FROM HYBRID MAPPING TO INTEGRATED QUERY AND PROCESSING: TWO SDI CASES FROM THE NETHERLANDS

Marian de Vries and Theo Tijssen

GIS Technology, Research Institute for Housing, Urban and Mobility Studies/
Faculty of Technology, Policy and Management, Delft University of Technology
The Netherlands

ABSTRACT

Consistent and 'transparent' combination of geo-information from different data sources, that each have their own existing data model, is not unproblematic. For shared use of information in a (national or European) Spatial Data Infrastructure, a certain degree of data model harmonisation often seems necessary. However, the degree of harmonisation and also the techniques that can be used, are not fixed.

Luzet (*2003*) suggests using the term 'interoperability' when discussing SDI developments and requirements, instead of 'harmonisation', to show the wide spectrum of options. He distinguishes two main levels of interoperability to be accomplished in the case of SDI projects: Hybrid Mapping and Consistent Mapping.

In this paper we will report on two recent SDI-like interoperability initiatives in the Netherlands. In both cases a certain degree of harmonisation is considered necessary, but the requirements and technology used are different.

The first project (Web access to 'seamless' maps of regional planning geo-information of four Dutch provinces) is a case of 'same thematic content, other area'. The second project (a harmonised information model for the cultural heritage domain) is an example of 'same area, partly overlapping content'.

1. INTRODUCTION

Geo-data (like other data) is stored according to a certain conceptual view of that part of (geographic) reality that is considered relevant to the business processes within an organization. The purpose of collecting or creating (digitizing) that specific set of geo-data will influence modeling decisions. As a consequence also the actual data models (the database schemes) will differ from one application to another: names of tables and attributes, granularity (many object types with few attributes, or few object types with many attributes), domain values, etc. (e.g. *Hart, 2003*).

Apart from differences in data structure also differences in information semantics will stand in the way of unproblematic multi-source data integration in a SDI (see *GINIE, 2003*).

The heterogeneity of the data sources can result in unsuccessful queries and, in the worst case, can lead to wrong interpretations.

For successful combination of geo-information from multiple sources it is necessary to have:
- unambiguous metadata about geo-information resources and (Web) services;
- consistent visualisation i.e. the same cartographic representation (colors, line width, symbology) for 'things' (objects) on the map that are the same;
- integrated query and selection possibilities and transparency in case of spatial and thematic analysis of the geo-information content (*Kap, 2004*).

SDI initiatives can aim at different ambition levels, and depending on the goals of a specific project, data model harmonisation will be a part of it, or will get a low priority (*Luzet, 2003*; *Riecken, 2003*). Apart from the ambition level, also the methods and techniques used for harmonisation can vary, and in many cases the data structure of the stored geo-information can be left intact (*Kap, 2004*).

In this paper we will report on two recent SDI-like interoperability initiatives in the Netherlands. In both cases a certain degree of harmonisation is considered necessary, but the requirements and technology used are different.
The first project (Web access to ‘seamless’ maps of regional planning geo-information of four Dutch provinces) is a case of 'same thematic content, other area'. The second project (a harmonised information model for the cultural heritage domain) is an example of 'same area, partly overlapping content'.
In these ‘all-Dutch’ pilots there are no differences in natural language and in spatial reference system (the multilingual and co-ordinate system aspects of setting up a SDI). The focus is on the information integration aspect, i.e. the combination of data sources from geo-data providers that all have their own 'legacy' data models and their own map classifications.
In both projects the thematic content of the geo-information is more important than the seamless combination of the geometry itself. Questions like: geodetic quality of the geometric data, precision and scale will not be addressed.

2. CONSISTENT MAP CLASSIFICATION: THE STREEKPLAN PILOT

The first project is a pilot between four Dutch provinces to publish geo-information on regional (urban and rural) planning ('Streekplannen') via the Web using OpenGIS compliant Web services. The provinces are: Gelderland, Brabant, Limburg and Overijssel. The project can be considered as a spin-off project of the X-Border GDI pilot between the border provinces of the Netherlands and North-Rhine Westphalia (Germany) (Interreg IIIA, also see *Riecken, 2003*).
One of the requirements in the 'Streekplan' pilot is that the efforts for the provinces to make their data available via Web services should be as little as possible (without extensive conversions). Nevertheless there should be an integrated view presented to the end-users, with consistent names for map layers and with consistent classification schemas (legendas). This is important because the 'Streekplan' information is used by e.g. municipalities as input for their long-term regional planning decisions.
Main issue in this pilot is therefore how to obtain uniform map classifications in the combined maps of the four provinces, without having to change the data structure of the geo-data that is put online.

