A Prototype Cross-Border GML Data Service

Lassi Lehto[1], Tapani Sarjakoski[1], Anders Hvas[2], Pär Hollander[3], Reino Ruotsalainen[4] and Andreas Illert[5]
Finnish Geodetic Institute (1), National Survey and Cadastre – Denmark (2), National Land Survey of Sweden (3), National Land Survey of Finland (4), Federal Agency for Cartography and Geodesy, Germany (5)

SUMMARY

Introduction of the Geography Markup Language (GML) has opened new opportunities for creating integrated cross-border spatial data services. As an XML application GML facilitates data integration by providing open, vendor-independent syntax for encoding spatial data. However, individual GML Application Schemas are inherently dependent on the local data content and thus vary from country to country. Those differences are also reflected in the corresponding Web Feature Service (WFS) queries. Consequently, a cross-border data service must support two-way transformations: those applied to incoming queries and those carried out on the delivered dataset. As the traffic in both directions is encoded in XML syntax, the same techniques can be employed in both tasks. In this paper an approach based on a general-purpose XML processing tool, XSL Transformation (XSLT), is proposed and first experiences gained from a four-nation cross-border prototype service are described.

KEYWORDS: *Cross-border, Data Integration, Schema Transformation, GML, WFS, XSLT*

INTRODUCTION

The recent developments in the GI standardization efforts emphasize the importance of the network-based solutions in spatial data delivery and processing. Many of the activities of the Open GIS Consortium (OGC) have recently been concentrated on various issues related to the use of generic Web technologies in geospatial domain. These include experiments on technologies like Simple Object Access Protocol (SOAP) and Universal Description Discovery and Integration (UDDI) (OGC, 2003). Extensible Markup Language (XML) technologies are routinely used in various data modeling tasks (XML, 2003). The intensive work being currently invested by various parties on the development of the Geography Markup Language (GML), a proposed XML application for spatial data, clearly indicates that Web has become an important platform for different GI-related processes (GML, 2003).

GML is widely seen as an important means for geospatial data encoding in the future. This is reflected for instance by the fact that the GML specification is currently being edited in the ISO TC211 process (ISO 19136) with the goal to make it an ISO standard (ISO, 2003). Some European national mapping agencies are working to establish GML as the default data delivery format, the Ordnance Survey of UK being the most prominent example (Ordnance Survey, 2003). Also the European Commission initiative 'Infrastructure for Spatial Information in Europe' (INSPIRE) proposes in its Architecture and Standards Position Paper that GML would be a mandatory element for the spatial data delivery and the related e-Business applications in Europe (INSPIRE, 2003).

NEED FOR DATA INTEGRATION SERVICES

The ongoing general European integration development creates increasing demand for cross-border data services. This need is exemplified by the wide set of European projects and initiatives aimed at improving the data compatibility and developing integrated services. These include individual projects with generic approach like GINIE (GINIE, 2003), domain-specific developments like EuroRoadS (EuroRoadS, 2003) or EULIS (EULIS, 2003), and initiatives with wider scope like the INSPIRE framework.

European National Mapping Agencies (NMA) have also done extensive work in this area through their co-operation body EuroGeographics. The past and ongoing projects of EuroGeographics in this area have aimed at creation of Pan-European centralized datasets like SABE, EuroGlobalMap and EuroRegionalMap.

One of the recent initiatives of EuroGeographics, called EuroSpec, has a goal to harmonize the European reference datasets in scale range 1:10000 (EuroSpec, 2003). The magnitude of national topographic datasets in this scale range makes the creation of a central database an impossible target. Creation of a Pan-European large-scale reference data service requires distributed database approach. Similarly, the huge amount of data involved and the vast investments spent while collecting the data makes it improbable that these datasets would ever be harmonized in the database level.

The only answer for a Pan-European provision of the reference data in large scale seems thus to be the on-line integration of various individual distributed data services providing heterogeneous data content. This approach requires an integration service, capable of carrying out real-time data transformations. The heterogeneity of the databases must also be reflected in the data queries being sent to the local services. Thus the integration service must apply appropriate transformations to the queries as well.

XSL TRANSFORMATION

One of the various technologies developed on top of the basic XML syntax is a technique for transforming XML documents from one data model to another. This method is called Extensible Stylesheet Language Transformation (XSLT, 1999). The main demand for this kind of tool comes from the need to transform datasets expressed as XML data in different application-specific data models into the XHTML to be displayed in a Web browser.

XSL transformation is a declarative process. The transformation is specified in a separate XML-encoded file that is provided to a software component, called XSLT processor, together with the source XML file. XSLT processor is supposed to read the transformation declarations, apply the specified rules to the source data, and produce the resulting XML structure as the output. Transformation rules are expressed in the form of individual templates that are matched against the data items in the source dataset. The contents of the template are then instantiated to create the corresponding structure in the result dataset. A rich set of functions is available to select items from the source data, process them as required and then formulate the results into the output data stream. There are plenty of free XSLT processors available in the Internet. As one is also included in the Internet Explorer Web browser, the XSLT functionality is available practically on every desktop.

