Using Semantic Web Technologies to Improve Accessibility to SDIs

Inge La Rivière[1], Rob Lokers[2], Ale Ganzeveld[3], Arend Ligtenberg[4], Jandirk Bulens[5]
[1]Wageningen UR, Alterra, Centre for Geo-Information, inge.lariviere@wur.nl
[2]Wageningen UR, Alterra, Centre for Geo-Information, rob.lokers@wur.nl
[3]TNO, ale.ganzeveld@tno.nl
[4]Wageningen UR, Laboratory of Geo-information Science and Remote Sensing, arend.ligtenberg@wur.nl
[5]Wageningen UR, Alterra, Centre for Geo-Information, jandirk.bulens@wur.nl

Abstract

In this paper we will focus on the use of semantic web technologies to support easy, user-oriented access to SDIs, and on the specific issues that occur because of the spatial nature of data. A key-register for the soil and subsurface domain is being set up to provide access to soil and subsurface data to users with a wide variety in background and an equal large variety in vocabularies. The data models the data is stored in are defined in professional terminology.

An infrastructure was designed and developed, that provides a semantic shell on top of an OpenGIS based spatial data store for soil and subsurface information. A first version of a web application (BRON portal) has been developed to support (end) users to pose a query related to the soil and subsurface domain in the language of their own specific domain (e.g. environment, agriculture, hydrology) and to their level of expertise (e.g. professionals, policy-makers, inexperienced).

The semantic shell translates user questions to the underlying datasets terminology with the use of a domain ontology and relevant information is retrieved from a catalogue. The relevance of the retrieved information can be evaluated by the users based on the metadata and by previewing the spatial dataset (available as a service) in a map viewer.

In the future, the portal also needs to provide functionality for disclosing individual features of the datasets (as WFS). Querying these datasets by building an ontology guided query will be a challenge.

Keywords: ontology, SDI, search term, Semantic Web, soil, subsurface, thesaurus

1. Introduction

In The Netherlands, enforced by law a system of basic services are being set up that provide access to information obtained by public funding for Dutch citizens at marginal costs. This system consisting of key-registers with authentic data, authentication, one stop integral services etc. are in Dutch called "*Basisregistraties".*

One of these key-registers concerns soil and subsurface data. The soil and subsurface data has a spatial orientation. This key-register will contain a huge set of (2 and 3-dimensional) spatial data. The interface to this system will be based on OpenGIS or Open Geospatial Consortium (OGC) standards like Geography Mark-up Language Encoding Standard (GML), Web Feature Service Interface Standard (WFS), Web Map Service Interface Standard (WMS) and Catalogue Service for the Web (CSW).

Although the availability of a national register is an indispensable condition, it does not guarantee that everyone has easy access to the underlying information. Available data on soil and subsurface in The Netherlands is very diverse and widespread across organizations. It concerns a huge amount of raw and processed spatial data which is published through various information systems. In the near future these issues will be tackled by the setup of one national service for soil and subsurface data.

What remains is the fact that all this data is stored in data models that use professional terminology, which is in a lot of cases hardly comprehensible by end users. IT developers and specialists that make the data available to end users have their own vocabulary. The project "*Toegang tot de BRON*" - literally translated as "access to the source", but also referring to the name of the system (portal) for soil and subsurface data - attempts to bridge this gap.

The developed ontology serves as the knowledge base that drives translation of user queries to technical queries on the metadata of underlying data sources. In this process, the seamless handling of the spatial nature of the underlying data was an additional challenge.

The structure of the remaining paper is as follows: Section 2 describes the relevant concepts, section 3 contains a description of the use case, the results are discussed in section 4, and section 5 states some relevant future work.

2. Concepts

2.1 Semantic Web

The Semantic Web is a vision for the future of the Web in which information is given an, for computers, well-defined and explicit meaning, and context. This meaning is making it easier for machines to automatically process and integrate information available on the Web. The actual data is hidden in databases and files. Through a web application you have access to what designers and developers allow you to see. Some web applications are more advanced in gathering and combining data from other sources but only in the way they allow it. On so called 'mash-up' sites this gathering and combining of data offers a way of personalisation.

The original vision of the Semantic Web was by Tim Berners-Lee: "I have a dream for the Web [in which computers] become capable of analyzing all the data on the Web – the content, links, and transactions between people and computers. A 'Semantic Web', which should make this possible, has yet to emerge, but when it does, the day-to-day mechanisms of trade, bureaucracy and our daily lives will be handled by machines talking to machines. The *'intelligent agents'* people have touted for ages will finally materialize." (Berners-Lee, 1999). The Semantic Web used to be called Web 2.0. Nowadays Web 2.0 (sometimes called Social Web) is used for interactive web applications users can use to interact with one another or to contribute content. The Semantic Web is now denoted with Web 3.0.

