given river unit can be (re-)classified using a different classification scheme.
- A given classification can be translated into a different classification scheme.

Thereby the interoperability of organisations using various schemes will be guaranteed.

Scenario 2: Search for the reference river unit

Based on the service developed in the first scenario a very important task can be carried out more easily: the search for the reference river unit. The rating between “bad” and “very good” is not an absolute but a relative assessment which is based on a comparison to another river or river unit. This river with a given rating has to be as similar as possible, i.e. it should own the same biological and geomorphological characteristics. It can run, however, – maybe partly – in a country which uses a different classification scheme. This leads to the demand of translating a particular classification from one scheme to another one as described in the scenario above.

Some additional information is needed to decide whether a selected river unit is suitable as a reference, including parameters as the altitude and the climate. The result of this scenario is intended to be a map of Europe with all eligible rivers or river units being marked as potential candidates.

Using the technology of OGC-conformant services, a service chain has to be implemented and extended by the service of searching for the reference unit.

Hereby, two different sub-scenarios are conceivable. In the first one a classification is configured and arranged based on the whole range of known schemes and parameters. In the second scenario an already classified river or river unit acts as the starting point for searching for all equivalents. In both sub-scenarios the query can be modified in several steps including the degree of accordance in order to control the number of resulting hits.

Scenario 3: Semantic translation of the water quality criteria

In contrast to the factors that are used to divide a river into several river units, the parameters that have to be measured to determine the water quality are already specified in the WFD. Although the parameters themselves are fixed, the methods to perform these measurements are not. For example, when counting elements of plankton, organisms from different levels of the biological taxonomy can be selected as indicators. Since institutes performing the examinations have used and will use different organisms, a translation service is needed.

This scenario is similar to scenario 1 and will produce in a first step a single service. In the next step this service will be inserted into the existing service chain to allow a transparent mapping between the used quality criteria schemes.

Scenario 4: Remotely sensed data for monitoring tasks

Another very important task prescribed by the WFD is the permanent monitoring of the water quality and the anthropogenic changes of the water bodies. Especially sources of pollution (point sources or diffuse sources) need to be controlled intensively in order to improve the basic conditions of the rivers. This is an expensive task, both in terms of time and money.

The usage of remotely sensed data for monitoring campaigns will be evaluated in this scenario. In a first step the applicability of normal optical data will be examined, which includes the detection of pollutants like oil and chemical substances. In a later step infrared and hyperspectral data could be used to detect pollutions that are not visible in the normal spectrum. As with the other scenarios above these tasks will be provided as services, with the opportunity to include them into static or ad-hoc service chains.

Conclusions

The Water Framework Directive of the European Union, which has to be implemented within three years by all member states, causes a variety of semantic problems. In order to establish interoperability between both governmental and non-governmetal organisations through all administrative and political borders and on several levels, the syntactical and semantical heterogeneity has to be overcome, which arises from the different classification schemes used by the participants. Therefore the WFD is not only an ideal basis for further research in semantic translation and for meaningful uses cases but also a concrete task which solution is required within a restricted period of time.

References

Bernard, L. (2002): *Experiences from an implementation Testbed to set up a national SDI,* in: Ruiz, M., M. Gould & J. Ramon (Ed.): 5th AGILE Conference on Geographic Information Science 2002, pp 315-321.

Böcker, K. et al (2002): *Zwischenergebnisse der Erarbeitung des Flussgebietsplans des Wupperverbandes*. In: 5. Symposium Flussgebietsmanagement beim Wupperverband - Regionales Wasserwirtschaftsforum, pp 27-32. Wupperverband, Wuppertal.

Bundesanstalt für Gewässerkunde (2000): *Internationaler Workshop im Rahmen der Machbarkeitsstudie für ein computergestütztes Entscheidungsunterstützungssystem* (Decision Support System, DSS) für die Elbe, URL: *http://www.bafg.de/html/aktuell/ workshop_dss.htm*

European Commission (2000): *Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 establishing a framework for Community action in the field of water policy*, URL: *http://europa.eu.int/comm/environment/water/water-framework/index_en.html*

Gemeindetag Baden-Württemberg (2000): *Zum zweiten Mal innerhalb eines Jahrzehnts: Verheerende Stürme in Baden-Württemberg*.

ISO/TC-211 (2001): *Text for DIS 19115 Geogaphic information – Metadata*, International Organization for Standardization.

ISO/TC-211 & OGC (2002): *Geogaphic information – Services*, Draft ISO/DIS 19119. OpenGIS Service Architecture Vs. 4.3. Draft Version, International Organization for Standardization & OpenGIS Consortium.

Kuhn, W. et al. (2001): *Referenzmodell 3.0 - GDI Geodaten-Infrastruktur Nordrhein-Westfalen*. Land NRW (Ed.), Media NRW, Band 26.

Matthies, M et al. (2001): *Georeferenced simulation and aquatic exposure assessment*, in: Wilderer (Ed.): Water Science & Technology 43(7), pp 231-238.

Ministerium des Innern des Landes Brandenburg (Ed.)(2001): *Brandenburg kommunal*, 34, pp 3-9.

OGC (2002a): *Styled Layer Descriptor Implementation Specification (GML)*, Version 1.0.0 OpenGIS Project, URL: http://www.opengis.org.

OGC (2002b): *Web Feature Server Interface Implementation Specification*, Version 1.0.0 OpenGIS Project, URL: http://www.opengis.org.

OGC (2002c): *Web Map Server Interface Implementation Specification*, Version 1.1.1 OpenGIS Project, URL: http://www.opengis.org

OGC (1999): The OpenGIS Abstract Specification. Topic 14: *Semantics and Information Communities*, Version 4 OpenGIS Project, URL: http://www.opengis.org

Vogel, W. R. (2002): *The EU – Water Framework Directive – A Challenge for Information Management*. In: Pillmann, W. & Tochtermann, K, (Ed.): Environmental Communication in the Information Society, EnviroInfo Vienna 2002 16th Int. Conference: Informatics for Environmental Protection, part 2, pp 556-560. GI Gesellschaft für Informatik, Bonn.