Proceedings of the 6th AGILE
April 24th-26th, 2003 – Lyon, France

SEMANTIC EXTENSION OF GEO WEB SERVICE DESCRIPTIONS WITH ONTOLOGY LANGUAGES

Rob Lemmens[1], Marian de Vries[2] and Trias Aditya[1]

[1] Department of Geoinformation Processing
International Institute for Geo-Information Science and Earth Observation (ITC)
P.O. Box 6
7500 AA Enschede, The Netherlands
phone: +31 (53) 4874529
fax: +31 (53) 4874335
e-mail: lemmens@itc.nl

[2] Geodesy department, section GIS-technology
Delft University of Technology (TU Delft)
Delft, The Netherlands
e-mail: m.e.devries@geo.tudelft.nl

1. INTRODUCTION

In the fast changing world of GIS and information technology, a growing number of users require tailored solutions, which they can define and build themselves on short notice or even in real-time, without having to make use of an entire pre-built GIS software package with proprietary data. This requires an easily accessible developing environment that allows for scalable solutions.

The answer to the problem of scalability can be found in providing GIS operations and data as components. In a world of component GIS, users will be able to discover the components they need and assemble them into their own application, on demand. Users may be thin clients, but eventually find themselves connected to a huge server-side GIS component space. This rather simple concept is not so easy to implement, because we deal with different types of users, ranging from GIS experts to users that are not aware of the spatial data handling processes required. In recent years, significant effort has been put into the development of interoperable GIS software. However, for the end-user the assembly of a working GIS solution takes more than just a set of interoperable GIS components. The main concerns are the description of the functionality of software components, the characteristics of the data, the laws of proper spatial data processing and the connectivity between components.

The new paradigm of Web Services is a promising technology that will enable us to use XML based technology to model these connections. Web Services are self-describing, self-contained applications that can be published and invoked over the web. Institutions and individuals tend to interact dynamically not only inside but also outside their network. That is the reason why the number of Web Services is growing rapidly. Moreover the W3C recommendations on SOAP and WSDL provide very good building blocks for ensuring basic (syntactic) interoperability. Semantic interoperability however remains unsolved and will be the topic of this paper. Our discussion will be supported by experiments with a prototype registry explorer / description mediator.

2. INTEROPERABILITY DEPENDS ON SERVICE CAPABILITY DESCRIPTIONS

The interoperability between operations, often referred to as system interoperability, relies on how well the output data of operation A matches the input requirements of operation B. Therefore, a thorough understanding of the correspondence between data set and data requirements of operations through clear descriptions, is important to solve interoperability problems.

For example in a geospatial context a Web Service that offers a buffer operation has to specify whether or not it can handle data that is not projected but in a geographic reference system (lat/lon). The calculation of distance does not make much sense in some reference systems, and a coordinate transformation into a projected system must be performed first. This can be transparent to the end user (the buffer operation calls the coordinate transformation service itself); but it can also be an operation that has to be called explicitly by the user, before the buffer operation is requested. Whichever is the case, whether or not all spatial reference systems are supported needs to be somewhere in the Web Service description for the buffer operation. Other items that have to be specified are: the units of measurement the service is able to handle (only meters or kilometers or also feet etc.) and whether only the geometry of the buffer is given back as output or a complete clip of the data within the buffer.

Another example of the need for service metadata and adequate expressiveness of the metadata description language has to do with visualization aspects and graphic capabilities of web map services (see also [1], [2]). 'Simple' web map services can decide only to offer default styling, that is defined beforehand by the data provider. More advanced map services however can offer the possibility of interactive classification and styling by the user. Does the map service offer interactive classification and if so, what types of classification is it able to perform: unique value, interval ranges, classes with an enumeration of codes? Can it handle classification based on the values in more that one attribute (multi-attribute), can it handle null values, what about the symbology sets etc.

From the design of the functionality of the Web Service there is a direct link to the specification of the interface to the calling software: what input parameters must be supplied in the user request, what values can exist for each parameter, and how will the output be returned. At runtime this information is used in two ways: to inform the end user of the capabilities of the service, and to function as input for an automated check on the fitness of use of data and service in that particular combination.