The Semantic Web is moving from having just human-readable information to being a world-wide network of cooperating processes on computer-readable information, a collection of standard technologies to realize a Web of Data. It strives after facilitating intelligent searching, combining, and reasoning. An ontology deals with the concepts and techniques for formatting data in a kind of semantic mark-up to be understood by 'software agents', and to be accessible to humans through their natural language.

The shortest definition of an ontology is that of Gruber: “an explicit specification of a conceptualization” (Gruber, 1993). An ontology describes basic concepts in a domain and defines relations among them. The basic building blocks of an ontology design include: concepts (classes), properties of each concept describing various features and attributes of the concept (roles), restrictions on slots (facets or role restrictions).

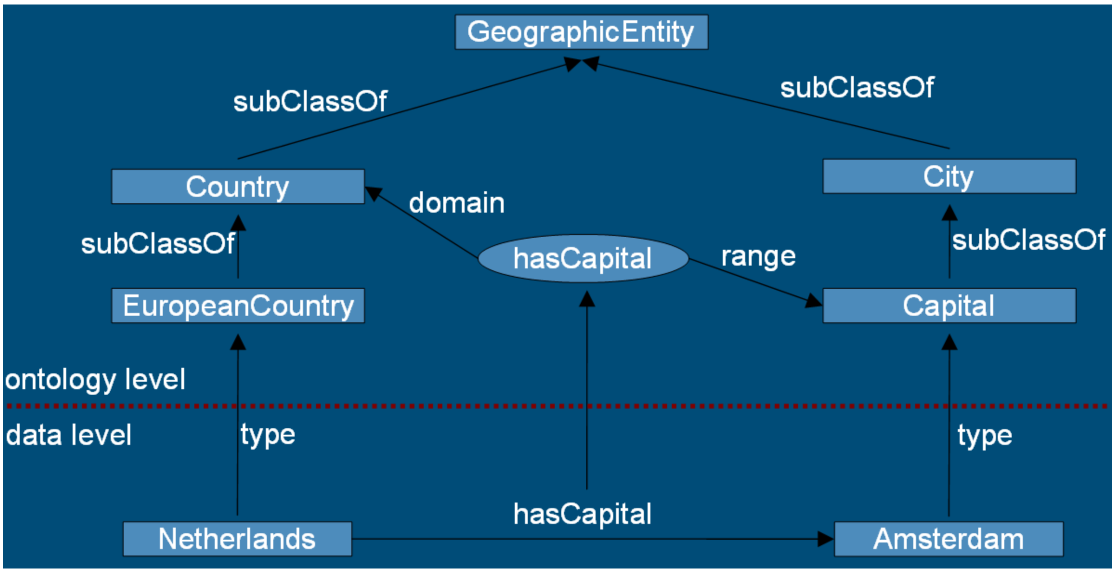
An ontology is intended to provide richer integration and interoperability of data among descriptive communities. An ontology together with a set of individual instances of classes form a knowledge base with the purpose to enable knowledge sharing and reuse. It not only defines a shared vocabulary but also it can bridge the gap of differences in vocabulary. Information is given an explicit meaning. By using a common vocabulary, through an ontology, interoperability will improve.

An ontology is defined by using The Web Ontology Language (OWL). OWL is one of the W3C (World Wide Web Consortium) recommendations related to the Semantic Web (McGuinness, 2004). The language is built upon the Resource Description Framework (RDF). RDF is a modelling language for representation of metadata and describing the semantics of information in a computer accessible and usable way. The value of information increases as this information becomes accessible to more users and more applications across the entire internet.

RDF is based on XML. XML defines the alphabet of valid characters (tags). XML Schema defines the rules for spelling, how can these characters be combined and nested to form valid structures or sentences. RDF is the grammar of semantics, how can we read these structures in meaningful sentences. OWL is richer and adds more vocabulary to RDF for describing properties, classes and relations.

RDF statements are formed by RDF triples consisting of subject, predicate (property) and object (value) and written as (subject, predicate, object). In the statement ‘Netherlands has capital Amsterdam’ is ‘Netherlands’ the subject. It has a relation ‘has capital’ with the object ‘Amsterdam’, as triple (Netherlands, hasCapital, Amsterdam), see figure 1. The statement describes a property of a resource. A resource is an object that can be identified by a URI, like a document, a webpage or a picture. It can also be a book or a person.

