Web Mapping and XML Technologies
"A Close Relationship"

Maria Spanaki, Byron Antoniou, Lysandros Tsoulos
Surveying Eng. MSc Geomatics PhD Student, Lieut. AMS, Assoc. Professor NTUA
Cartography Laboratory - National Technical University of Athens, Greece
{spanaki,rs02626}@mail.ntua.gr, lysandro@central.ntua.gr

SUMMARY

The role of XML technologies in GIS and Cartography becomes more and more critical offering great advantages in web-based applications. The objective of this study is to elaborate on the utilization of the XML specifications used in modern web cartographic applications. Emphasis is given to GML, XSL and SVG standards that constitute the major XML encodings used in web Cartography. The Geography Markup Language (GML) constitutes a powerful means of coding and processing spatial data. Extensible Stylesheet Language Transformation (XSLT) is an XML codification concerning the conversion of XML documents from one form to another, while SVG is a two-dimensional vector XML based standard that ensures the visualization of XML data in a web browser environment. The process leading to a working example of the utilization of these technologies is presented, in order to provide proof of concept.

KEYWORDS: XML, GML, XSL, SVG

INTRODUCTION

The Standard Generalized Markup Language (SGML) became popular to the users with the advent of World Wide Web. In fact, what really people were utilizing was the HyperText Markup Language (HTML), which is a SGML application used for the construction of Web pages. HTML is already overburdened with dozens of interesting but incompatible inventions from different manufacturers, because it provides only one way of describing information. As the web grew, the need for better organizational and searching methods came to the fore. It became apparent that a simpler version of the full SGML standard would be advantageous. Thus XML was created. XML is an abbreviated version of SGML, to make it easier for the users to define their own document types and to make it easier for programmers to write programs to handle them. It omits all the options and most of the more complex and less-used parts of SGML enabling the easier development of applications deliverable and interoperable over the Web (Goldfarb 2002).

The main reasons for migrating from SGML and HTML to XML are listed below (Flynn 2003):

- Authors and providers can design their own document types using XML, instead of being stuck with HTML. Document types can be explicitly tailored to an audience, so the cumbersome fudging that has to take place with HTML can become a thing of the past: authors and designers are free to invent their own mark-up elements.
- Information content can be richer and easier to use, because the descriptive and hypertext linking abilities of XML are much greater than those of HTML.
- XML can provide more and better facilities for browser presentation and performance, using CSS and XSL stylesheets.

- Information will be more accessible and reusable, because the more flexible mark-up of XML can be used by any XML software instead of being restricted to specific manufacturers as has become the case with HTML.

GML: THE XML STANDARD FOR GEOGRAPHIC DATA

Geographic data is the digital representation of the world in spatial terms that is independent of any particular form of visualization. Geographic data contain information about the geometry and properties of the spatial objects. Just as XML helps to clearly separate content from presentation for web pages, so Geography Markup Language (GML) does the same in the world of geography and mapping. GML provides an XML-based encoding for the modelling, transport, and storage of geographic information including both the spatial and non-spatial properties of geographic features and it can be viewed also as a basic application framework for handling geographic information in an open and non-proprietary environment.. By leveraging related XML technologies (e.g. XML Schema, XSLT, XLink, XPointer, SVG) a GML dataset becomes easier to process in heterogeneous environments and it can be readily utilized with other types of data: text, video, imagery, etc. Since GML documents are both human-readable and machine-parsable, they are easier to understand and maintain than proprietary binary formats (Lake 2003).

GML is designed to support interoperability among different data models and feature representations by providing a common data model, a set of basic and sophisticated geometry tags to describe the spatial features and a mechanism for creating and sharing application schemas. GML also provides standards to describe features, geometry and feature associations by developing numerous schemas as it is documented in detail in GML 3.0 standard (OGC GML 2003). The value of the use of an application schema is better understood when the GML-coded geospatial data is transported and all the mark-up elements that describe the spatial and the non-spatial features, geometry and spatial reference systems of the data are also transported to the recipient. Based on the standard mark-up elements, the receiving party knows exactly what each data component means and how to extract it so that nothing gets lost or distorted in the transport and translation process. This is particularly important for real-time data access and transport in the Internet environment. Besides encoding data either for storage or transport purposes, the GML-coded spatial and non-spatial data enable vector mapping and even spatial analysis on standard web browsers and establish the foundation for Internet GIS (Zhong-Ren Peng and Ming-Hsiang Tsou 2003).

