APPLICATION OF MARKUP LANGUAGES IN CARTOGRAPHY

Otakar Čerba[1]

[1]Department of mathematics, Faculty of Applied Sciences, The University of West Bohemia in Pilsen,
Univerzitní 23, 306 14, Plzeň, Czech Republic
ota.cerba@seznam.cz

Abstrakt. Během Mezinárodního sympozia GIS... Ostrava 2006 byl přednesen příspěvek Cartographic e-documents & SGML/XML. Tento příspěvek shrnoval možnosti formátu XML (eXtensible Markup Language) a jeho potencionálních aplikací v digitální kartografii. Vztah kartografie a XML se během dvou let změnil. Proto tento příspěvek nabízí inovovaný přehled „kartografických" XML a příbuzných formátů a jazyků, jejich výhod a nevýhod. Tento dokument je konkrétně zaměřený na

- ukládání, kódování (prostorová data, vektorové grafické formáty, formáty popisující komponenty mapy),
- popis (metadatové formáty),
- transformace (transformační a stylové jazyky) prostorových dat určených pro kartografickou vizualizaci.

Tento příspěvek není určený k propagaci konkrétního formátu nebo technologie. Mnohem důležitější je implementace a dodržování jednotlivých standardů. Aplikace standardů z oblasti kartografie a informačních technologií přispívá k vyšší úrovni kartografických produktů. Tato vyšší kvalita zapříčiňuje lepší nezávislost, navigaci, orientaci uživatele, čitelnost, přístupnost apod. Standardizace by měla vést také k vyšší unifikaci relativně rozšířeného spektra XML formátů (např. jazyků pro popis dokumentu nebo metadatových formátů).

Klíčová slova: Značkovací jazyky, XML, digitální kartografie, XML schémata, standardy.

Abstract. Through the International symposium GIS... Ostrava 2006 there was published the paper Cartographic e-documents & SGML/XML. This paper recapilutated the possibilities of XML (eXtensible Markup Language) format and its potential applications in digital cartography, concretly for web maps. The situation of XML and cartography interaction has changed through two years. Therefore this paper offers a another review of new and inovated "cartographic" XML and XML related formats or languages and its benefits and advantages. This document is concretly focused on

- storing, coding (spatial data, vector graphics formats, map components describing formats),
- describing (metadata formats),

- transformating (transformation and style languages) of spatial data intended for cartographic visualization.

This paper is not intended for a propagation of concrete technology or format. An implementation and observance of standards is more important. An application of cartographic and information technology standards makes for a high-quality of cartographic products. The high-quality is the cause of better independence, navigation, user orientation, readability, accessability etc. The standardization could assist an unification of relatively disintegrated spectrum of XML formats (e.g. schema languages or metainformation formats).

Keywords: Markup languages, XML, digital cartography, XML schemas, standards.

1 Introduction

At the present time it is turning out that the main problem of digital communication is the vast bulk of digital formats, platforms and mutual compatibility. Cartographers do more and more meet with the questions of heterogeneous standards, formats, platforms, data models, specifications, technical directives, specifications of user requirements etc., too. Language XML (eXtensible Markup Language) shows one feasible solution of the actual state of an electronic communication. This paper should present the integration XML standards to cartographic processes.

2 About XML

XML Language ranks among the group of markup languages or metalanguages, which marks the meaning of particular parts of documents, but does not specify their design. In some publication there is the term "self-describing languages" - the languages, which know to describe data (documents) and their structure, too. Just the needfulness of data interchange between incompatible hardware architectures was the original reason for the development of markup languages.
In the year 1986 the works on a "lite version" of the older markup language SGML (Standard Generalized Markup Language) started. SGML was very popular markup language in the eighties. SGML was standardized by ISO organization (International Organization for Standardization), ISO 8879:1986.
From SGML the vast number of redundant parameters (abstract syntax, e.g. special characters, minimized syntax, labeling of tags, choice of length of tag names etc.) were removed. After this manner XML language came into existence under the direction of W3C organization (World Wide Web Consortium). XML represents a subset of SGML not a substitution of SGML.
The following list summarizes some versions of XML specification:

- 1997 - 1st Working Drafts.
- 10.02.1998 – XML 1.0 W3C Recommendation (definite specification).
- 2001 - 1st Working Draft of XML 1.1.
- 04.02.2004 – XML 1.1. W3C Recommendation.
- 16.08.2006 – XML 1.0 Fourth Edition (4th revision of XML 1.0 standard).
- 16.08.2006 – XML 1.1 Second Edition.

