Integrating Knowledge Bases into SDI's

Ingo Simonis
Institute for Geoinformatics, University of Münster
Robert-Koch-Str. 26-28, 48149 Münster, Germany
e-mail: simonis@uni-muenster.de
fon: +49 251 8333084
fax: +49 251 8339763

SUMMARY

Spatial Data Infrastructures (SDI) define a framework that allows on-line access to distributed geographic data and geoprocessing functionalities, though only very few standardized processing capabilities are yet usable. For the modeler, providing a sophisticated application in form of a web service using interoperable frameworks means that he has to be in expert not only in his knowledge domain but in web service orchestration and chaining additionally. Simple questions with answers based on integrated sensor results, like "do I have smog at location x" have to be modeled using GIS techniques before being publishable as a web service. This paper prepares the ground for integrating expert systems that are based on autonomous sensor networks into spatial data infrastructures and frees the expert of the burden of in-depth knowledge of service coupling and managing. Most of the geoprocessing capabilities will be part of the – the expert system wrapping – service. Chaining and orchestration will be mostly performed by the backward chaining system of the inference machine.

KEYWORDS: *Knowledge Bases, Spatial Data Infrastructures, Sensor Web*

INTRODUCTION

Current spatial data infrastructures are typically based on distributed web services and aim to replace the predominating local management and processing of data sets. SDIs develop their real performance on national or international scales only in case that cross-system and cross-institutional interoperability is achieved. International accepted standards help to realize such large-scale distributed infrastructures assembled by numerous heterogeneous, but interoperating services. Primarily based on the efforts of the Open GIS consortium standards had and have been developed to support the cooperative use and processing of spatial data. These services are currently being improved and enlarged in functionality (OGC 2002). A large step towards real-life GI-systems was achieved in the course of the Sensor Web Initiative. Its initial focus was to develop standardized interfaces to access data which is provided by live sensors operating in near-real-time, rather than by conventional static data store (Simonis et al. 2003a). During the last OGC Open Web Service Testbed the initiative has been extended to a complex, multi web service based framework facilitating all stages of observation planning, feasibility testing, and measurement as well as simulation execution and execution (Simonis and Wytzisk 2003; Simonis et al. 2003b).

Still, most of developments within OGC focus the end user, the one consuming maps or feature data from web map or web feature servers. There exist no mechanisms helping the modeler to compile complex chains made up of individual web services, so called service chains. To provide a simple digital map which is composed of data from different data providers, the business logic has to be hard coded at first. No description language to define service chains is found. All data providers have to be known in advance because no automated publish-find-bind is realized yet.

This article aims at showing how SDIs can be equipped with functionalities to generically derive aggregated intention or decision relevant information on the basis of distributed information sources. This requires capabilities to dynamically link distributed spatially enabled services and to jointly analyze the information they provide. The formal frame of the underlying work is constituted by approaches performed with rule based expert systems that were extended by specific spatial related operations. The presented approach shows how expert systems can act as decision support tools and upvalue interoperable SDIs enormously.

THE LORRY USE CASE

Imagine the following scenario: At three o'clock in the morning, a lorry loaded with Acrylnitril is overturned in a car accident within the city limits of a larger city. It has to be supposed that the entire tank is leaked: Nearly five thousand kilogram of the highly toxic liquid poisons the surrounding area. The following questions have to be answered immediately: How the current situation has to be assessed? Are there any areas that have to be evacuated or any streets that have to be blocked? An underestimation will almost certainly lead to disastrous results; an overestimation will lead to high costs that would have been evadable. Either way, a decision support system, based on a variety of parameters or actual observations that allow the calculation of danger levels and danger zones would be helpful.

How could a standards based web service be modeled to support the person in charge of operations in our scenario illustrated above? On the one hand, a wide range of spatial data stored in rather static data stores and accessed using standardized web feature services exists and could be visualized using web map services. Even highly dynamic data like current wind speed, precipitation and other observable are available at web services. On the other hand, the integration or better the integration process has to be performed in advance by the service modeler. He is in charge to chain the services in the correct manner, deal with all conversions and taxonomic or ontological imponderabilities and incompatibilities, and does all processing using highly complex filter systems or has to incorporate convenient predefined GIS functionalities into his web services.