Figure 1: RDF Schema for ‘hasCapital’



Answering a (full) natural language question is a goal of the semantic web. A natural language interface doesn't require domain or system knowledge from the user. The learning curve for the user should be limited or nil. Interpreting the question semantically by the use of an ontology makes it possible to translate the question into a (keyword) search based on a common vocabulary. Because a natural language is not a formal language, the interpretation of the query sentence are troubled by ambiguities.

2.2 SDIs

A "Spatial Data Infrastructure" (SDI) is the relevant base collection of technologies, policies and institutional arrangements that facilitate the availability of and access to spatial data (Nebert, 2004). The SDI provides a basis for spatial data discovery, evaluation, and application (use). Users and providers are from within all levels of government, the commercial sector, the non-profit sector, academia and by citizens in general. This means that a SDI is used by users with a wide variety in expertise and domain knowledge.

The infrastructure of an SDI offers reliable access to hosted geographic data and means to discover, visualize, and evaluate the data through e.g. catalogues and Web mapping. The data collections need not to be part of the SDI but the provision of services is. The infrastructure provides the ideal environment to connect applications to data – influencing both data collection and applications construction through minimal appropriate standards and policies.

On different scales there are initiatives to develop SDIs. The initiative to achieve an European Spatial Data Infrastructure is INSPIRE. The initiative for a National Spatial Data Infrastructure (NSDI) in The Netherlands is the executive committee Geonovum.

INSPIRE means INfrastructure for Spatial InfoRmation in Europe. The main goals of INSPIRE are Establishing European Spatial Data Infrastructure and Exchange of spatial information between public services for the performance of public tasks with a direct or indirect impact on the environment. The target users of INSPIRE include policy-makers, planners and managers at European, national and local level and the citizens and their organisations.

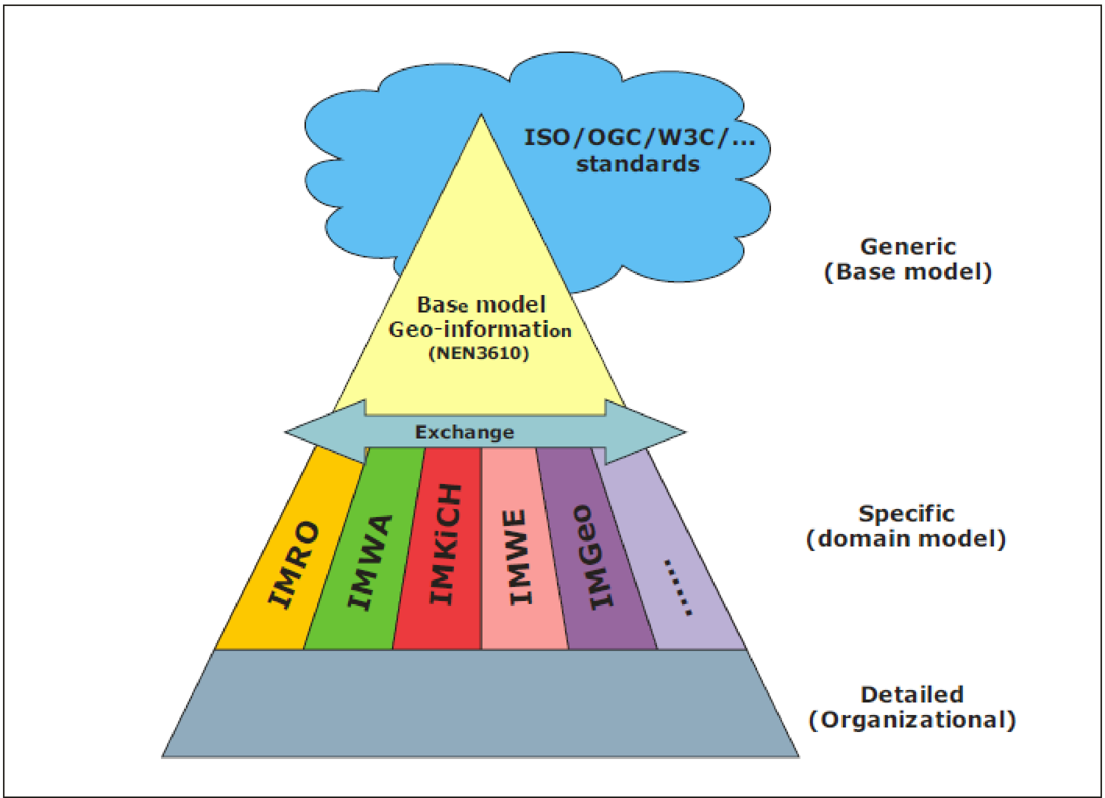
INSPIRE works on the interoperability and harmonisation of spatial data sets and services. The general situation on spatial information in Europe is one of fragmentation of datasets and sources, gaps in availability, lack of harmonisation between datasets at different geographical scales and duplication of information collection. These problems make it difficult to identify, access and use data that is available. INSPIRE strives after delivering integrated spatial information services to the users from a wide range of sources, from the local level to the global level, in an interoperable way.