3 XML, semantic web, Web 2.0 and cartography

The terms "semantic web" and "Web 2.0" are frequently mentioned on the present. This section does not focus on "classic" question, if these terms are only modern words or really innovative processes. The aim of this section is to show

- XML as a tool for semantic web and Web 2.0,
- the web cartography and new conception of web relationship.

At first it is needed to define above-mentioned terms:

- The Semantic Web is not a separate Web but an extension of the current one, in which information is given well-defined meaning, better enabling computers and people to work in cooperation. [Ber2001]
- Web 2.0 (the fundamental principles of platform Web 2.0 are detailed in the publication [Ore2005]):
  - The movement from centralized processes or services to decentralized processes or services. [Web], [Skl2007]
  - The conversion of "web of documents" to "web of services" – Web 2.0 should be the platform for data sharing, where an user control data originated from separate sources. [Web]
  - The sequence of some relatively independent technological changes on the web and the movement from "passive consummation" of the web to active usage of web space. [Cim2005]

In agreement with [Ose2002] there are ranked among the parts of semantic web these following technologies: XML, XML Schema, XML Namespaces, XHTML (eXtensible HyperText Markup Language), SVG (Scalable Vector Graphics), SMIL (Synchronized Multimedia Integration Language), RDF (Resource Description Framework), RDF Schema, DAML + OIL (DARPA Agent Markup Language, Ontology Integration Language) and MyOwnML (an own XML application).
The fundamental elements of Web 2.0 platform are in agreement with [Skl2007] and [Ore2005]: tags (content marked by users, so-called folksonomy, collaborative tagging, social classification, social indexing, social tagging etc.), social relations, RSS (Really Simple Syndication; content syndication), Long tail (self-service customer processes, e.g. RSS, blogs or Google AdSense), Blogs, Wikipedia (contain is created by users), program modules for data processing (e.g. AJAX /Asynchronous JavaScript and XML/ or web interface for an access of different sources applications) and web services.
Web 2.0 in agreement with some experts represents just the one phase of development of network communication. It starts to talk about platforms Web 3.0 (the connection of semantic web and Web 2.0 [Skl2007]) and Web 4.0. It is interesting to cartographers, that XML scheme SVG, which is very often used for the creation of web maps (see [Čer2006b]), is one of fundamental parts of next generation of the web along with other technologies (e.g. Artificial intelligence, Automated reasoning, Cognitive architecture, Composite applications, Distributed computing, Human-based genetic algorithms, Knowledge representation, Ontology, Semantic Web, Semantic Wiki or Software agents).
Web 2.0 and semantic web have much in common. The both technologies try to share and reuse data within the scope of different communities. An effort to harmonize data is common to both platforms. Both approaches are complementary each other – Web 2.0 represents platform or technology and semantic web represents semantics. Semantic blogging, semantic wiki, semantic desktop or working with folksonomy or ontology are considered as a example of a combination of Web 2.0 and semantic web.