SCALABLE VECTOR GRAPHICS (SVG)

GML upholds the principle of separating content from presentation, so it does not address the visualization of encoded data. A styling mechanism is required to display GML data on a particular device. To make a map from GML the only thing needed is the styling of the GML elements into a form, which can be interpreted for graphical display in a web browser. Potential graphical display formats include W3C Scalable Vector Graphics (SVG), the Microsoft Vector Markup Language (VML), and the X3D (Lake 2003). Up to now, any serious attempt concerning Internet Cartography was limited by technical specifications, which made high quality cartography rather impossible. Most cartographers at this particular front have been busy creating workarounds preventing them from doing their actual work. For the first time SVG - the new Internet vector standard - reduces this strain. It opens ways for cartographers to concentrate on contents, on interactions, still typical for monitor cartography (Neumann and Winter 2003). The open character of Internet along with the success of the open source model allows for new advances in Cartography, which are of major significance for other fields as well like data processing and sales/promotion.

SVG is an adopted W3C standard, which is under constant development. SVG's upgraded version 1.1 was launched on 14 January 2003 as an official specification replacing SVG 1.0. SVG is a standardized XML language for describing 2D graphics via vector graphics, text and raster graphics (W3C SVG 2003).

Graphical formats, as we know them from graphics software, can be defined by CSS in order to confer standard appearance to a group of graphical objects. This way the appearance of several similar objects may also be modified effectively. Grace to style (format) compatibility, SVG files can simply be integrated in existing web applications. Integration of SVG into existing DOM (Document Object Model) means that SVG elements can be controlled and modified by the usual JavaScript/Java interfaces (Neumann and Winter 2003). The SVG standard was developed and is supported by all major graphics and software companies and organizations that are web relevant. The most notable advantages of SVG are (Mansfield and Fuller 2001):

- SVG graphics are resolution and device independent (can be scaled to match different devices)
- SVG format creates smaller size files that reduce download times compared to bitmapped graphics. That is why they are better suited for devices with low bandwidth and limited memory
- SVG images can be panned and zoomed without degrading image quality
- SVG specification is fully compatible with existing technologies like HTML and XHTML, XLink, XML Namespaces, DOM, CSS and XSL
- SVG graphics can use scripting to provide interactivity and animation
- SVG images can be animated by two methods: the use of declarative SVG elements (mostly using elements borrowed from SMIL Animation) or the use of ECMAScript (JavaScript) or another scripting language to manipulate the Document Object Model (DOM) of an SVG image.

SVG offers all the advantages of Flash, the de-facto standard of the day, plus the following features: embedded fonts, extensible mark-up language (XML), stylesheets (CSS), interactivity and animation (Neumann and Winter 2003).

DATA EXTRACTION AND MAP COMPOSITION PROCESS

As the Web has grown in popularity and complexity, users and content providers have demanded more Web sites with visual sophistication. The low-resolution GIFs that populate Web pages today have limited capabilities. XML family of technologies constitutes a revolution in the way geographic information is encoded, transferred and rendered. These different languages are all built in the same syntax, giving powerful new possibilities for interchanging information between computer applications and composing maps on the Web.

In the framework of the research carried out, the objective of map composition on the Web through the exclusive use of XML technologies has been sought and achieved through a process, which includes the transformation of raw XML data to GML encoding format and the utilization of SVG graphic representation. The idea behind this approach emanates from the application specific spatial model in relation with the desirable visual mapping outcome. Therefore a model has been developed taking into account the spatial data characteristics and the scope of the spatial database, where spatial data will be stored in an object-oriented environment. The database utilized has been implemented in Oracle Spatial 9i, which in its current version includes an extension of the object-relational model enabling the storage of an SDO_GEOMETRY type for the geometry part of the objects constructs (Spanaki and Tsoulos 2003). The geometry is stored using what the OGC calls simple features and composite collections of these primitive types. The same standards are also used in GML and in the database conceptual model. This compatibility is enviable, since the geo-database is developed to work as a repository where from the GML data is generated.

Cartographic requirements identify the necessary datasets to be chosen for subsequent extraction in XML. Three layers of information are selected through spatial queries from each group of geographic objects namely points, lines and polygons. The point layers describe spatial information such as elevation points, churches etc.. The line layers are used to describe roads, rivers, contours, coastlines and the polygon layers the settlements. Furthermore, a layer for annotation is used for the toponyms.. In the next step, a GML application schema must be developed to include all basic geometries (points, lines and polygons). This holistic approach has been adopted in order to achieve a group transformation of source data for all basic geometry types. Transforming XML data to GML ensures that the outcome will be geographically correct. This correctness is crucial in order that the result of the next transformation (from GML to SVG) to be an image to a SVG-enabled web browser. A special software program (XmlSpy) that helps the design, editing, parsing and debugging of XML applications has been used for the creation of the GML schema file.