Geonovum was founded in spring 2007 and devotes itself to providing better access to geo-information in the public sector. To implement this goal Geonovum develops a framework in which a set of necessary geo-standards, to be used in The Netherlands, are described. Geonovum also is involved in the implementation of INSPIRE in cooperation with the Netherlands Ministry of VROM (Ministry of Housing, Spatial Planning and the Environment).

Geonovum manages the Dutch standards and develops new ones in relation to international developments (International Organization for Standardization (ISO), OGC, W3C). Geonovum maintains the Base model for Geo-information (NEN3610) and the domain model for spatial planning (IMRO). They also provide support for the maintenance of other domain specific information models. An information model is the formal definition of a set objects, attributes and rules. It is an abstraction of the real world and schematically visualizes the concepts and definitions.

With the use of these and other standards it becomes more and more easy and efficient to discover, exchange and use geo-information. As described in the introduction, the Dutch Government is taking great efforts in creating and improving the Spatial Data Infrastructure by setting up a system of key-registers. Part of these contain spatial data, like the registers of addresses and buildings, the Large scale base map of The Netherlands (GBKN), and the register for soil and subsurface.

Figure 2: Information model pyramid (Bulens, 2007)



In the picture above (figure2), the 'pyramidal' relation between the international standards, the generic, specific and detailed information models is outlined. The domain specific information model for soil and subsurface could be entered on the dots, and carries the abbreviation IMBOD.

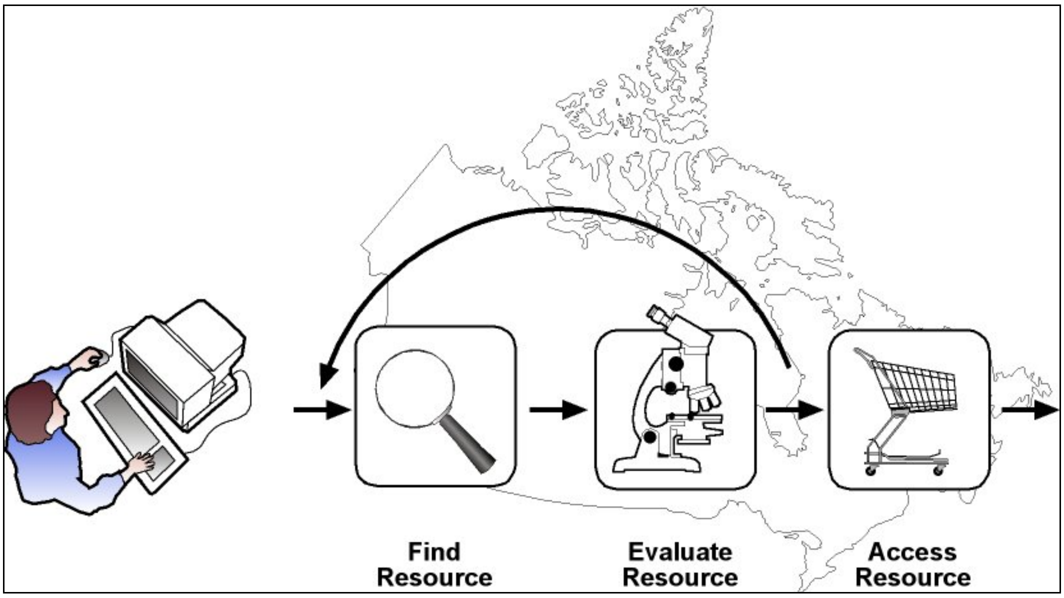
2.3 Metadata

Making the spatial information in a SDI identifiable and accessible is crucial. Otherwise all that valuable information is just sitting on the shelf, collecting dust, while waiting for someone to use it. Searching the information itself when looking for spatial information is not feasible. Instead metadata, 'data about other data', is used to describe resources in agreed and clearly defined attributes. These descriptions are gathered in catalogues.

The INSPIRE Metadata Implementing Rules support the three mechanisms of metadata as required by the INSPIRE Directive (INSPIRE, 2007), i.e. metadata needed for discovery, evaluation and usage, see also figure 3.

1. Metadata should make it easy to discover available spatial data.
2. Metadata makes it possible to evaluate the spatial data suitability for the purpose.
3. Metadata holds the information to know the conditions applicable to the use and access of spatial data.