The paragraphs above described both platforms show that XML and related technologies are the fundamental of semantic web and Web 2.0. XML can produce the following benefits:

- The data is self-describing.
- The data can be manipulated with standard tools.
- The data can be viewed with standard tools.
- Different views of the same data are easy to create with style sheets. [Har2001]

The connection between cartography and modern approaches to web is widely discussed. Like some examples there are two paper published on the XXIII International Cartographic Conference (4.-10.8.2007, Moscow, Russia) [Zen2007], [Car2007] – the paper of the current president of ICA Professor William Cartwright.
On the Internet there are some examples of usage Web 2.0 principles [Zen2007]:

- Project WikiMapia (www.wikimapia.org), which combine advantages of Google Maps (maps.google.com) and wiki systems.
- Placeopedia (www.placeopedia.com) – the other type of an integration of Wikipadia (en.wikipedia.org) and Google Maps.
- Blogs with cartographic themes: geobloggers (www.geobloggers.com), Mapperz (mapperz.blogspot.com) or Map the Universe (www.maptheuniverse.com).
- Google Maps or Google MyMaps represents the cartographic type of very popular mushup applications (other examples of mushup applications: eBay, Amazon, Flickr, del.icio.us, YouTube).

4 Review of XML applications in cartography

The list of XML technologies applicable in a work of cartographers is the last part of this paper. This list is an upgrade of the paper "Cartographic e-documents & SGML/XML" [Čer2006a], which was published during the international symposium GIS... Ostrava 2006. This review of XML technologies could serve as a source of information on the XML. The goal of this section is also to present the breakthrough of markup languages and digital cartography over the last years.
The use of XML in cartography is very wide. This document is concretely focused on three fundamental groups:

- Storing, coding (spatial data, vector graphics formats, map components describing formats),
- describing (metadata formats schemas),
- transformations (transformation and style languages) of spatial data intended for cartographic visualization.

Data storing and coding
A contemporary digital map is not a product, which stay in itself. The map is composed of several components, which form a complete digital cartographic product. Accordance with [Čer2006a] there are cartographic elements, text elements, multimedia elements, interactive elements and metadata elements. Every previous item could contain some different parts – e.g. multimedia parts of a digital map could be formed from graphic files, sound files or video files. XML could be applied for the coding of all parts of a digital map. Only a coding of raster graphic components is very complicated and not very efficient. Except the coding of a map or some component of a map XML could be used as a format for data to map generating. The last possible of using of XML in digital cartography is an application of XML as a format

for cartographic processing (e.g. office document, project management, process diagrams etc.).
The following list shows formats (languages, technologies, applications) based on XML language and other related technologies applicable in cartographic project (information sources: [Čer2006a], [Čer2007], [Sal2007], [dev]):

1. Basic XML technologies
    1. XML, Binary XML
    2. XML Namespaces, XML Catalogs.
    3. Odkazy: XLink (XML Linking Language), XML Base, XPointer (XML Pointer Language), XBL (XML Binding Language).
    4. Abstract models for XML documents: XML Information Set (XML Infoset), XPath Data Model, DOM (Document Object Model), XQuery 1.0 and XPath 2.0 Data Model.
    5. Describing characters in different natural languages of the world – Unicode, ISO/IEC 10646 (International Organization for Standardization / International Electrotechnical Commission).
    6. Denoting character sets: IANA-CHARSETS (Internet Assigned Numbers Authority Character Sets).
    7. Denoting languages and countries: IETF (Internet Engineering Task Force) RFC 1766, ISO 639, ISO 3166.
    8. Identifying Internet resources: IETF RFC 2396, IETF RFC 2732.
2. Text and hypertext data
    1. Office application formats: ODF (OpenDocument Format; ISO/IEC 26300, guaranteed by OASIS /Organization for the Advancement of Structured Information Standards/; open source software Open Office, from the version 2.0), OOXML or OpenXML (Office Open XML; Ecma 376 standard; Microsoft's office software), Mars (XML-based representation of PDF /Portable Document Format/ documents; Adobe software).
    2. DocBook schema is intended for writing (coding) of a technical documentation. DocBook documents could be transform via style languages to many standard formats (RTF /Rich Text Format/, PDF, HTML, Postscript, TeX etc). DCF (Document Composition Facility) is a format similar to Docbook, but DCF is more older and based on GML (Generalised Markup Language, the predecessor of SGML).
    3. Web publishing languages (formats applicable for creating of web pages): HTML XHTML, XForms, XFrames, XML Events (to represent asynchronous occurrences, such as mouse clicks, in XHTML or in other XML markup langugaes), WebDoc, DIAL (Device Independent Authoring Language; for writing web pages that can be presented by a range of devices).
    4. Languages for mobile devices: WML (Wireless Markup Language), HDML (Handheld Device Markup Language).
    5. Format for presentation slides – Slide ML (Slide Markup Language).
    6. Formats for a describing of special text stuctures – MathML (Mathematics Markup Language), CellML (a language describing mathematical models), CML (Chemical Markup Language) etc.
3. Graphical data
    1. Vector graphics – SVG is the best known vector graphics format based on XML. But the group of similar formats is more wider: VML (Vector Markup Language), WebCGM (Web Computer Graphics Metafile), DrawML (Drawing Meta