3. CONNECTIVITY LEVELS BETWEEN DATA AND OPERATION

Within services the relationships between data and operations are materialised at different connectivity1 levels similar to a protocol stack (see figure 1). A dataset embodies characteristics at all levels. A road dataset for example can be characterised by its road class definitions such as ‘primary road: road width = 12 meters’ (semantic level), its inherent line feature objects (data structure level) and its data format. However, descriptions of datasets do not necessarily contain information on all levels. For data processing, operations always need to ‘know’ the identity of the data at the lowest level (data format) before any processing can take place, but this is not necessarily so at higher levels in the stack (for example when we perform a simple merge of datasets). The level of connectivity

1 We use the term connectivity with reference to [3], where data format, data structure and semantics are residing in one layer, half way the Services interoperability stack between connectivity and interoperability.

that we can anticipate on, depends on the availability of meta information linked to dataset and operation at those levels. In proprietary systems this meta information is often not exhibited, in open systems we need this information to link data to operations. In an automated environment this is expected to be carried out by a mechanism or agent that

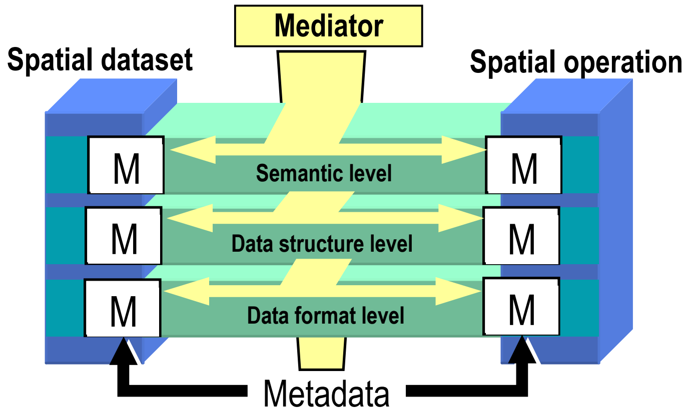

Fig. 1 Connectivity stack: levels of connectivity between data sets and operations. A mediator evaluates the potential connectivity based on the metadata (at each level).

mediates between data and operation. Such mediators must rely on the meta information of the components between which they mediate and they need a framework in which they reference this information.

An interesting development is currently taking place at the data format level. Historically, data formats do not reveal much meta information on the semantic level. With the advent of XML-based dataset formats, such as GML2, it is easier to derive higher level descriptions from the data, through text-based tags, namespaces and schema definitions. GML will stimulate domain-specific spatial schemas and allows for publication of these schemas on the web, thus improving discovery of specific types of datasets. Nolan et al. [5] submit that GML serves as a starting point for an ontology3 for geospatial data.

4. ONTOLOGIES AND TAXONOMIES : A REFERENCE FRAMEWORK FOR MEDIATORS

A good description framework is flexible and at the same time rigid enough to enforce interoperability between operations and data. It has to fulfil the needs of different types of user as well as the requirements of the component providers. In this respect, XML has the advantage of being extensible, and both machine and human readable, but is open enough to trigger developers in writing their own, again inflexible, XML-based description language. The XML based RDF provides and prescribes more specific handles for resource descriptions with the help of Uniform Resource Identifiers. This allows for uniform discovery of globally unique resources (data and software). A mediation framework based on RDF is

2 The Geography Markup Language (GML) is an XML encoding for the transport and storage of geographic information, including both the spatial and non-spatial properties of geographic features [4].
3 An ontology contains a taxonomy and a set of inference rules. The taxonomy defines classes of objects and relations among them.

promising, but still needs a common semantic understanding of GIS terms amongst the global user community.

Consider an RDF triple that prescribes a coordinate system to be of type ‘Geographic’:

```
<?xml version="1.0"?>
<RDF:RDF xmlns:RDF=http://www.w3.org/1999/02/22-rdf-syntax-ns#
xmlns:GEODATA="http://www.itcdata.nl/rdf#">
    <RDF:Seq about="url:http://www.itc.nl/projection_service:conditions">
        <RDF:li>
            <RDF:Description about="url:http://www.itc.nl/projection_service:input">
                <GEODATA:coordsys>geographic</GEODATA:coordsys>
            </RDF:Description>
        </RDF:li>
    </RDF:Seq>
</RDF:RDF>
```

The pitfall is that in this example the value ‘Geographic’ is just a term that is not supported by any referenced meaning. A more advanced version of this triple would refer to an external description or ontology, giving the meaning of the term Geographic. When this description has a well known location, for instance materialised with a URI, then other resources can refer to the same description and have the same understanding of the value ‘Geographic’. This could be enforced by the use of ontologies. It is preferred that such ontologies become a standard accepted by the user community. Different communities can share ontologies or parts of it to form information communities [6] and different but similar ontologies can even co-exist when we define mappings between them. An advantage of such an approach is the extensibility of interoperating systems after building the ontologies [7]. Compared to other domains, the domain of GIS is characterized by the geospatial connotation of its data and operations and the different representations of geospatial features. Description methods for GIS components and their mediators therefore rely heavily on the inclusion of spatial feature ontologies.