Figure 3: Metadata mechanisms (GeoConnections Secretariat, 2004)



For the first mechanism keywords, classification and geographic location (geographic extent) are needed. When matching on a search action the dataset is 'found'. The second mechanism needs information about quality and validity of spatial data sets. Also the responsible authorities for creating, managing, maintaining and distributing the spatial data sets and services can be relevant for the potential user. Access and usage conditions for data sets and services are needed for the third.

Catalogues provide searchable repositories of information descriptions. The keyword based search doesn't solve the heterogeneity of natural languages and different domain backgrounds. With some background knowledge (and maybe some luck) some well chosen keywords may lead you to the jackpot of (geographically linked) information. The search can be put into a geographic context. The search is limited for a specific geographical area by extending it with a geographic search by extent.

The metadata keywords in the catalogue can be generated from the contents of the dataset. Also keywords can be abstracted from other parts in the metadata like title and abstract. The completeness and correctness of metadata is viable. Missing and misspelled keywords reduce the number of hits (false negatives). Incorrect keywords give false positive hits and bother the user with irrelevant information. The metadata standard in The Netherlands requires that the extent is specified.

The mechanism of Geo-tagging (Jones, 2007) can be used to generate geo-metadata. Geo-tagging consists of subsequently geo-parsing and geo-coding. Geo-parsing is described as recognising genuine geographic references (place names, addresses, post codes, phone codes) and ignoring non-geographic uses. Geo-parsing comprises of looking for patterns and context and detecting spatial propositions in natural language text. Place names can be recognised because they often start with a capital. Another way is the detection of a spatial propositions like in, near, south of, outside etc. for example, “clay *in* Utrecht”. A place ontology, which encodes knowledge of terminology and structure of geographic space, might add to the geo-parsing of a question.

Geo-coding couples the geographic location (e.g. coordinate, bounding box, polygon) to the geographic reference from the geo-parsed information. The geo-coded location is called a footprint. The geographic reference can be translated through locator and gazetteer functionality. Queries and documents can be parsed and coded in a similar way. For the document this is the document footprint and for query this is a query footprint.

A wider use of the geo-parsing and coding mechanisms would be to extend these principles to a wider range of toponyms. A toponym is a name that refers to local conditions in an area in the past. It can have a reference to geomorphologic conditions, soil, vegetation, land use or anthropogenic landscape characteristics (manmade objects). This is a more strict definition of a toponym. In geo-information it more often is used for place names and administrative areas. Geographically linked terms, like some soil types (e.g. loess) only occur in certain areas of The Netherlands (loess: Limburg). Names of places, areas and other toponyms are located at specific coordinates and could be geo-coded when suitable locator services exist to translate a toponym to its coordinate(s).

Currently, locators or gazetteers for addresses, zip codes, place names, countries and other geographically linked areas can be accessed through services to convert toponyms to geographical coordinates.

3 Use case – fitting it all together

3.1 Problem

The goal of the BRON portal is to accommodate (end) users with a way to pose a query related to the soil and subsurface domain in the language of their own specific domain (e.g. environment, agriculture, hydrology) and to their level of expertise (e.g. professionals, policy-makers, inexperienced). Often an end user will not be familiar with the expert soil and subsurface domain jargon. The portal must function as an intelligent searcher with some kind of semantic wrapper to lead the user to the intended information.

3.2 Workflow

To deal with the wide variety of vocabularies a workflow was designed. The workflow for processing the user’s natural language question consists of three steps. For the first version of the BRON portal the question will be restricted to the interpretation of one single search term. This term can consist of more than one word. This approach is supported by analysis of web queries. The average number of words used in a web query is just above 2, and about 30% of the queries consist of 1 word. Also a lot of queries counted as multiple words contain correlated words and are actually single term queries (Spink, 2001). The goal is to translate this single term from the user’s vocabulary to the common BRON vocabulary.

First, the validation of the entered word must be performed by checking of the existence of the word in the thesaurus. In the workflow the use of a wildcard will be allowed. In the case of wildcards a list of matching terms needs to be retrieved from the thesaurus from which the user can make a choice. Only terms that exist in the thesaurus (validated terms) can be used in the further steps.

In the second step the validated term will be translated through the ontology (ontology mapping service). In the ontology, terms, synonym terms and related terms are described. The type of related terms are broader, narrower, 'strongly associated' and 'weakly associated'. The resulting ontology terms, with their relation and description, need to be presented to the user

In the user guided process the user has the possibility to choose up to ten ontology terms to be used. Alternatively up to ten additional terms can be entered manually. At least one term must be provided. Also it should be possible for the user to define the geographic extent which optionally could be taken into account. The provided information is used in the third step.