Language) or the unfinished PDF- and Postscript-based format PGML (Precision Graphics Markup Language).

2. Raster graphics. Presently it is not available any standardized or widely dispersed XML format used for describing of raster graphics. The reason why these formats are missing, is large size of raster graphics data files, which need a binary form for their processing.
   In the University of Washington there was created the format IMT (Image Markup Language) applicable for storing of text comments on GIF (Graphics Interchange Format) and JPEG (Joint Photographic Experts Group) files. This format is used above all in public health.
3. 3D graphics – X3D (the alternative to the very popular format VRML /Virtual Reality Modelling Language/).

4. Multimedia data
   1. The formats SMIL (Synchronized Multimedia Integration Language) and SMIL Animation make for a connection between XML applications and binary multimedia files.
   2. VoiceXML and CCXML (Call Control eXtensible Markup Language) a standard designed to provide telephony support to VoiceXML represent XML languages used for a spoken langugage processing.
5. Geodata (spatial data, geospatial data)
   1. GML (Geography Markup Language) guarantee by Open Geospatial Consortium (OGC) represents the base formats for the storing and describing of geospatial data.
   2. Other less common formats describing common geodata:
      1. cGML (compact GML), the CRS4 (Center for Advanced Studies, Research and Development in Sardinia) standard is derived from GML. cGML is applicable above all for mobile devices (shortcut tags, server side computing).
      2. JML (JUMP GML) represents the simplified variant of GML. This format is used in JUMP and OpenJUMP software.
      3. KML (Keyhole Markup Language) is in principle the copy of GML 1.0 and 2.0 designed for Google Earth. This format describes 3D spatial data.
      4. LandXML is the language for describing and sharing the architecture, surveying, urban planning, civil engineering data.
      5. GPX (GPS eXchange Format) is an XML schema designed for transferring GPS data between software applications. It can be used to describe waypoints, tracks, and routes.
      6. Other formats: G-XML, SOTF (Spatial Object Transfer Format), TML, (Terrain ML), cityGML etc.
   3. Formats describing of some specific type of geodata applicable for creating of thematic maps: OMF (Weather Observation Definition Format), XMML (eXploration and Mining Markup Language), NVML (NaVigation Markup Language), CaveMap DTD (format used in speleology), EML (Ecological Metadata Language), MayDay ML (format for data sharing in risk management) etc.
   4. Formats described the digital maps or some items of digital maps:
      1. The ESRI company uses XML derivates (ArcXML. ArcPadXML) in its products. These formats are instrumental towards map composition definition or data processing. [Jed2004]
      2. Some formats like MDML (Map Description Markup Language) or DiaML (Diagram Markup Language) describe some parts of maps (e.g. diagrams).

Data describing

These formats make for the "new" XML definition. XML schemes formats describe elements, attributes, their connections, data types, and other rules for a designing XML documents.