Taxonomies have been used widely to model application domains and can be of help in our geo services model when they are integrated with the abovementioned ontologies. A taxonomy of datasets at the data structure level classifies for example datasets in vector and raster structures. A taxonomy of GIS operations would enable us to include statements on the connectivity of types of operations for the current dataset. Further, such statements can also be embodied in descriptions of operations. Although the OpenGIS Consortium (OGC) initiative on OWS Service Registry considers providing a common taxonomy of OGC services which is based on ISO19119, an operational taxonomy of GIS services is not in place yet. For this reason, we extend the RDF Schema of the OGC taxonomy with DAML+OIL encoding to ensure the better representation of this taxonomy (e.g. for describing not-subclass of). DAML+OIL is a set of constructs to create ontologies and to markup information so that it is machine understandable [8]. When it is deployed to build a taxonomy, DAML+OIL can support useful modelling primitives such as subclass of, disjoint decomposition, exhaustive subclass decomposition, and not subclass of (see [9]).

5. GEO WEB SERVICE DESCRIPTIONS : INVOCATION SYNTAX AND SEMANTICS

A Web Service description informs the human user and automated mediator on what the service provides and how to access it (describing the service as a collection of ports). With the intention of providing a mechanism for the user to find the service, the service description in the format of WSDL is published in a UDDI registry service. Basically there are two types of descriptions of services being published (see [10]). The WSDL service interface is published as tModel in UDDI, and the WSDL service implementation is published for describing business services, which can contain one or more binding templates [11]. These binding templates must refer to a specified tModel. Then in the UDDI model, one can search the intended service in a registry by submitting keywords based on name of service, name of binding or tModel.

The full capability of Web Services would be better used if the interoperability can be assured at all levels (format, structure, semantic). WSDL and UDDI do not support semantic description of services [12,13]. We therefore decided to investigate more advanced methods to implement semantic descriptions in Web Services. Our efforts on resolving semantic interoperability are making use of the achievements of the semantic web community and new developments in ontology languages such as DAML+OIL. Worth mentioning is DAML-S which is a DAML+OIL based language for describing the properties and capabilities of Web Services [12]. To overcome the gap between UDDI and these semantic descriptions. [14] proposes the mapping of DAML-S profiles into UDDI business services.

The OpenGIS Consortium (OGC) initiative on OWS Service Registry is aiming to provide a standard for searching, classifying, and finding geo services. This draft specification is heavily based on ebXML RIM instead of UDDI [15]. While the OWS Service Registry is still in a conceptual stage and considering the maturity of UDDI and the fact that it is widely-used, we choose UDDI registry as a starting point for implementing a registry explorer / description mediator for our geo Web Services.

6. PROTOTYPE REGISTRY EXPLORER / DESCRIPTION MEDIATOR FOR SERVICE CHAINS

Our prototype is based on the extension of the OGC taxonomy schema using the ontology language DAML+OIL. In order to overcome the UDDI limitations we provide semantic service descriptions and matching capabilities upon these descriptions. And once we have multiple semantic service descriptions, we can build possible service compositions through content-aware service chaining. In the conceptual stage, the possible solution can be established either using XDD-XET [13] or DAML-S. This composition should be able to be defined using a “shared” ontology of geo Web Services.

In our test environment we establish a chain between the PlaceFinder service by ESRI [16] with a closest-point locator (developed by the authors) resulting in a chained service that provides information on the nearest point of interest for a given place name. In order to realize the semantic description we create a DAML-S service profile for both of these services, and then we compose these two services as a new geo processing service. According to the ISO19119 taxonomy we can define PlaceFinder as an instance of the geocoding class, and the closest-point locator as an instance of the spatial proximity class and their combination results in a new instance of the spatial proximity class. All these classes are maintained with the consistent relationship and individual properties (slots) within the geo Web Services ontology.