The third step is the retrieval of references to matching datasets. The matching will be performed by the use of a (keyword) search through a catalogue service (metadata discovery mechanism). The dataset references contain metadata that provide the user with the information to evaluate the resource suitability (metadata evaluation mechanism). Previewing the spatial dataset in a map viewer supports this process. To complete the search process, the document or dataset can be retrieved (metadata use mechanism).

3.3 Implementation

To start with, an analysis of the professional terminology in various sub domains was performed and existing thesauri from different organisations were harmonized. Using several distributed thesauri was considered but the organisation of the maintenance was not feasible. The harmonisation resulted in the central BRON thesaurus, covering terminology of the soil and subsurface domain.

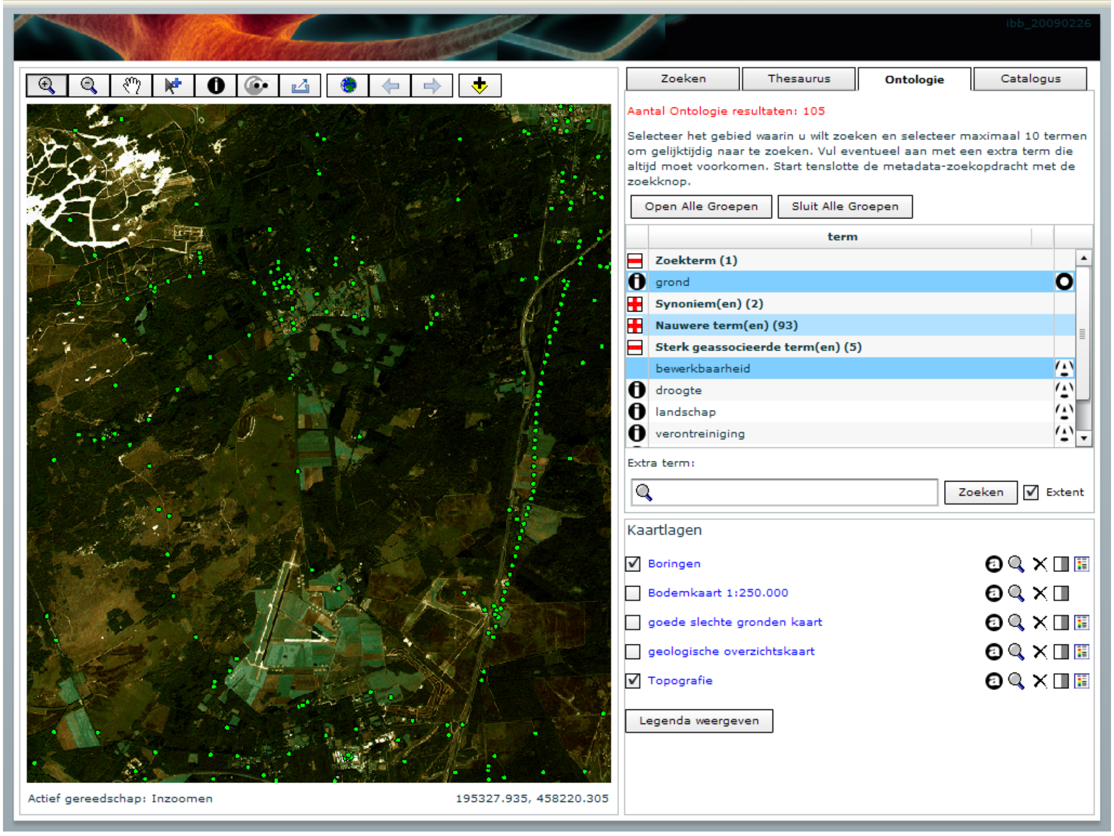
Subsequently, an ontology was developed that enriches the terminology of the thesaurus with knowledge and relationships required to support translation of terminology between user domains and the professional soil and subsurface domain. Protégé was used as a tool for editing and visualisation. It is a free, open source ontology editor and knowledge-base framework. It is considered the de facto standard for the development of knowledge systems

A semantic shell was developed that provides an interface between the OpenGIS based soil and subsurface SDI and the end user. It is a rich internet application developed using the Adobe Flex platform. Adobe Flex is platform independent and is quite suitable to develop highly interactive internet GIS applications as Flex offers a rich set of user interface components.

For the translation of the search term from the user domain to the soil and subsurface domain we use semantic web technologies. The ontology mapping step is used to obtain the relevant keywords that will lead to the soil and subsurface or geographic information as described in the catalogue. Normally the user will not be fully aware of this mapping. In the BRON portal all steps performed are intentionally made visible and the user can influence each step of the validation, mapping/translation, discovering/evaluating and obtaining process.