Data transformation from XML to GML format is accomplished with the utilization of XSLT stylesheets in a DOM class parser environment. XSLT is considered as a transformative style language and provides a set of tools for transforming documents from their original structure to a new structure. Instead of adding information to the original document structure, it creates a new document structure based on the stylesheet's rules that are applied to the content of the original file. More specifically, an XSLT stylesheet is a collection of template rules. Each template rule includes a pattern that identifies the source tree nodes to which the pattern applies and a template that is added to the resulting tree when the XSLT processor applies a specific template rule to a matched node. An XML specification is affiliated with XSLT in the above procedure. XML Path Language (XPath) is a specialized language for addressing parts in an XML document. XPath provides a simple and concise syntax for the identification of nodes in an XML document, based on the node's type, name, content and context in relation to other nodes in the document's tree (W3C XPath 1999). XSLT provides a grammar in which the results of XPath queries are associated with templates to describe the transformation - parsing of data in the XML source document as a new XML document. Like the core of the SQL languages used to manipulate relational databases, XPath is important because it provides a flexible way to identify the desirable information that someone might want to pull out from a larger collection. XSL transformations are implemented using XSLT processors. More specifically, XSLT processing involves the following steps:

a. Reading the XML source document and the associated XSLT style sheets
b. Parsing XML files and then their associated XSLT files into XML Document Object Model (DOM) objects that represent the XML elements as trees of nodes. Each node corresponds to an XML document element or attribute
c. Applying XSLT transformation to the source trees
d. Producing resulting trees in accordance with the specification of the XSLT style sheet
e. Serializing the resulting trees as output files. These files output XML, GML, SVG or other formats

The last part of the map composition procedure is the generation of SVG graphic results from GML. Once again, XSLT is involved in order to perform the transformations that lead to the visualization of GML data through SVG graphics. XSLT processors may work in a chained automatic mode responding to data transformation demands in an online environment, or they can just be used in an off-line mode for the generation of base maps that are used in online applications. This case study deals with the second aspect offering predefined cartographic displays in a web browser's environment. The SVG elements' basic attributes should be considered when transforming GML files to SVG. The correspondence between the GML geometry types and the SVG structures is easily implemented as they refer to the same geometric constructs (Spanaki and Tsoulos 2003). For the creation of the SVG graphic output through the transformation of XML files, the Cascading Style Sheets – CSSs can be used. A CSS is commonly described as an annotative style language meaning that it adds formatting information to the document tree rather than changing the document itself. CSSs don't use XML or XSL syntax formatting in a different way the final graphic result. CSS is easier to use than XSLT and requires less memory because it cannot reorder a document or build a tree representation of the document. SVG supports various methods for the implementation of CSS styling, such as external CSS stylesheets, internal CSS stylesheets or inline styling. All these options give considerable control capabilities in web mapping implementations, which is also a very important issue in cartography, where independence of the data transformation procedure from the drawing process is needed and must be taken into account. On the other hand, since the content is separated from presentation, different stylesheets can produce alternative symbolization scenarios for the same data set according to the users' preferences. Resulting SVG images from the above described process for points, lines and polygon features and annotation, are finally presented in an SVG-enabled web browser. The application is enhanced with multiple interactive capabilities implemented using Java script. An HTML web page is constructed (figure 1) including a number of nested SVG elements. This fact facilitates the accurate positioning of the elements composing the application interface and the relative position between the elements themselves, when at the same time the advantages of HTML specifications are being implemented. The application's interface - apart from the common used logos and navigation bar (implemented in SVG, figure 2) - includes a separate SVG window for the display of the geographic datasets. In addition, there is a legend, which interacts with the datasets and a tool-set with two basic functions: pan and zoom. It must be mentioned that the SVG viewer provides such functions with the combination of keyboard and mouse (i.e. Ctrl + Right click), but a straightforward interface is much more comprehensive.