Following the *Emergency Response Planning Guidelines* which are developed by the Organization Resources Counselors (ORC) and the American Industrial Hygiene Association (AIHA), it is very complicated to take actions and procedures in a well structured manner. Help could be provided by knowledge based systems that have expertise of complex connections and a problem solver which allows the calculation of (disaster) scenarios.

INTEGRATION OF KNOWLEDGE BASED SYSTEMS

Knowledge based systems distinguish themselves from conventional systems by explicit instead of implicit storing of the domain specific knowledge. They excel in having a clear interface between specific knowledge and the problem solving strategy (so called inference machine). Therefore it becomes possible to change the explicitly stored knowledge or the knowledge base without modifying the hard coded algorithms as it is the case at conventional systems (Gonzales and Dankel 1993). This article will not address knowledge based systems in detail, but will represent an approach to include such systems into a web service based infrastructure. We will use rule based expert systems as one of many forms for the purpose of representing knowledge. Rule based systems express the knowledge by means of rules. For our scenario, the following (simplified) rules apply:

- If the leaked chemical is toxic, calculate its distribution.
- If the distribution cloud covers any streets, mark these streets as necessary to block.
- If the distribution cloud is highly toxic, and if it covers inhabited area, evacuate this area.

Analyzing the nature of these rules, the following problems have to be solved to offer knowledge as a web service within a standardized framework:

- How does the conceptual integration of the knowledge base into a standardized spatial data infrastructure look like?
- Which extensions to existent web services or web service interfaces become necessary to make the formalized knowledge accessible via a web service?
- How can knowledge bases be applied to a priori unknown fact sources?
- How does an intuitive non-logical description language look like that allows the formalization of the spatial and temporal knowledge?

The conceptual framework

The integration of knowledge bases into a web service requires functionality on different stages of the exploitation process. A domain expert has to formalize his knowledge in an appropriate way before he will be able to register his knowledge at the service.

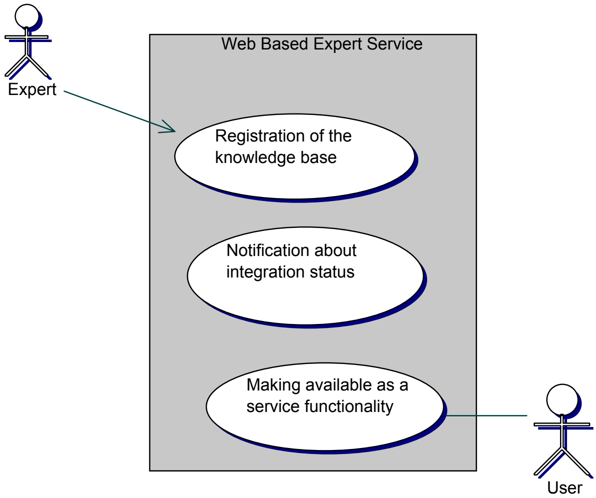

Figure 1: Knowledge exploitation in UML notation

The formalized knowledge (regarding the knowledge formalization see section *Description language*) will be registered at the Web Expert Service (*WES* in the following), parsed and analyzed for errors or inconsistencies and finally made available to the public (see *Figure 1*). If a user wants to make use of the service, all he has to do is to define the location, as well as the kind and amount of toxic material that is released. The WES makes then use of two further components, the fact seeker and its inference machine (see *Figure 2*) to (try) to solve the problem.

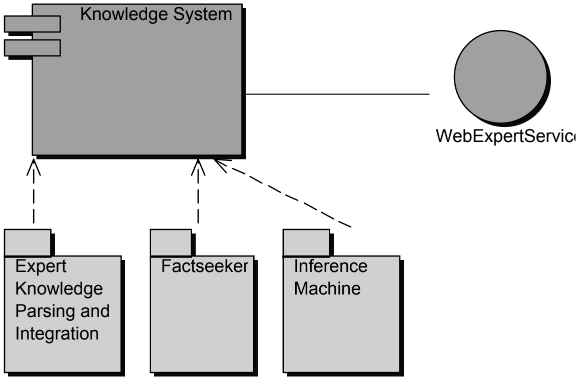

Figure 2: Components of the web expert service in UML notation

The fact seeker is scanning catalogues in the web for suitable sensor and other feature data services, posts the necessary requests and integrates the data into the fact base. It is than the duty of the inference machine to solve the problem.