In another test, we are setting up two Web Services: a Web Service that offers a simple buffer operation, another that offers a coordinate transformation between a small set of (European) spatial reference systems (both lat/lon and projected). The two services can be called separately (from a simple test client), but also used in a service chain: first a coordinate transformation, then a buffer operation, and after that again a coordinate transformation, so that the output buffer is in the same reference system as the original input dataset.

We deploy the registry explorer for searching the service that has been published. The registry explorer is using UDDI API for searching UDDI tModel just like a common UDDI explorer. We extend this model by adding capabilities of searching semantic service description. This prototype should extend the possible keywords not only by name, by bindings, by tModel as it is in UDDI, but also has the capability to search by functionalities. Searching based on functionalities will be possible since every tModel in the registry system refers to a specific class in a geo Web Services ontology.

7. REFERENCES

[1] Brändli, M. and Sparenborg, J., 2002, SVG as graphical metadata for distributed spatial data processing. *SVG Open Developers Conference*, Zürich, Switzerland, July 15-17.

[2] Keller, S.F., 2002, Towards Open Geo-Visualization Services. *SVG Open Developers Conference*, Zürich, Switzerland, July 15-17.

[3] OGC, 2002, OWS 1.2 Architecture Appendix D: OGC Interoperability Program Service Model (IPSM). Available at: http://ip.opengis.org/ows1.2/docs/020314_OWS1.2_Annex_B_appD.doc

[4] OGC, 2002, OpenGIS® Geography Markup Language (GML) Implementation Specification, version 2.1.1. Available at: http://www.opengis.net/gml/02-009/GML2-11.pdf

[5] Nolan, James J., Simon, Robert, Sood, Arun K., 2001, Developing an Ontology and ACL in an Agent-based GIS. In (Short Paper): *Proceedings of the Workshop on Ontologies in Agent Systems*, May 29, Montreal, QC, Canada. pp. 99-101.

[6] Bishr, Y., Pundt, H., Kuhn, W. and Radwan, M., 1999, Probing the concept of information communities - a first step toward semantic interoperability. In: Goodchild, M., Egenhofer, M., Fegeas, R., Kottman, C. (eds.), Interoperating Geographic Information Systems, pages 55-69. Kluwer Academic.

[7] Hakimpour F. & Timpf S., 2001, Using Ontologies for Resolution of Semantic Heterogeneity in GIS. Paper presented at the *4th AGILE Conference on Geographic Information Science* in Brno, April 19-21.

[8] DAML Website, 2002, at http://www.daml.org/about.html

[9] Gomez-Perez, A., Corcho, Oscar., 2002. IEEE Intelligent Systems vol. 17, no. 1, Jan/Feb, pp. 54-60.

[10] Schmelzer, R., Travis Vandersypen, Jason Bloomberg, Maddhu Siddalingaiah, Sam Hunting, Michael Qualls, Chad Darby, David Houlding, Diane Kennedy, 2002, *XML and Web Services Unleashed*. SAMS.

[11] Brittenham, P., Cubera, C., Ehnesbuke, D., Graham, S., 2002. Understanding WSDL in a UDDI registry part 1. IBM developer works. Available at : http://www-106.ibm.com/developerworks/library/ws-wsdl/

[12] Anupriya, A., Burstein, M., Hobbs, J.R., Lassila, O., Martin, D.L., McIlraith., S.A., Narayanan, S., Paolucci, M., Payne, T., Sycara, K., Zeng, H., 2002. DAML-S: Web Service Description for the Semantic Web. Presented at The First International Semantic Web Conference (ISWC). Available at: http://www.cs.cmu.edu/~terryp/Pubs/ISWC2002-DAMLS.pdf

[13] Suwanapong, S., Anutariya C., Wuvongse, V., 2002. *Proceedings of Working Conference on Engineering Information Systems in the Internet Context 2002*, September 25-27, Kanazawa, Japan, pp. 177-201.

[14] Paolucci, M., Kawamura, T., Payne, T.R., Sycara, K., 2002, Importing the Semantic Web in UDDI. Web Services, E-Business and Semantic Web Workshop, May 27-28, Toronto, Ontario, Canada.

[15] Martell, A., Atkinson, R. (ed.), 2002, OWS Service Registry 1 (OGC 02-30). OpenGIS Consortium. Available at: http://member.opengis.org/tc/archive/arch02/02-030.doc

[16] ESRI Place Finder Web Service overview, 2002. Available at: http://www.geographynetwork.com/services/arcwebonline/services/pf_overview.htm