2.1. Hybrid Mapping

First step in the pilot was to make the regional planning data of the four provinces accessible via Web services. Four separate Web services were set up, in this phase at the server side all with the same software, i.e. ESRI's ArcIMS with WMS connector. The WMS connector acts like a wrapper around the proprietary ArcIMS core, which makes the ArcIMS service 'behave' like an OpenGIS WMS service.
At the client side we used three clients for access to the WMS services: a simple Web client (based on HTML and JavaScript) developed for testing at TU Delft, the Intergraph WMS viewer that is available on the Web (http://www.wmsviewer.com), and ESRI's ArcMap with WMS extension.

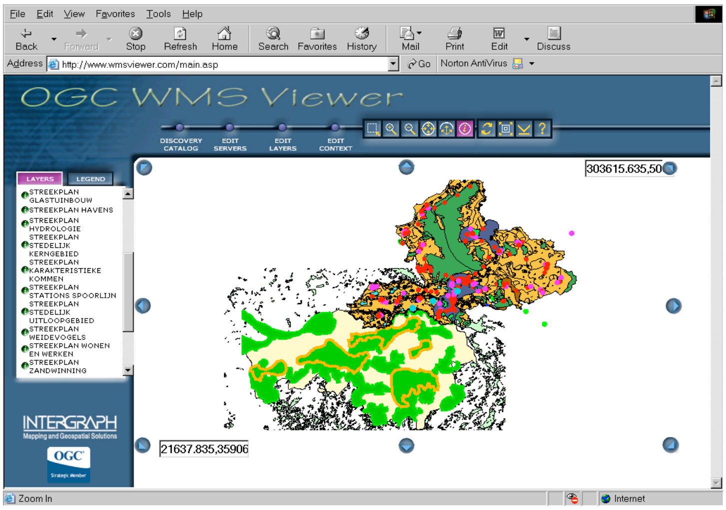

Figure 1: Hybrid Mapping.

Figure 1 shows the Intergraph WMS viewer with the combined output (jpeg or png) of two different sources: the WMS service with Streekplan data of Gelderland and the WMS service with (part of) the Streekplan data of Brabant. What is clear from this screenshot is that although the combination of output from different Web services in one client (in real-time) is technically possible, there is no seamless visualisation and uniform classification.
We could call the end product of this first step a case of Hybrid Mapping.

2.2. Web Map Services with Styled Layer Descriptor

To offer end-users a consistent view on the regional planning geo-information of the four provinces a second step is necessary: WMS services with a uniform map classification will have to be set up.
One of the techniques that can be used for harmonisation of the cartographic aspects (visualisation and classification) is based on the OpenGIS Styled Layer Descriptor specification (*SLD, 2002*). The SLD specification was created for use in a WMS context, to allow users to request other display styles per map layer than the default ones. SLD can also be used however by the data providers themselves (at the server side of the Web Map Service architecture), to fine-tune styles within a map layer. This can be accomplished by using conditional statements based on the OpenGIS Filter Encoding specification (*OGC, 2001*).

Below is an example of a styling rule with Filter condition in SLD notation.

```
<sld:Rule>
  <sld:Name>weg-in-ontwerp</sld:Name>
  <sld:Title>Weg in ontwerp</sld:Title>
  <ogc:Filter>
    <ogc:Or>
      <ogc:PropertyIsEqualTo>
        <ogc:PropertyName>brab:Status</ogc:PropertyName>
        <ogc:Literal>gebiedsontsluitingsweg tracé vastgesteld</ogc:Literal>
      </ogc:PropertyIsEqualTo>
      <ogc:PropertyIsEqualTo>
        <ogc:PropertyName>brab:Status</ogc:PropertyName>
        <ogc:Literal>stroomweg tracé vastgesteld</ogc:Literal>
      </ogc:PropertyIsEqualTo>
    </ogc:Or>
  </ogc:Filter>
  <sld:LineSymbolizer>
   ...
  </sld:LineSymbolizer>
</sld:Rule>
```

And for another province in the pilot, the Filter condition to create the harmonized cartographic style and legend would look like this:

```
  <ogc:Filter>
    <ogc:Or>
      <ogc:PropertyIsEqualTo>
        <ogc:PropertyName>gld:Lijnsoort</ogc:PropertyName>
        <ogc:Literal>2</ogc:Literal>
      </ogc:PropertyIsEqualTo>
      <ogc:PropertyIsEqualTo>
        <ogc:PropertyName>gld:Lijnsoort</ogc:PropertyName>
        <ogc:Literal>5</ogc:Literal>
      </ogc:PropertyIsEqualTo>
    </ogc:Or>
  </ogc:Filter>
```

