Geography Markup Language: a 'de facto' Standard

Exchanging Geo-data on The Web

The internet and the World Wide Web have rapidly become increasingly important in the field of geo-information technology. Such exchange of geo-information requires standards and implementation of these. A suitable language is the Geography Markup Language (GML); now endorsed by the OGC, this has become the 'de facto' standard. The authors present ongoing developments.

By Wilko Quak and Mathias Lemmens, Delft University of Technology, The Netherlands

Vast amounts of digital geo-data for use in infrastructure maintenance, urban planning, environmental, natural and technical disaster and land and natural resources management and location based services has been created over the past twenty-five years in many more than 25 data formats, most proprietary to the various GIS vendors. In the early nineties people began to worry about the waste of resources arising from so much incompatibility. Interoperability became the recipe, resulting in 1994 in the founding of the Open GIS Consortium (OGC), recently renamed the Open Geospatial Consortium. An open software interface for geo-information specified by the OGC has now been co-operatively adopted by the main GIS vendors. Today, the Geography Markup Language (GML) plays a key role in both geo-information encoding and transport and in the description of geographic objects for geo-information Web services.

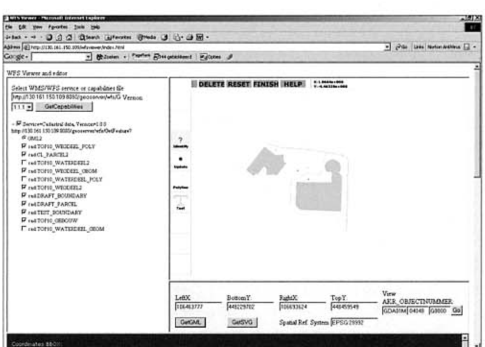

Figure 1, Red line shows new boundary as sketched by notary, stored in temporary cadastral database and returned after validation.

Design Goals

GML is a geo-information exchange language based on XML (eXtensible Markup Language), the standard language of the internet. GML is not tied to any proprietary GIS or database software but is specifically designed for feature-based geo-information. It is an open standard, which anyone can use. GML has been designed to support interoperability. OGC design goals include more specifically:

- providing a means of encoding spatial information for both data transport and data storage, especially in a wide-area internet context
- being sufficiently extensible to support a wide variety of spatial tasks, from portrayal to analysis
- establishing the foundation for internet GIS in an incremental and modular fashion
- allowing for efficient encoding, such as data compression of geo-spatial geometry
- providing easy-to-understand encoding of spatial information and spatial relationships, including those defined by the OGC Simple Features model
- enabling separation of spatial and non-spatial content from data presentation (graphics or otherwise)
- permitting easy integration of spatial and non-spatial data, especially for cases in which the non-spatial data is XML-encoded
- enabling ready linkage of spatial geometric elements to other spatial or non-spatial elements
- providing a set of common geographic modelling objects to enable interoperability of independently developed applications.

Characteristic Contents

A GML file consists of two types of documents:

- one containing the actual data
- one providing information about how this data is structured: schema definition. This tells the computer how to read and interpret the data, basically referring to the Digital Landscape Model (DLM).

Since one design goal of GML is its primary concern with geo-information exchange – the DLM – content and presentation are separated. In order to make a map from GML data, this needs to be converted into symbols, colours, line and area styles: for short, a Digital Cartographic Model (DCM). Since a GML document is composed of text (like XML) it can be edited by any WordPad-type text editor. The appearance of geo-data as text necessitates an abundant use of tags, resulting in high network load and commensurate computational demands. Compressing techniques may reduce the size by a factor 10 to 20. However, compression results in loss of performance. Another disadvantage of the text format – inability to process gridded data – has been resolved with GML3.

New Version GML3

The first version of GML was introduced in 2001 and GML2 was introduced in February 2001. Like GML2, GML3 will play a key role in both geo-data encoding and transport and in the description of geographic objects for geospatial Web services. The latest version GML3 is modular so that users can pick out the schemas or schema elements that apply to their work, thus simplifying and minimising the size of implementations. New features designed specifically to address real-world problems are:
- complex geometry, like curves, especially necessary for motorway and other transportation design issues
- time information, essential for tracking applications like monitoring locations of emergency services and exploring movement and growth of natural disasters such as floods and fires over hours, days or months
- topology: relationships between features, a key requirement for National Mapping Agencies and for Location-Based Services (LBS) routing applications.

Other new features include:
- units of measure
- spatial reference systems
- metadata
- gridded data, like aerial and satellite imagery
- default styles for feature and coverage visualisation; rules on visualisation of data can be exchanged with styling module.