1. Primarily it is necessary to introduce the quaternion of the most used XML schema languages: DTD (Document Type Definition), XML Schema, RELAX NG (Regular Language Description for XML - Next Generation) a Schematron.
2. But the set of schema languages is very good-sized. The following list shows as an illustration another format for data describing (the majority of these format does not assert oneself presently): RELAX (Regular Language Description for XML), TREX (Tree Regular Expressions for XML), XML-Data, XDR (XML-Data Reduced), DCD (Document Content Description), DSD (Document Structure Description), SOX (Schema for Object-oriented XML), DDML (Document Definition Markup Language, formely XSchema), XDuce, DTD++, DTD4DTD (Data Types for DTD), Examplotron, Hook. [Vli2005]
3. DSDL (Document Schema Definition Languages) can be seen as a framework and set of languages to check the quality of XML documents. Because the validation of XML documents is a subject too wide and complex to be covered by a single language. It also acknowledges that the industry needs a set of simple and dedicated languages to perform different validation tasks - as well as a framework in which these languages may be used together. DSDL could thus be a technology which is absolutely indispensable for most of XML applications. [Vli2003]
4. Metadata formats and formats for semantic web (these group is also very extended and hardly classifiable): RDF (Resource Description Framework), OWL (Web Ontology Language), DCMI (Dublin Core Metadata Initiative), METS (Metadata Encoding & Transmission Standard), MODS (Metadata Object Description Schema), MARC XML, P3P (Platform for Privacy Preferences), PICS (Platform for Internet Content Selection), TEI (Text Encoding Initiative), DAML + OIL, Topic Maps etc.

Data transformation (XML Transducers)

1. CSS (Cascading StyleSheet, kaskádové styly) is very simple style language used in relation to HTML or XHTML for the definition of visualization rules.
2. XSL (eXtensible Stylesheet Language) represents the tool for XML transformation to other not only XML formats. XSL is composed of
    1. XSLT (eXtensible Stylesheet Language Transformation),
    2. XSL-FO (eXtensible Stylesheet Language - Formatting Objects).
3. DSSSL (Document Style Semantics and Specification Language), FOSI (Formatted Output Specification Instance), XPP (XML Professional Publisher) or STX (Streaming Transformations for XML) group some older or less common XML transducers.
4. Query languages: XPath (XML Path Language), XQuery, Ouilt.
5. Canonical XML defines a process to create a specified physical representation, a canonical form, to an XML document or a document subset.
6. XML Fragment Interchange language includes capabilities to specify a part of a whole XML document as a fragment to be sent to a receiver.
7. XInclude is a language for specifying merging of a set of XML documents represented as Infosets to a new Infoset.

5 Conclusion

This paper contains 111 technologies, formats and languages based on XML. This list does not represent complete and close set of possibilities XML application in cartography. Cartographers could use many other XML technologies in their branch:

- The large number of local proprietary formats used by concrete software products exists for a geodata describing.
- In the field of creation of web pages there are the initiatives of WHATWG (Web Hypertext Application Technology Working Group) organization except W3C standards.
- XML formats are used as communication standards in the sphere of web services - standards WMS (Web Map Service), WFS (Web Feature Service), WCS (Web Coverage Service), WPS (Web Processing Service) etc.
- XML is very important to e-business, which infiltrates more and more into digital cartography, too.
- Administration, control, planning and running the processes represents the other possibility of XML usage in cartography – formats PSLX (Planning and Scheduling Language for XML), XProc (XML Pipeline Language), XMI (XML Metadata Interchange) etc.
- The special format for the interchange of scientific information also exists – XSIL (Extensible Scientific Interchange Language).

XML and markup languages in general represents very useful tool, which enables to integrate inputs and outputs of different phases of the creation of a digital map. It is necessary to awake to hazards consequent upon the main property of XML. Extensibility of XML format makes possible to create an infinity of XML subsets or more precisely an infinity of new markup languages. A mutual communication between new created, not standardized and poorly described formats make difficult for the accessing and sharing of decentralised informations. That is why it is necessary to put the accent on

- an integration of existing XML schemas and adding of these schemas instead a production of new application,
- a full and exact description of new schemas (including a standardization), which makes easy a sharing and implementing of XML languages to own software products.

Bibliography

[Ber2001] Berners-Lee, T., Hendler, J., Lasila, O. The semantic web. A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities [online]. Scientific American, 2001, roc. 284, s. 35–43. [cit. 2007-10-12]. Resource: http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21.