The Filter part of SLD makes it a good technique for map classification harmonisation. Both the classification items, their labels and ranges, and the colors and symbology for each feature object in the map can be specified. This can be handled at the server, before the data is transformed into the map images that are sent to the Web clients. This way it is possible to create map classifications that are consistent between data providers. The data structure itself can remain intact, the (cartographic) harmonisation is accomplished during retrieval and rendering.

3. A COMMON INFORMATION MODEL: THE IMKICH PROJECT

The second SDI project we present in this paper is an example of 'same area, partly overlapping content'. In this project, called IMKICH (Information Model Knowledge Infrastructure Cultural History), the aim is to harmonize the existing data models of a number of organisations involved in cultural heritage (archaeology, monuments, landscape patterns) into an overall information model.
In the IMKICH project one of the requirements is to provide integrated query possibilities on the different data sources in a multi-source Web client, i.e. selection queries and spatial and thematic analysis possibilities. Where the data originates (which data provider) should be as transparent as possible to the end-user. It must, in other words, be possible to pose a question like “select all 18th century buildings on the topographic map that are also official national monuments”, where the geo-information to answer this query is stored in separate data sources.
Because of this requirement two issues have to be solved: differences in implemented (‘legacy’) data structure, and differences in the semantics of the information (attribute content).
There are also other requirements: the new overall data model should be compatible with the new TOPNL data model of the Dutch NMA (Topografische Dienst Kadaster), and with the rules for content and structure of geo-data sets of the Dutch Normalisation Institute (NEN3610) (see e.g. *Bulens, 2004*).

3.1. IMKICH data sources

To give an idea of some of the harmonisation issues in the IMKICH project, we give a short impression of each of the data sources.

The first data source is called ‘Monuments’. This data set contains all objects that have the status of ‘national monument’ in the Netherlands. Most of these monuments are buildings (churches, chapels, 17th century houses, city halls etc.), but also 'complexes' (country houses with stables, gardens), and other kinds of artifacts that are considered of cultural-historical importance (waterways, bridges, dock works, staircases, statues). They are monuments in the sense that they are officially registered as such. Changes to these monuments or parts of monuments are restricted and can only be carried out after procedures to get a permit. For urban and rural planning (zoning plans) it is therefore very important to know where these monuments are located.
What kind of (physical) object the monument is, can be seen in a free text attribute. The values are at the moment a mix of type names ('castle', 'military compound') and individual names ('castle so-and-so').
All instances in the ‘Monuments’ data set are national monuments. Therefore, in this data set, there is no need for an attribute 'legal status'.

The second data set, maintained by another government agency, is called ‘Archaeological Monuments’.
This data set contains all archaeological sites in the Netherlands. Apart from sites that have been assessed as having archaeological importance, the data set also contains sites of which the importance still has to be established. This can for example be a forest clearance where an ancient burial ground is suspected.

There is one attribute with the name 'waarde' (= 'value') that is actually a combination of three properties:
- whether or not the importance of the site is assessed (procedure);
- after it is assessed, the archaeological importance ('important', 'high importance', 'very high importance');
- whether or not it is an official national monument (legal status).

Only a small part of the archaeological sites have the legal status of national monument. The meaning of the word 'monument' in the ‘Archaeological Monuments’ data set is therefore another than in the ‘Monuments’ data source (see above).
The kind of physical object(s) the archaeological site contains (or maybe contains) is stored in a free text attribute. As with the ‘national monuments’ data set, also in this case there is no code list (e.g. based on an UML enumeration type) that contains the permitted values for the attribute.

The third data source, maintained by yet another agency, has to do with landscape patterns and landscape elements that are considered of cultural-historical importance. This data set contains historically interesting objects in rural areas, e.g. characteristic village settlement patterns, evidence of certain agricultural systems, clearances, enclosures, but also water mills, canals or little streams. These objects do not have a legal status as monument, but have a certain cultural-historical importance according to the experts that maintain the data set.
To designate the importance or relevance of a landscape object there is an attribute called ‘kenmerkendheid’ (i.e. ‘how characteristic is this object’. The domain values for this attribute are: 'provincial', '(inter)national', or the attribute is left empty (has a null value).