Transactional WFS

One of the many GML applications is the Web Feature Service (WFS), a GML protocol for exchanging data between client and server. This OGC standard enables assessment of geo-information over the Web using one vendor's GIS functionality or browser whilst querying and displaying using software from another vendor. In addition to querying and displaying, editing is also possible using transactional WFS. To test the functionality of this in practice we built an experimental system for use by notaries. When a property parcel is split up into two parcels the notary draws the rough location of the splitting boundary on a paper map as part of the transaction. The cadastre surveyor uses this draft boundary as a preliminary one when measuring its exact location as described by seller and buyer. Since the exact location of the boundary of parcels is determined only on completion of the legal transaction, a notary may make do with a sketch; in case of disagreement, a judge will decide upon the actual location of the boundary. Direct registration of the draft boundary in a temporary cadastral database would have several advantages. First of all, it would ease cadastral (surveying) processes. In addition, it would improve legal security because the new situation would become instantaneously available to others active in the property market. Direct registration becomes possible by sketching the new boundary on a digital map in a Web client and sending the digital boundaries to the cadastral database via a Web service. A possible scenario would proceed as follows.

An experimental system for use by notaries

1. Login to the server from the client side.
2. Provision by notary of ID of property parcel concerned.
3. Return by server of boundaries and labels from the cadastral database.
4. Return by server (when applicable) of boundaries and labels from a temporary database resulting from previous notary action not yet processed by the cadastre.
5. Drawing up by notary of new boundary and provision of deed information about associated parcels.
6. Validation of data sent to server, for example regarding uniqueness of labels.
7. Insertion after validation of new features into temporary cadastral database and server return to client of result of transaction (Figure 1).

System and Results

The aim of our experimental system was to test the feasibility of such an approach. The open source transactional WFS server GeoServer (http://geoserver.sourceforge.net) was used in developing the system, a full implementation of the OpenGIS Consortium WFS specification. We configured GeoServer as a Transactional Web Feature Server on top of an Oracle Spatial database. The WFS client running at the notary was implemented on standard web technology so that no complicated GIS software had to be installed at the client side.

WFS technology is promising and has many advantages; for example, data like ortho images and topographic data from other servers can be added in the client as additional background information. But a few problems still remain.
- In an interoperable environment in which client and server come from different vendors the interpretation of information exchanged between them should be fully transparent. However, some functionality is interpreted differently by different providers. As the standards mature, so these issues will be resolved.
- Security warranting access to the database only by authorised notaries.
- Prevention of simultaneous transactions, so that a parcel is locked whilst editing by authorised person.

Complexity Issues

In its attempt to provide a modelling tool for any type of geo-data, GML has become increasingly complex and hard to implement completely. The size of the specification has grown from 85 pages for GML1 to 548 pages for the GML3 standard. This makes it hard for any GIS vendor to be fully GML-compliant. A proposal to split the complete standard into profiles, where 0-profile is a simple, widely accepted profile, should make GML more convenient. The 0-profile could also help in the development of WFS clients; there are currently some WFS systems operational but few generic clients. This is because a WFS Server can make the GML server very complex and the client must be able to consume any GML document produced by the server, thus laying a heavy burden on the client side. But it is feasible to build a client that can read any level 0 document. A new GML3 raster format has been defined to support gridded data. Is this feasible, given the many popular raster-encoding formats such as jpeg, tiff and gif? JPEG2000, for example, is an imaging format that supports lossless compression of images and offers many interesting options for the storage of Remote Sensing Data (bigger bit depths, tiles, resolution progression, quality progression and fast access to spatial locations). JPEG2000 also has the option of encoding GML fragments inside the image. It will be possible to geo-reference an image (like GeoTiff). But much apart from camera position and extent of image can be encoded in the GML part of the image. Annotating the image with vector data will be possible.

Concluding Remarks

As so often, developing and implementing standards like GML is a necessary but not sufficient condition for exchanging map products over the internet. Indeed, standards and technological capabilities form the fundament. But it will be the awareness of decision-makers and their commitment to incorporating new technology in existing work processes that will make it all work.

Acknowledgement

Thanks are due to Marian de Vries and Thijs Brentjens for providing the WFS example.

Further Reading

- Brentjens, T.J., 2004, OpenGIS web feature services for editing cadastral data; analysis and practical experiences, MSc thesis, Delft University of Technology, Section GIS Technology, http://www.gdmc.nl
- De Vries, M., Quak, W., 2004, GML 3, X3D and Topology, paper presented at GML Days 2004, July 25th to 29th, Vancouver, British Columbia, http:// www.gdmc.nl
- Insider's View, this issue of GIM International

Websites

- The GML 3 specification is available for download at www. opengis.org/techno/implementation.htm
- The ISO TC211 website reads: www.isotc211.org/◆

Biography of the Authors

Both authors are with Delft University of Technology. **Wilko Quak** holds a MSc degree in Computer Science from Utrecht University in The Netherlands. His research interests include standardisation of spatial data and spatial DBMS performance and modelling. **Dr Mathias Lemmens** is assistant professor in GIS Technology and editor of GIM International. His research interests include acquisition and processing of 3-dimensional point data and GIS quality issues.

Wilko Quak

Wilko Quak, Delft University of Technology, Section GIS Technology, P.O. Box 5030, 2600 GA Delft, The Netherlands, e-mail: w.quak@otb.tudelft.nl