Many innovative ways to apply the same tool have been proposed. These include for example transformations from different kinds of statistical datasets to Scalable Vector Graphics (SVG)-based vector images to visualize the data as graphs. Similarly, a spatial dataset expressed in GML can be transformed into an SVG map image (Lehto et al, 2001). Two new ways to apply XSLT process in the domain of geospatial data are proposed in this paper: transformation from one GML application schema to another, and transformation of a spatial data query formulated against the first data model into a query expressed in terms of the other.

INTEGRATING WFS AS A TWO-WAY TRANSLATOR

The OGC has developed a spatial data delivery service specification called Web Feature Service (WFS). The WFS document has been accepted by the OGC as an official implementation level specification in September 2002 (WFS, 2002). The have also been discussions about starting the work to establish WFS specification as an ISO standard.

WFS can be seen as a translator transforming incoming requests to the internal query mechanism of the data management software used and subsequently transforming the delivered internal data representation to GML. The details about how to constrain the query are specified in a separate OGC document called Filter Encoding (FE, 2002)). According to WFS and FE specifications, the query requesting spatial data from the service is encoded in XML form.

WFS query from Global to Local Schema Transformation

The concept of a cascading map server has been introduced in the OGC's Web Map Server (WMS) specification. Although not explicitly mentioned in the WFS specification, the same idea could be equally well be also applied in the case of a spatial data service. In this approach a WFS acts as a client to several other WFS nodes and provides the resulting integrated dataset as a service to the real clients. As it is expected that those individual WFS nodes contain a heterogeneous set of data, the integrating WFS assumes the role of a Translating WFS, a term given by OGC to a WFS capable of schema transformations. In that case a common data model, called Global Schema, has to be specified for the integrated service.

An interesting fact is that the service query in the case of a Translating WFS is naturally expressed in terms of the translated data model. This makes it necessary for the Translating WFS to take the role of a two-way translator, as also the incoming queries need to be translated from the Global Schema to the original data model of the underlying WFS. In the case of an integrating cross-border WFS server, the query must actually be converted to reflect various different local variants of the dataset.

As both the incoming query and the produced output data in the case of a WFS server are encoded in XML, interestingly, the same tool can be applied for the transformations in both ways. A declarative XSL transformation can equally well be defined for the WFS queries as for transforming the data from one GML Application Schema to another.

The use of a declarative approach makes the system flexible. If the original local data model is changed, only the corresponding declaration needs to be changed. There is no need to edit the executable code of the integrating WFS server. The same applies in the case when a new local WFS node is added to the service. Only a new pair of XSLT files are needed, one for query transformations and one for data transformations. Being text files, XSLT declarations can be edited with the simplest text processing software.

GML from Local to Global Schema Transformation

GML specification does not propose any detailed Feature level schema, but rather only sets a generic framework for building application-specific local data models. This framework consists of basic concepts like Feature, Property, Geometric Property and FeatureCollection. GML also defines a set of rules about how those concepts are to be encoded using the standard XML structures. Thus each individual data provider is free to develop a local data content-specific application schema and deliver the dataset encoded according to that schema.

To integrate heterogeneous datasets, it is clearly necessary to define a common Global Schema into which the individual dataset are to be transformed. This schema can be developed as a largest possible common intersection of the Feature types available in the participating datasets, or it can be determined from of the user requirements point of view.

In the transformation process the local Feature types might need to be combined to create a Feature type in the Global Schema. In another case Features in a single Feature type in the local database need to be split into two Global Schema Feature types. There might be a need for filtering out individual Features because of varying collection criteria applied in the local data collection. The Properties of the Features might need to be processed, because different scales and units of measure are used. Schema transformation is also a natural place to add reasonable default values to Properties for which there is no value available in the source dataset. More challenging operations include those needed to convert the Geometric Property of a Feature from one dimension to another, for instance to collapse an area to a point.

An additional processing step is needed in most of the cross-border applications – that of the coordinate transformation. National coordinates need to be translated to the selected common coordinate system during the GML schema transformation. The same need also applies in the other direction; the spatial extent defined in the WFS query in the common coordinate system needs to be translated to the local systems. Complicated coordinate transformations can be introduced into the XSLT process by defining an extension function for the purpose.

CASE: GiMoDig

The European Commission-funded GiMoDig project develops methods for the provision of vector maps to mobile devices from heterogeneous distributed databases in a cross-border setting (GiMoDig, 2003a). The participant organizations in the GiMoDig project include four NMAs (Finland, Sweden, Denmark and Germany). The data available from the service represent the best topographic datasets with national coverage in the scale range 1:10000.