Description language

The language to describe the knowledge should be readable be machines and humans. To make it easier for the domain expert to formalize his knowledge, you should refrain from using complex logic based styles or languages. Therefore the primary task the description language should meet is human readability. A closer look at predicate logic languages shows that they are rather illegible. By contrast, XML based description with expressive tag names are primarily designed to fulfill intuitive human understanding and readability without being rising bigger problems for machines. The formalized rules will be based on XML schemata to ensure proper syntax usage. The WES is then responsible to parse and compile the XML encoded rules to make them available for its system shell (the inference machine). To optimize the level of interoperability, all rules are substitutes for the GML (Geography Markup Language) substitution group "_Feature", so they are in compliance with the GML guidelines.

Three different kinds of rules exit:

- **ObjectRules** define simple or complex Objects which shall be used within the knowledge base.
- **RelationRules** define topological relation between individual objects, e.g. Object_A intersects Object_B or Object_C isNorthOf Object_D
- **ProcedureRules** define procedures that are started in case that the rule will fire, e.g. `(Object_A => Printout("Please Evacuate this city")`; or describe GIS-operations, e.g. (BUFFER (BaseObject InnerDist OuterDist)).

For some objects the temporal aspects are of major importance, others are completely time independent. If you want to calculate the toxic cloud distribution of our scenario, there is not much sense in using three days old wind speed and wind direction information. Time dependency is supported by the XML representation for open and closed time intervals as well as for specific moments in time.

Statistical distributions or means often play a major role in defining objects in case that more than one observation was or is performed in the defined time interval. E.g. stipulates the German smog decree that in summer time, smog warning level one has to be announced in case that the averaged ozone values in the air of the last hour at certain locations exceed 180 g/m³. Therefore, the object definition can be refined by statistical instructions, e.g. take the arithmetic, geometric, or harmonic mean of the specified time interval.

Sensor data of unknown fact sources

A web service offering calculated scenarios which shall support decision making in the above mentioned accident have to integrate actual measurement data independent of any institutional or organizational borders. To provide a distribution prognosis, at least current wind speed and wind direction as well as precipitation are mandatory parameters for any kind of distribution calculation. The domain expert formalizes his knowledge without knowing any fact sources. Therefore it is never for sure that the necessary sensors will be found in the area of interest. Even if a dense sensor network exists, usually no sensor will be exactly placed at the inquired position. Those problems can be solved by interpolation rules. The inference machine uses backward chaining mechanisms to inference new knowledge based on current facts. Rules exist that do an interpolation of any given object value. These rules are predefined by the web expert service. The domain expert has to define in his ObjectRules what kind of interpolation mechanism should apply and has to specify the necessary parameters only.

To ensure that all participants have the same understanding of individual ObjectRule parameters, they are organized in web based dictionaries on principle. Those dictionaries exist for observables, units of measure and interpolation algorithms. Therefore it is ensured that the observables defined by the domain expert are the same the fact seeker finds within the web. Every entity has its universal identifier. The dictionaries are defined in compliance to GML3. *Error! Reference source not found.* illustrates a reference to a unit of measure; microgram in this case. All units of measure are based on the SI unit system (Taylor 1991). The usage of the XLink technology allows making use of other units defined in arbitrary but GML3 compliant dictionaries as far as they are traceable down to the original SI unit. In this case, microgram references a unit called gram which is defined in another dictionary. This unit gram is not a core SI unit either. The GML3 schema ensures that even gram will have a reference to its corresponding SI unit (which is kilogram).

```
<UIDObject>
   <...>
   <UnitOfMeasure>
      <gml:ConventionalUnit gml:id="µgm-3">
         <gml:name>MicrogramPerCubicmeter</gml:name>
         <gml:quantityType>Concentration</gml:quantityType>
         <gml:conversionToPreferredUnit
              uom="http://mars.uni-muenster.de/Units.html#g">
              <gml:factor>0.001</gml:factor>
           </gml:conversionToPreferredUnit>
           <gml:unitDerivation>
              <gml:unitTerm
                 uom="http://mars.uni-muenster.de/Units.html#m3"
                 exponent="-1"/>
           </gml:unitDerivation>
         </gml:ConventionalUnit>
   </UnitOfMeasure>
   <...>
</UIDObject>
```