[Car2007] Cartwright, W., Miles, A., Morris, B., Vaughan, L., Yuille, L. Affective Atlas – Constructing an Atlas Using Web 2.0 and Social Software. In XXIII International Cartographic Conference, 4.-10. August 2007, Moscow, Russia. International Cartographic Association, 2007.

[Cim2005] Cimprich, P. Akta X 0509 [online]. In ROOT.CZ, 10.10.2005. [cit. 2007-10-12]. Resource: http://www.root.cz/clanky/akta-x-0509/.

[Čer2006a] Čerba, O. Cartographic e-documents & SGML/XML. In International Symposium GIS... Ostrava 2006. Ostrava: Vysoká škola báňská - Technická univerzita, 2006. [cit. 2007-10-12]. Resource: <http://gis.vsb.cz/GISEngl/Conferences/GIS_Ova/GIS_Ova_2006/Proceedings/Referaty/default.htm>.

[Čer2006b] Čerba, O. SVG v kartografii [online]. In Geoinformatics FCE CTU 2006. Praha: 2006. ISSN 1802-2669. [cit. 2007-10-12]. Resource: <http://geoinformatics.fsv.cvut.cz/wiki/index.php/SVG_v_kartografii>.

[Čer2007] Čerba, O. XML Technologies for Cartographers. In XXIII International Cartographic Conference, 4.-10. August 2007, Moscow, Russia. International Cartographic Association, 2007.

[dev] developersWorks: XML: Standars [online]. IBM. [cit. 2007-10-12]. Resource: http://www.ibm.com/developerworks/views/xml/standards.jsp.

[Har2001] Harold, E. R. XML Bible. 2. vyd. New York: Hungry Minds, 2001. 1015 s. ISBN 0-7645-4760-7.

[Jed2004] Jedlička, K. Tipy a triky pro ArcPad. In ArcRevue. 2004, roč.13, č.1, s.17-19, ISSN 1211-2135.

[Ore2005] O'Reilly, T. What Is Web 2.0 [online]. Sebastopol: O'Reilly, 30.9.2005. [cit. 2007-10-12]. Resource: http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html.

[Ose2002] Ossenbruggen, J. van, Hardman, L., Rutledge, L. Hypermedia and the Semantic Web: A Research Agenda [online]. In Journal of Digital Information, Volume 3 Issue 1 Article No. 55, 2002-05-17. [cit. 2007-10-12]. Resource: http://jodi.tamu.edu/Articles/v03/i01/VanOssenbruggen/?printable=1.

[Sal2007] Salminen, A. XML Family of Languages. Overview and Classification of W3C Specifications [online]. 18.5.2007 [cit. 2007-10-18]. Resource: <http://www.cs.jyu.fi/~airi/xmlfamily.html>.

[Skl2007] Sklenák, V. Web 2.0 vs. sémantický web [online]. In INFORUM 2007: 13. ročník konference o profesionálních informačních zdrojích, Praha 22.-24. května 2007. Praha: Albertina icome Praha, 2007. [cit. 2007-10-12]. ISSN 1801-2213. Resource: http://www.inforum.cz/pdf/2007/sklenak-vilem.pdf.

[Vli2003] Van der Vlist, E. ISO DSDL Overview [online]. In XML Conference & Exposition. Philadelphia: Pennsylvania Convention Center. 7.-12.12.2003 [cit. 2007-10-12]. Resource: http://www.idealliance.org/papers/dx_xml03/papers/04-03-03/04-03-03.html.

[Vli2005] Van der Vlist, E. XML schema languages compared [online]. In XML Prague 2005. Praha: Institut teoretické informatiky, 25.6.2005 [cit. 2007-10-12]. Resource: htpp://dyomedea.com/papers/2005-xmlprague.

[Web] Web 2.0 [online]. [cit. 2007-10-12]. Resource: http://suewebik.net/webove/web20-def.html.

[Zen2007] Zentai, L. Application of Web 2.0 in Cartographic Education: Is It Time for Cartography 2.0?. In XXIII International Cartographic Conference, 4.-10. August 2007, Moscow, Russia. International Cartographic Association, 2007.