The project applies XML-based techniques extensively. WFS implementations are used in the four NMAs to provide GML data. The integration service run by the Finnish Geodetic Institute, the coordinator of the project, transforms the data streams from the local GML schemas to the common Global Schema defined in the project (GiMoDig, 2003b). At the same time a coordinate transformation is carried out to go from the local coordinate systems to the selected common ETRS89 system. The integration service is designed to implement a cascading WFS node, serving topographic map data from the four participating countries. Currently two cross-border test areas are available, one from Tornio-Haparanda region on the border between Finland and Sweden and the other from Flensburg area on the Danish-German border.

The queries need to be transformed in a similar way. Incoming queries are to be formulated in terms of the Global Schema. These queries must be transformed into four different local variants reflecting the organization of data in the local level. An exemplary Global Schema query is shown in the code snippet in the figure 1.

```xml
<?xml version='1.0' encoding='iso-8859-1'?>
<GetFeature outputFormat='GML2' handle='GiMoDigQuery'>
 <Query handle='query00' typeName='Road' version='1.0'>
  <PropertyName>centerLineOf</PropertyName>
  <PropertyName>intendedUse</PropertyName>
  <PropertyName>name</PropertyName>
  <Filter xmlns:gml='http://www.opengis.net/gml'>
 <BBOX>
  <PropertyName>centerLineOf</PropertyName>
  <gml:Box srsName='EPSG:4258'>
   <gml:coordinates>
  366500,7305000 368500,7307000
   </gml:coordinates>
  </gml:Box>
 </BBOX>
  </Filter>
 </Query>
</GetFeature>
```

Figure 1: A WFS query requesting Road data, expressed in the GiMoDig Global Schema.

The Finnish and Swedish versions of the local query corresponding to the Global Schema query in code snippet 1 are shown on the code snippets in the figures 2 and 3.

```
<?xml version='1.0' encoding='iso-8859-1'?>
<GetFeature outputFormat='GML2' handle='geoserverQuery'>
 <Query handle='query00' typeName='liikenneverkot_viiva' version='1.0'>
  <PropertyName>the_geom</PropertyName>
   <Filter xmlns:gml='http://www.opengis.net/gml'>
  <And>
   <BBOX>
    <PropertyName>the_geom</PropertyName>
    <gml:Box srsName='EPSG:2393'>
   <gml:coordinates>
    3372120,7308556 3373621,7310057
   </gml:coordinates>
    </gml:Box>
   </BBOX>
   <Or>
    <PropertyIsEqualTo>
   <PropertyName>luokka</PropertyName>
   <Literal>12111</Literal>
    </PropertyIsEqualTo>
    <PropertyIsEqualTo>
   <PropertyName>luokka</PropertyName>
   <Literal>12112</Literal>
    </PropertyIsEqualTo>
    <PropertyIsEqualTo>
   <PropertyName>luokka</PropertyName>
   <Literal>12121</Literal>
    </PropertyIsEqualTo>
    <PropertyIsEqualTo>
   <PropertyName>luokka</PropertyName>
   <Literal>12122</Literal>
    </PropertyIsEqualTo>
    <PropertyIsEqualTo>
   <PropertyName>luokka</PropertyName>
   <Literal>12131</Literal>
    </PropertyIsEqualTo>
    <PropertyIsEqualTo>
   <PropertyName>luokka</PropertyName>
   <Literal>12132</Literal>
    </PropertyIsEqualTo>
    <PropertyIsEqualTo>
   <PropertyName>luokka</PropertyName>
   <Literal>12141</Literal>
    </PropertyIsEqualTo>
   </Or>
  </And>
   </Filter>
  </Query>
</GetFeature>
```

Figure 2: A WFS query requesting Road data, expressed in the Finnish local schema.

```
<?xml version='1.0' encoding='iso-8859-1'?>
<GetFeature outputFormat='GML2' handle='geoserverQuery'>
 <Query handle='query00' typeName='vag_jarnvag' version='1.0'>
  <PropertyName>the_geom</PropertyName>
   <Filter xmlns:gml='http://www.opengis.net/gml'>
  <And>
   <BBOX>
    <PropertyName>the_geom</PropertyName>
    <gml:Box srsName='EPSG:2400'>
   <gml:coordinates>
    1877062,7329068 1878679,7331396
   </gml:coordinates>
    </gml:Box>
   </BBOX>
   <PropertyIsLike>
    <PropertyName>detaljtyp</PropertyName>
    <Literal>VÄG*</Literal>
   </PropertyIsLike>
  </And>
   </Filter>
  </Query>
  <Query handle='query01' typeName='ovr_vag' version='1.0'>
   <PropertyName>the_geom</PropertyName>
   <Filter xmlns:gml='http://www.opengis.net/gml'>
  <And>
   <BBOX>
    <PropertyName>the_geom</PropertyName>
    <gml:Box srsName='EPSG:2400'>
   <gml:coordinates>
    1877062,7329068 1878679,7331396
   </gml:coordinates>
    </gml:Box>
   </BBOX>
   <PropertyIsEqualTo>
    <PropertyName>detaljtyp</PropertyName>
    <Literal>ÖVÄGTRA.M</Literal>
   </PropertyIsEqualTo>
  </And>
   </Filter>
  </Query>
 </GetFeature>
```