In the interface two main components are visible, see figure 4. One component is the search component and the other is a map viewer. The map viewer is used to define the extent for the search and to preview the spatial datasets. Several background and orientation layers are preloaded in the map viewer based on a configuration file. Also for specialist users there is functionality to add their own. There is also a smaller component to control the presentation of the layers in the map viewer. The map viewer used in the BRON portal is called Luigi (Vanmeulebrouk, 2008).

Figure 4: BRON web application



Luigi is a user-friendly internet GIS application framework and is developed using the Adobe Flex development framework. It is one of the spin-offs of the Geoloketten project, which supports the use of geographical information in The Netherlands. During the Geoloketten project not all the necessary tools were not available or sufficient to demonstrate the geo-spatial enabled web-services.

The Luigi framework currently supports map services such as WMS, WFS and ArcIMS. Also catalogue services, geo-coding services, a coordinate transformation service and charting services are supported. The usual pan, zoom and identify tools are at the users disposal. Because of the modular design of the framework, adding new functionality or support for additional data formats is easy for Flex developers.

The search components supports the three steps through four tabs in the navigator. After entering a search term in the first tab, a web service is accessed for validating the search term or retrieving a list of matching terms (step 1). In case of a wildcard search, the list of matching terms will be shown in the second tab and the user can make a choice from them. The validated or chosen term is used in the translation through the ontology (step 2). The OWL file produced by Protégé is queried through the JENA-api and accessed from the client application as a web service.

The results from the ontology mapping service are presented in the third tab. The user can choose up to ten from them. Also up to ten additional words can be entered. The map viewer can be used to zoom in to the relevant extent. A checkbox can be ticked to indicate that the extent needs to be used for the further search. The ontology terms are used in a logical disjunction (OR relation) with a logical conjunction (AND relation) for the optional extent and the additional supplied words. All metadata text fields in the catalogue are searched and the metadata's bounding box is used for the extent (step 3). The results of the catalogue are presented in the fourth tab.

The catalogue server used for the BRON portal is the eXcat CSW server developed by Rob van Swol of the National Aerospace Laboratory (NLR). This is also a spin-off of the Geoloketten project (Swol, 2008). eXcat supports the OGC CSW versions 2.0.0, 2.0.1 and 2.0.2. Luigi provides the client-side catalogue query services. In the fourth tab, one can apply the metadata evaluation mechanism based on the full metadata to decide if the discovered information is the right information for the intended purpose. The BRON datasets (as WMS and WFS) can be (pre)viewed in the Luigi map viewer by pressing the 'add' button.

The WMS and WFS services can be used in compliant tools because the URL of the service is presented in the metadata. Also an identify can be performed on a feature in the map viewer. The feature information will become available for downloading as GML.

4. Observations and discussion

Adding a semantic layer on top of datasets that are searched by users from the same domain provides little or no added-value. Often, the users are experienced in finding data, especially when they know for certain that 'it must be there somewhere'. On top of that they use the same language as the specialists that supplied the data. Ambiguities are either solved in the software (in the middle tier) or by workarounds by the experienced user.

People unfamiliar with the soil and subsurface domain find the semantic layer helpful. A crucial dependency is the completeness of the underlying ontology with respect to the contents (the number of terms) as well as its structure (the number and kind of the relations between the terms). In the ontology only basic relations are used. Supporting other relations, like 'consists of' or 'is made of' and adding the appropriate reasoning will make it more powerful but also requires extending the interface to facilitate it. A more static snapshot of the ontology in the form of a guided categorized search would be helpful as well, though less dynamic.

Another point observed during demonstrations and training sessions, is that the current combination of text and the extent of the map in the current interface is not useful. Users tend to enter only textual search criteria, of which some are geographical. Also the current datasets all have a bounding box that covers the full extent of The Netherlands and the datasets can't be distinguished on that aspect. The map only becomes important when the catalogue results are evaluated in the map as a layer. Offering support for translating domain specific and generic toponyms to geographic regions might me worth coming.

A third observation is that a better understanding of the metadata attached to a dataset has emerged. The keywords in the metadata catalogue describing each dataset were entered manually by domain specialists. Preferably, all keywords in the catalogue should exist in the ontology. A periodical check on the keywords could be performed to optimize the coupling between the ontology and the catalogue.

Of course in the interface it is possible to manually enter some keywords to overcome this. These terms can also be used to limit the number of results, when the number of hits grow. The use of more ontology terms widen the search. With the metadata for the current small number of datasets this is not an issue. When the number (from this same domain) increases, it will not become easy to distinguish one dataset from another. Future work needs to address some kind of ranking based on the calculation of relevance in relation to the ontology results.