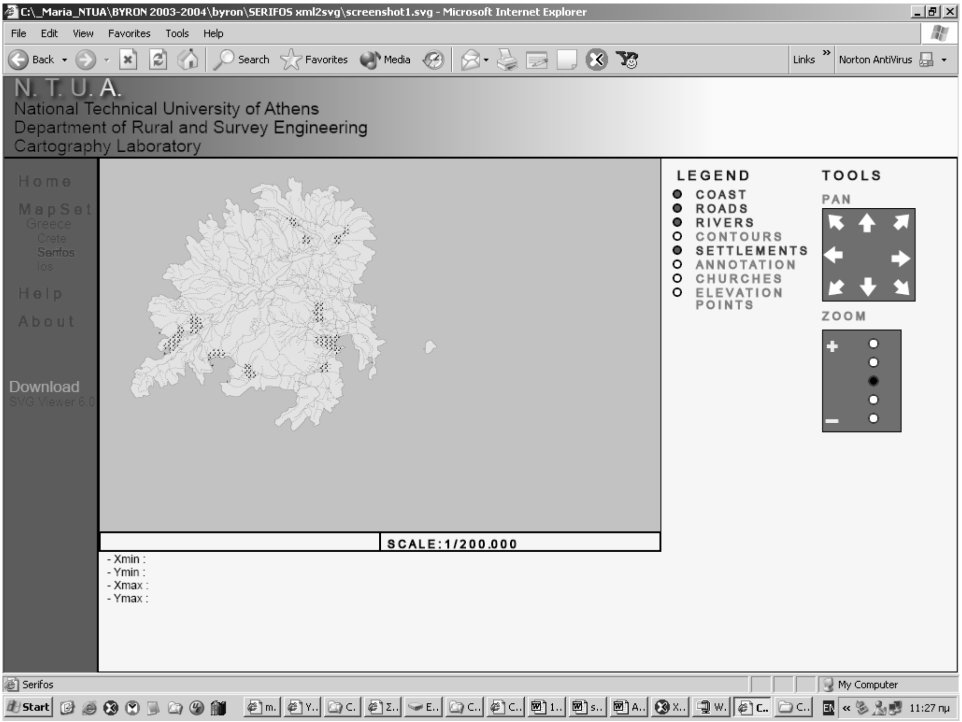

Figure 1: HTML web page with a number of nested SVG elements

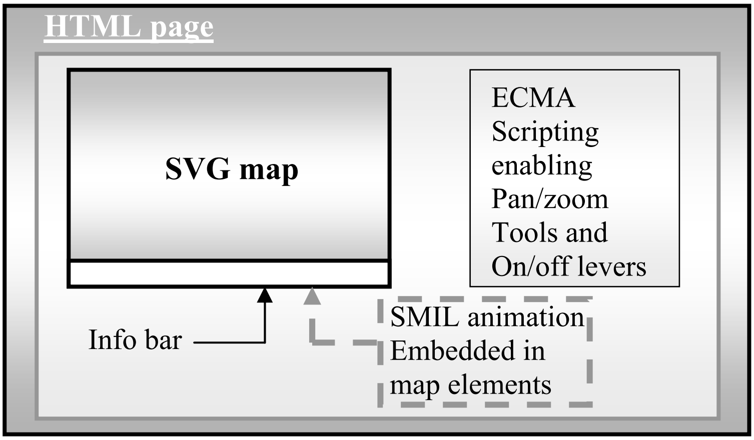

Figure 2: Technologies utilized for the application interface

CONCLUSIONS

The World Wide Web (WWW) is the most recent new medium to present and disseminate geospatial data. In this process the map plays a key role, and has multiple functions. Because of the nature of the WWW the map can also function as an interface or index to additional information (http://kartoweb.itc.nl). The process of map composition on the Web requires the utilization of the family of XML technologies, which provide all those tools needed for static or dynamic map composition in a cohesive and stable environment.

BIBLIOGRAPHY

Flynn P., Frequently Asked Questions about the Extensible Markup Language, www.ucc.ie/xml/faq.html 2003.

Goldfarb C., A Brief History of the Development of SGML, sgmlsource.com/history/sgmlhist.htm 2002.

Guarino N., Semantic Matching: Formal Ontological Distinctions for Information Organization, Extraction, and Integration. In M.T. Pazienza (ed.). Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology. Springer Verlag: 139-170, 1998.

Lake R., What is a Map Style? Building on GML and SVG. In Proceedings of SVG Open Conference 2003.

Mansfield P, Fuller D, 2001 Graphical Stylesheets Using XSLT to Generate SVG, In proceedings of XML Conference 2001.

Neumann A., Winter A., 2003 Vector-based Web Cartography: Enabler SVG, www.carto.net

OGC GML, 2003 Geography Markup Language (GML) 3.0 Implementation Specification, OpenGIS Consortium.

Spanaki M., Tsoulos L., 2003 Holistic Approach of Map Composition Utilizing XML. Proceedings of SVG Open Conference 2003.

Zhong-Ren Peng, Ming-Hsiang Tsou, 2003 Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Networks, ISBN: 0-471-35923-8, pp 720.

W3C CSS 1988, Cascading Style Sheets, level 2, World Wide Web (W3C) Specification.

W3C SVG 2003, Scalable Vector Graphics (SVG) 1.1 Specification, World Wide Web (W3C) Recommendation.

W3C XPath 1999, XML Path Language (XPath) Version 1.0, , World Wide Web (W3C) Recommendation.

W3C XSLT 1999, XSL Transformations (XSLT) Version 1.0, World Wide Web (W3C) Recommendation.