3.2. Creating the core model

The first step was to explore the existing data models of the different cultural-history data sources (see the impression in the previous section).
We did this by:
- talking to domain experts;
- analysing the data structure (the database schemas: tables/files, attributes, domain values);
- looking at the actual content (explore the datasets);
- looking at the cartographic products and end applications (legends and map classifications, labels, links to document sources).

After this preliminary phase, aimed at understanding the current models, the second step was to establish the commonalities between the different models.
At the lowest level of (thematic) detail the data sources appeared to have much overlap: in the attribute for the type of object (‘chapel’, ‘wind mill’, ‘old road’) we often see the same domain values.
Nevertheless is was not possible to ‘merge’ the different feature types into one combined object class, mostly because of the fact that different properties were important in the different data sources. For example, for a (legal) ‘Monument’ the exact postal address is essential, together with other administrative attributes (cadastral parcel numbers and details about ownership). For an historical landscape element without a legal status as monument (e.g. a 15th century canal or village settlement pattern) the exact address does not exist (and is not important, because it is not an official monument).
Therefore we created one (abstract) super class to hold the properties (attributes and relationships) that the feature types in the different data sets have in common. The feature

types that could be distinguished were made into subclasses of this super class. The super class has all the common properties; the subclasses inherit these properties and have additional properties of their own.

Deciding about the common 'core' properties for all feature types was not easy (c.f. *GiMoDig, 2004*). If the set of common properties is too small, there is hardly any gain in the new core model. If the set is too large, the new data sets will have many attributes with either ‘dummy’ default values, or null values.

3.3. Implementing the core model

The third step in the project was to create translation rules from the existing data models into the new core model. Sometimes this involved splitting one feature type (UML object class) into two or more, based on values for one or more distinctive properties. It could also mean combining two object classes into one. To avoid semantic ambiguity some of the names of tables and attributes were changed (e.g. archis:ArchaeologicalMonument became imkich:ArchaeologicalSite).
To evaluate the adequacy of the new integrated information model, and to test the translation rules, a number of prototype data sets were created. These data sets are now tested by stakeholders (potential data receivers) and by the three data providers involved in the project.

In this project we used database views to remodel the existing data sources into the new model. Of course this is only possible when the geo-information is stored in (object-)relational databases like Oracle Spatial (*Oracle, 2004*), PostGIS (*2004*) or MySQL (*2004*).
In other situations (file-based GIS or CAD systems) this is not feasible; in that case new files, that conform to the common model, have to be created by the data provider, as derived data sets, especially for data exchange. This makes harmonisation at the server side (and at the data layer) maybe less attractive. A solution would be to have some automated process that takes care of replication and synchronization of the production data sets (for internal use) and the data sets that are meant for exchange with the outside world.
Another strategy is to have a mediator layer that is equipped with translation components from ‘local’ to ‘global’ model. This is also the approach taken at the GiMoDig project (*Lehto, 2004*).

4. CONCLUSIONS AND FUTURE RESEARCH

Both projects discussed in this paper can serve as proof of concept for the idea that 'strict' harmonisation (of the source data, at the data storage level) is often not necessary, because interoperability can also be achieved in other ways.
In the Streekplan pilot the original aim was to make the regional planning information of the four provinces available via Web Map services. Now, in an additional step, the intention is to harmonize the map styles and legends.
In the IMKICH project the aim goes beyond this: also integrated querying (selection and analysis) should be possible. In this case a harmonized ‘core’ information model is necessary.
This does not mean however that the original data structure in the production databases of the contributing data providers has to be changed. We showed that, by applying database views, the core model can be implemented in a ‘virtual’ way.

For the construction of the core model it was important to look at the actual content of the data (the thematic attribute values) as a way to understand the structure and the (implicit) conceptual model. Parallel to this exploration of the geo-data content we talked to developers to get insight in the application domain.
Combining the information from these different activities we could 'reverse engineer' the conceptual data models of the organisations. The data structure (which feature types, which attributes and relations between feature types) was of course relatively easy to establish. It was not so easy however to capture the implicit semantics: which 'things' are relevant in the cultural heritage domain, how can these 'things' be categorized/grouped into sets, e.g. in one or more hierarchical taxonomies, and what semantic conflicts exist currently between the data sets.
For integrated query and processing of the combined data sources it is precisely this aspect that is important. In this project we made only a start to make the semantics of the existing data models more explicit. Semantic ambiguity in domain values for example can only be discovered by intensive use of the combined data sources and by feedback of end-users and domain experts.
Causes for semantic ambiguity ('same terms, other meaning', 'same meaning, other terms') have to be discovered and dealt with. This may lead to an adaptation in the permitted values for (thematic) attributes. It can also be decided however to keep the content 'as is' and to construct translation rules based on semantic relations that are detected between the different models.
Future research will concentrate on this aspect. We will use techniques from the Semantic Web, especially the ontology language OWL (Web Ontology Language). OWL can be used to specify semantic relations between data models, and this would enable semantic reasoning during data retrieval (e.g. 'query expansion' using synonyms and hyponyms).
On-the-fly harmonisation has also been tested in the GiMoDig project (*Lehto, 2004*), where XSLT is used to translate from 'local' to 'global' model and back. It is an interesting research question to see whether or not tools from the Semantic Web, like ontology reasoners, can be used for this same purpose.