Listing 1: XML representation of unit of measure identifier

This mechanism allows the inference machine to solve a problem even if the defined value to compare has a different unit of measure. Predefined rules exist or are dynamically constructed in case that unit conversion becomes necessary. Due to the schema based well defined syntax, dynamic unit conversion rules generation can be performed.

Expert System

Expert Systems make use of two different types of rules: Forward-chaining and backward-chaining rules. Forward-chaining rules are somewhat like *if…then* statements in a procedural language (though without being used in a procedural way: They are not executed at a specific time or order). Backward-chaining rules are also similar to *if…then* statements but the rule (or better the inference machine) actively tries to satisfy the condition of its *if* part. The WES makes use of backward-chaining whenever a translation of an existing fact becomes necessary. This is a regular problem because facts of the fact base do not necessarily match all criteria of a rule. *Listing 2* shows such an example. The left hand side (LHS) of the rule cannot be satisfied by any fact of the fact base because all facts containing information about the amount of CO2 use the unit of measure g/m³. For this purpose, the WES contains a wide range of rules that solve those problems, like rules for transformation, interpolation, conversion, etc. Making use of this capability provided by the expert system, the domain expert does not have to take care about proper chaining of all services involved in the required operation.

```
( (> Amount_of_CO2_in_µgm-3 180 ) => ... )
```

Listing 2: Example of a rule

Extension of existing standardized services

Finding and accessing data is mostly performed using Sensor Collection Services (McCarty 2002) or Web Feature Services (Vretanos 2002). Both of them currently do not support any kind of universal identifier. Currently, there is a discussion if the Sensor Collection Service will be supplemented by an *Observation and Measurement Service* within the OpenGIS standardization consortium. We demand the integration of universal identifiers into the sensor based networks (for a beginning) to allow proper referencing and functionality. We are aware of the fact that consensus process of identifiers definition will be a difficult one.

CONCLUSIONS

This paper gives an overview how expert systems could be integrated into web service infrastructures. At this stage, the web expert system exists in a first prototype stage. First tests have shown that the intended simplification of the knowledge formalization will bring some restrictions to the entire system. Not all the power of logical programming languages can be reached by using a XML based syntax to describe the knowledge. It might be possible to achieve a very high level of expressiveness, but would be payable by readability. A compromise that levels readability and expressiveness is currently under development.

BIBLIOGRAPHY

Gonzales, A. and Dankel, D. D. (1993) The Engineering of Knowledge-Based Systems. Theory and Practice. Prentice Hall, Upper Saddle River, NJ

McCarty, T. (2002) OGC Web Services - Sensor Collection Service Version DIPR Version 0.0.6. Open GIS Consortium (OGC), Technical Report

OGC (2002) OGC Vision, Mission, Goals, Values. Open GIS Consortium, Inc., Technical Report

Simonis, I. and Wytzisk, A. (2003) Simulation Models in SDI. Integrating Simulation Standards in Geo-processing. GIM International 35(10)

Simonis, I., et al. (2003a) Interoperability of Simulation and Geoinformation Standards. Proceedings of: Spring-SIW, Orlando, Florida

Simonis, I., Wytzisk, A., and Lutz, M. (2003b) Integrating Live Sensors and Simulation Models in Spatial Data Infrastructures. Proceedings of: 9th EC-GI&GIS Workshop, La Coruna, Spain, pp. 12

Taylor, B. N. (1991) The International System of Units (SI). National Institute of Standards and Technology, Washington DC, Technical Report NIST Special Publication 330

Vretanos, P. A. (2002) Web Feature Service Implementation Specification Version 1.0.0. Open GIS Consortium (OGC), Technical Report