Ontology based chaining of distributed Geographic Information Systems

Rob Lemmens

Department of Geo-information Processing
International Institute for Geo-information Science and Earth Observation (ITC)
P.O. Box 6, 7500 AA Enschede, The Netherlands
e-mail: lemmens@itc.nl

1. Geographic Information Systems as components

For the last decade, Geographic Information Systems (GIS) have provided planners and geo scientists with tools to analyse, maintain and present geo spatial information (information that is, in one way or the other, referenced to the earth surface). In the early days of GIS, its software systems were sold as monolithic systems. As the software became more mature, the systems were offered as modules containing a module with basic functionality and a variety of plug-in modules with extended functions. Main software producers came to realise that specific users who wanted to customise their systems needed a development environment with smaller system building blocks (components).

Today, a product like ESRI's ArcObjects provides the software elements to create an entire GIS. However these building blocks in themselves do not provide executable GIS analysis capabilities, they have to be assembled by a programmer. Unfortunately, these 'GIS objects' are of little use to the common GIS end-users whose interest is to apply certain common GIS processing functions to give solution to their geographic problems. GIS applications can be characterised by the wide variety of datasets (themes and data structure) and the often complex, but reusable operation-data chains. Many GIS applications, in particular in environments that require ad hoc queries, can greatly benefit from the use of interoperable components. To enable on demand component chaining we need data components and software components that are well defined and well described in terms of functionality, together with a user interface that facilitates the user-interpretation of these descriptions. Component-based applications have been around for some time, but their deployment in GIS is still in its infancy. This can be explained by the fact that GISs have to deal with complex (spatial) data types and software manufacturers tightly couple their functional parts with internal data structures.

2. Supporting data-operation connectivity, a multi-layer approach

In order to construct a component chain, users seek for *meaningful* combinations of data and process components. The term meaningful can be interpreted on different abstraction levels of connectivity between data and operation and depends on possible other requirements in the component chain. For example, suppose we want to calculate the shortest route between two house addresses and we make use of a chain of distributed operations. There can be different reasons why a typical GIS operation such as an address matcher[1], as first part of the chain, would not meaningfully operate on a certain address dataset. First the address matcher may use only street names (and no house numbers) as reference entities. Thus the geographic resolution is not appropriate for this component chain. Further the, address matcher may output the coordinates in a coordinate system that is unknown to the subsequent components of the chain. Generally speaking, we can distinguish three levels of abstraction, namely conceptual model, data structure and data format, where connectivity appears on all three levels. In this layered approach an address appears respectively as a concept (meaning of an address as interpreted by the information provider), its representation in a database as field(s) and the actual field values as output in a string or file.

In a more generalised geographic point of view, the address is a possible absolute location of a phenomenon as depicted in figure 1. In order to identify the connectivity between an operation and a dataset, we need descriptions on these different levels. Whether descriptions are needed on all levels depends on the context of the component chain. For example, if we would like to convert a dataset from one geographic coordinate system to the other, we do not need to know whether we deal with street features or houses (information at the conceptual level). A mediator identifies

[1] An address matcher finds the location coordinates (e.g. X,Y) of an address (street address with or without house number).

potential connectivity, based on dataset and operation descriptions, referred to as metadata (see figure 2).

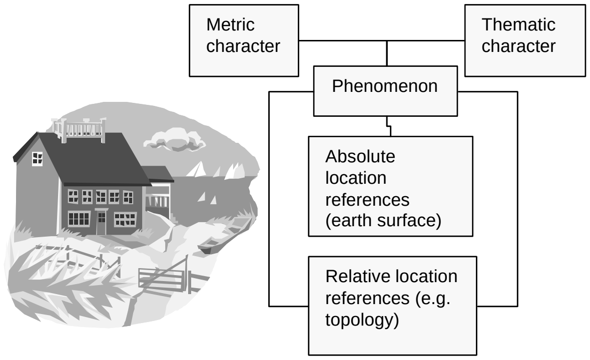

Figure 1. Generalised conceptual data model of a phenomenon in geographic space.

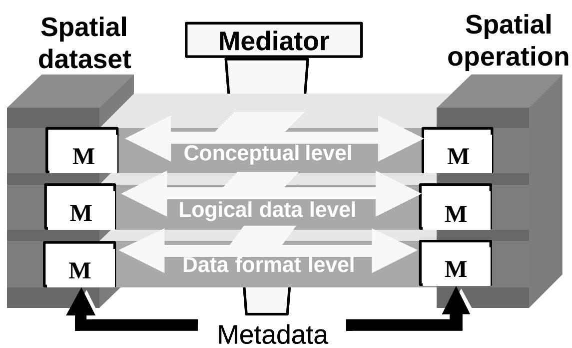

Figure 2. Connectivity layers and metadata, after [Lemmens *et al.*, 2003]

3. The role of geo ontologies

Descriptions of data and operations have to be measured against a reference frame of known artefacts and for the sake of automation such a reference frame must rely on machine processible information.

Reference framework topic	**Starting points**
Geographic coordinate systems	EPSG classification [EPSG, 2003]
Atomic and composite operations	ISO 19119 [ISO, 2002]
Location identifiers of geographic phenomena such as *address*	This research
Geodata structures	Geography Markup Language [OGC, 2002]
Thematic types of geodata (e.g. land use classification)	Domain specific taxonomies, e.g. CORINE land-cover classification [Bossard *et al.*, 2000]

Table 1. Reference frame topics for geo ontologies

Currently, the emerging Semantic Web provides several techniques to handle such reference frames with XML based ontologies. Table 1 indicates important reference frames for geo-information based processing that are partially existent, however not implemented yet in processible ontologies.

This research has initiated the creation and testing of a limited address ontology as partly depicted in figure 3.

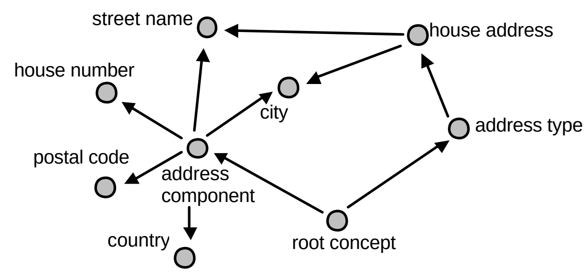

Figure 3. Graphic representation of a part of the *address* ontology

The address ontology is used in a natural disaster event scenario where multiple users need to identify the danger zone around their current location by providing an address. Depending on the kind of address they provide (e.g. with or without house number) a dedicated address matcher is selected. In the descriptions of the address matching components, the address ontology is referenced in RDF triples giving a conditional statement that clarifies which address type is used.

References

[Bossard *et al*., 2000] M. Bossard, J. Feranec and J. Otahel. *CORINE land cover technical guide– Addendum 2000* Technical report No 40 European Environment Agency Copenhagen Denmark. Available at http://www.eea.eu.int

[EPSG, 2003] European Petroleum Survey Group. *Geodesy Parameters Version 6.3 data model and data set*. February 2003. Available at http://www.epsg.org.

[ISO, 2002] *ISO Draft International Standard 19119 Geographic Information Services*. ISO, Geneva, Switzerland.

[Lemmens *et al.*, 2003] Lemmens, R., de Vries, M., Aditya, T. (2003). *Semantic Extension of Geo Web Service Descriptions with Ontology Languages.* In Proceedings of the 6th AGILE. April 24th – 26th, 2003, Lyon, France.

[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