Figure 3: A WFS query requesting Road data, expressed in the Swedish local schema.

The local queries reflect the structure of the road data in Finland and Sweden. In the case of Finland one local Feature type is queried with a set of conditions on one Property value, whereas in Sweden two local Feature types must be queried with appropriate conditions on a Property value. The coordinate values inside the gml:Box elements reflect the differences between the three reference systems involved.

The GML data transformations from the local national datamodels to the GiModig Global Schema naturally involve much bigger datasets. A query for a complete mapsheet might result in several megabytes of XML stream. The processing time spent on the schema transformation and the embedded coordinate transformation seems to be almost linearly dependent on the number of geospatial Features included. The first tests indicate processing times varying from 0.7 to 1.0 ms per Feature.

SUMMARY

Pan-European services providing large-scale geospatial data are going to be based on distributed service architecture. This vision is exemplified in various European initiatives, like in the Commission-supported INSPIRE, and in the recent work of EuroGeographics. Although GML is becoming a widely adopted spatial data encoding standard, the flexibility in creating the local GML Application Schemas results in a set of heterogeneous service nodes reflecting national data structures. There seems to be an urgent need for integration services able to carry out on-line schema transformations.

The cascading WFS concept can be applied to establish cross-border services integrating two or more national WFS nodes into a one seamless whole. The cascading WFS needs to be capable of carrying out two-way transformations: incoming data requests need to be transformed from the adopted Global Schema to the local variants, and resulting datasets must be transformed from the local data model to the Global Schema of the integrated service. As the messages in both directions are encoded in XML, a general-purpose XML processing technology, XSL Transformation could be applied in the required transformations.

As a case study the service developed by the GiMoDig project is presented. The project's WFS implementation combines topographic map data servers of the four participating NMAs into a one cascading WFS. The first results are promising. Although the performance of the XSLT-based process is not as good as a task-specific low-level executable code would be, the transformation step doesn't seem to be a bottleneck in the overall query processing as seen from the client side. Most importantly, the declarative way of defining the transformations makes the system flexible to adapt to changing requirements.

BIBLIOGRAPHY

EULIS, 2003, EULIS Home Page, www.eulis.org

EuroGeographics, 2003, EuroGeographics Home Page, www.eurogeographics.org

EuroRoadS, 2003, EuroRoadS Project, www.eurogeographics.org/eng/03_projects_euroroads.asp

EuroSpec, 2003, EuroSpec Initiative, www.eurogeographics.org/eng/03_projects_eurospec.asp

FE, 2002, Filter Encoding Specification, www.opengis.org/docs/02-059.pdf

GiMoDig, 2003a, GiMoDig Project Home Page, gimodig.fgi.fi

GiMoDig, 2003b, GiMoDig Global Schema, gimodig.fgi.fi/pub_deliverables/ GiMoDigD521_global.pdf

GINIE, 2003, Geographic Information Network in Europe, Home Page, www.ec-gis.org/ginie/

GML, 2003, Geography Markup Language version 3.0, www.opengis.org/docs/02-023r4.pdf

INSPIRE, 2003, INSPIRE Architecture and Standards Position Paper, inspire.jrc.it/reports/ position_papers/inspire_ast_pp_v4_3_en.pdf

ISO, 2003, ISO/OGC GML Working Group, www.interactive-instruments.de/isogml.html

Lehto L, Kähkönen J and T. Sarjakoski, Multi-purpose Publishing of Geodata in the Web, Proceedings of the 4th AGILE Conference on Geographic Information Science, Brno, Czech Republic, April 19–21, 2001, pp. 209–214.

OGC, 2003, Open GIS Consortium Home Page, www.opengis.org/specs/?page=discussion

Ordnance Survey, 2003, OS GML Schema Repository, www.ordnancesurvey.co.uk/ oswebsite/xml/schema/

WFS, 2002, Web Feature Service Specification, www.opengis.org/docs/02-058.pdf

XML, 2003, World Wide Web Consortium XML Page, www.w3.org/XML

XSLT, 1999, Extensible Stylesheet Language Transformations, www.w3.org/TR/xslt