Not only keywords but also other text fields in the metadata, like title and abstract are currently being searched for the existence of the mapped term. This leads also to false positive results because the term isn't used in its exact form. The Dutch word for clay is 'klei' and this word will occur as a common part of other words in text fields. A solution for this is to perform a search on the keywords only. By identifying all the relevant words and terms from the title, abstract and other text fields and adding them as keywords, there will be no need to search the other text fields.

The current version is an ontology guided search on a catalogue to help users in finding relevant datasets. The way the different steps are presented is confusing. Insufficient explanation about the search steps is provided to the user and the interface hasn't proven to be intuitive. The disclosure on the performed steps of the search process in the interface is intended for demonstration purposes. Also it gives domain specialists the means to see, validate and influence the underlying process. An expert review followed by an eye tracking usability test aims to enhance the application, making it accessible for a multitude of users, like policy-makers and engineers. This future development is surely backed by the enthusiastic response the application has generated so far, as the possibilities of the BRON portal go beyond the expectations of (end) users.

Logging of a user's search path is interesting for analysing which search terms did, and which did not, lead to a successful catalogue result. All the terms (entered, chosen and added) and the extent should be part of this logging. A log serial is completed by stopping the application (user gave up?) or an action on a catalogue result (presumed to be a successful find). Not for all catalogue results an action is possible The failing search terms could be evaluated as candidates for the enrichment of the thesaurus, ontology or catalogue.

5. Future developments

Using a semantic wrapper to help a user to bridge the gap between the user's and the domain vocabulary can be useful. In the future, the portal also needs to provide functionality for disclosing individual features of the IMBOD datasets (as WFS). Querying the IMBOD datasets by building an ontology guided query will be a challenge. The Luigi map viewer can provide added-value in this process.

The following observed and discussed issues need to be addressed in future work to solve the current implementation problems:

- extend number of search terms (to natural language sentence);
- focus on ontology completeness (terms, relations, reasoning);
- deal with ontology catalogue keyword coupling and ranking;

- search path logging for analysis of failing and successful search terms;
- enhance user interaction design aspect of the application.

REFERENCES

Berners-Lee, T. and M. Fischetti (1999). *Weaving the Web: Origins and Future of the World Wide Web*, San Francisco: Harper.

Bulens, J., Vullings, W., Crompvoets, J. and M. Reuvers (2007). The framework of standards for the Dutch SDI, *Proceedings 10th AGILE International Conference on Geographic Information Science, 2007, Aalborg University, Denmark,* at: http://people.plan.aau.dk/~enc/AGILE2007/PDF/146_PDF.pdf [acessed: 5 March 2009].

GeoConnections Secretariat (2004) , A Developers' Guide to the CGDI: Developing and publishing geographic information, data and associated services. at: http://www.geoconnections.org/publications/Technical_Manual/CGDI_Technical_Manual_0204_e.pdf. [accessed: 27 March 2009].

Gruber, T. R. (1993). A Translation Approach to Portable Ontology Specifications. *Knowledge Acquisition*, 5(2):199-220.

INSPIRE (2007), Directive 2007/2/EC of the European parliament and of the council of 14 March 2007: establishing an Infrastructure for spatial information in the European Community (INSPIRE), http://inspire.jrc.ec.europa.eu/directive/l_10820070425en00010014.pdf. [accessed: 27 March 2009].

Jones, C. (2007), Geographical Web Search Engines and Geographical Information Retrieval (GIR). *European GeoInformatics Workshop Edinburgh 2007*: at http://www.nesc.ac.uk/action/esi/download.cfm?index=3408. [accessed: 11 March 2009].

McGuinness, D. L. and F. van Harmelen (2004), OWL Web Ontology Language, *W3C Recommendation*. http://www.w3.org/TR/owl-features/, [accessed 5 March 2009].

Nebert, D. D. (Ed) (2004), Developing Spatial Data Infrastructures: The SDI Cookbook, Version 2.0, Technical Working Group, GSDI, at http://www.gsdi.org/docs2004/Cookbook/cookbookV2.0.pdf, [accessed 5 March 2009].

Spink, A., Wolfram, D., Jansen, B.J. and T. Saracevic (2001). Searching the web: The public and their queries. *Journal of the American Society for Information Science*, 53(2): 226-234.

Swol, R.W. van (2008), eXcat OpenGIS CSW server en clients - Handleiding, at http://gdsc.nlr.nl/gdsc/sites/default/files/publications/2008/2008-809-cr.pdf, [accessed 19 March 2009].

Vanmeulebrouk, B. and M. Kuyper (2008), Integration of geo-spatial web services using Adobe Flex. *FOSS4G 2008*. at: http://www.foss4g.org/index.php/foss4g/2008/paper/view/125. [accessed: 9 March 2009].