ACKNOWLEDGEMENTS
This publication is the result of the research program Sustainable Urban Areas (SUA) carried out by Delft University of Technology. The research was partially funded by the Province of Gelderland.

REFERENCES

Bulens, J. and W. Vullings (2004): National Spatial Information Models, 7th AGILE Conference, April 2004, Heraklion, Greece.
GINIE (2003): New Issues for the GI Research and Technology Development Agenda (D 4.3.1), Martin Klopfer (ed.), October 2003.
Hart, G. and J. Greenwood (2003): A component based approach to geo-ontologies and geodata modelling to enable data sharing, in: Proceedings of the 6th AGILE conference, April, Lyon, France.
Kap, A., B. van Loenen and M. de Vries (2004): Harmonized Access to Heterogeneous Content: Towards a European SDI, 7th AGILE Conference, April 2004, Heraklion, Greece.
Lehto, L., T. Sarjakoski, A. Hvas, P. Hollander, R. Ruotsalainen and A. Illert (2004): A Prototype Cross-Border GML Data Service, 7th AGILE Conference, April 2004, Heraklion, Greece.
Luzet (2003): EuroSpec - Providing the foundations to maximize the use of GI, in: 9th EC GI & GIS Workshop, ESDI Serving the User. Coruna, Spain, June.
MySQL (2004): http://dev.mysql.com/doc/mysql/en/Spatial_extensions_in_MySQL.html
OGC (2001): Filter Encoding Implementation Specification. http://www.opengis.org/docs/02-059.pdf
Oracle (2004): Oracle Spatial. URL: http://otn.oracle.com/products/spatial

OWL (2004): OWL Web Ontology Language, Overview, W3C Recommendation, 10 February 2004, http://www.w3.org/TR/2004/REC-owl-features-20040210/

PostGIS (2004): http://postgis.refractions.net

Riecken, J., Bernard, L., Portele, C. and Remke, A. (2003): North-Rhine Westphalia: Building a Regional SDI in a Cross-Border Environment, in: 9th EC-GI & GIS Workshop, ESDI Serving the User. Coruna, Spain, June.

SLD (2002): Styled Layer Descriptor (SLD) 1.0.0, Open GIS Consortium Implementation Specification, September 2002, http://www.opengis.org/docs/02-070.pdf

WFS (2002): OGC Web Feature Service Implementation Specification. URL: http://www.opengis.org/docs/02-058.pdf

WMS (2001): OGC Web Map Service Implementation Specification. URL: http://www.opengis.org/docs/01-068r2.pdf

CVS OF THE AUTHORS

Marian de Vries is scientific researcher at the section GIS Technology of Delft University of Technology (the Netherlands). Focus of her research is on Web service/client architectures that are based on interoperable software components and open (data exchange) standards. She was co-developer of the GML prototype of the new TOP10NL product of the Dutch Topographic Service. At the moment she is involved in a number of geo-information integration projects for large data providers in the Netherlands.

Theo Tijssen is a lecturer and researcher at Delft University of Technology, Section GIS-Technology. His current main research topics are geo-DBMS implementation and optimization, geo-DBMS and GIS, data exchange.

CO-ORDINATES

Marian de Vries and Theo Tijssen
GIS Technology – Research Institute for Housing, Urban and Mobility Studies
Faculty of Technology, Policy and Management
Delft University of Technology
Jaffalaan 9
2628 BX Delft
The Netherlands

Tel.	+31-15-278 4268
	+31-15-278 3670
Fax	+31-15-278 2745
E-mail	m.d.vries@otb.tudelft.nl
	t.tijssen@otb.tudelft.nl
Website	www.otb.tudelft.nl
	www.gdmc.nl