THEMATIC MAPS ON THE WEB

Otakar Čerba
University of West Bohemia in Pilsen
Univerzitní 22, Pilsen, 306 14, Czech Republic.
ota.cerba@seznam.cz

The author was supported by the project VisualHealth (2E08028)

Abstract

Thematic maps represents very useful and intelligible presentation of statistical data with geospatial linkage. Anybody can see the thematic maps on TV news or TV documentaries, in newspapers or magazines, subject publications etc. daily. Thematic maps are in used in the field of economy, demography, transport, physiography, tourist trade, international relationships, landscape management etc. The Internet represents the major information source for many people. Therefore on the Internet there is a large number of different types of thematic maps. It is very surprising, that the majority of these maps is coded in raster form, although some vector or hybrid format would be preferable for many reasons.

One of the activities of Geomatics section of the University of West Bohemia in Pilsen (Czech Republic) is focused on possibilities of a presentation of vector thematic maps on the Internet and methods of creation and description of these maps. As an objective of our activities there is the creation of a system in terms of a library of XSLT (Extensible Stylesheet Language Transformations) templates. These templates in conjunction with control file provide the generating of different types of thematic maps (e.g. choropleth maps or maps with proportional symbols) in a vector format being able to cooperate on the Internet. Now there are three formats under consideration – SVG (Scalable Vector Graphics), VML (Vector Markup Language) or KML (formerly Keyhole Markup Language).

The fundamental principle of our approach results from the methods of style languages using. This methods are based on XML (Extensible Markup Language) technologies and related applications (e.g. script languages or interfaces) making possible to convert separate formats or languages. In this case there is the transformation from source geospatial data format (e.g. Geography Markup Language) to some above-mentioned vector formats making for a presentation of thematic maps on the web. This conversion is realised by the new version of style language XSLT and query language XPath (XML Path Language). XSLT styles are processed by simple software product called XSLT processor. For our purposes there were used two types of these processors – external processor Saxon (version 9.X) or processors inbuilt in web page browsers. The

visualisation of created maps is managed by the the software Google Earth or by common web page browsers supported implemented standards. In browsers there is used an existing interface (e.g. Google Maps API) or own simple interface created in SVG.

The results of transformations, pros and cons of these processes are described and summarized in following paper. Except the creation of transformation templates we try to develop the data model necessary for a description of separate maps. This model is written in RELAX NG (Regular Language for XML Next Generation) format which comes within the ambit of schema languages. It could be used not only for map generating but also as a common exchange format of cartographic products. The XSLT styles were applied to the preparation of printed publication Atlas of International Relationships. On the present there are generated sets of thematic maps designed for the project VisualHealth focused on a visualization of health data by way of cartographic interpretation methods.

XML or markup languages in general terms represent one of the way of elimination of many heterogeneities in cartography. These heterogeneities embarrass the interoperability and mutual communication in cartography. The main goals of languages based on XML are high level of standardisation (XML standards are managed above all by generally accepted international organizations as the World Wide Web Consortium or Open Geospatial Consortium), independency (due to coding mode and strict abidance of rules is XML independent on hardware platforms, programming languages or operational systems) and large spectrum of existing applications, including cartographic applications.

Keywords: Thematic maps, cartographic visualization, XML, XSLT, web.

Introduction

Thematic map concentrates on distribution of a single attribute or the relationship among several [10]. When do you want to search some thematic maps through any search engine, you get a huge number of results. The different types of thematic maps (e.g. choropleth maps, cartograms, maps with isolines etc.) are very popular. They represent very simple, but very sophisticated tool making possible the transformation of complicated language of numbers (in the form of statistical data, geographic or cartographic coordinates) to simple graphical language, which is intelligible to almost everyone without exception. The another goals of the thematic maps is connected with dual view on the map. The user can see the global situation, but also can study and compare details. Both advantages are clear at the view on pages and screens of mass media. Thematic maps appear in newspapers, TV news, documentaries, Internet news etc. In these media there are a large number of map products every day. There is the important problem – the majority of thematic maps on the Internet is represented by raster graphics. But just the thematic maps are very often associated with vectors,

because the majority of the thematic maps is designed from the simple vector components. This paper is focused on the generating of vector thematic maps on the Internet. The proposed method is based on XML (Extensible Markup Language) and related technologies. The main reasons of an interconnection of digital cartography and XML technologies are a possibility of

- the extension of an using of wide spectrum of cartographic interpretation methods (cartography have many different tools and methods for geospatial data visualization, but the majority of current software products are unable to exploit the cartographic capabilities),
- the description of geospatial data, transformation templates and vector graphics by XML formats.

The following list introduces the structure of this paper:

- XML solution – principles and reasons: Description of XML and the principles of generating of the thematic maps by XSLT styles, examples of possible formats and technologies, pros and cons of this solution and examples of this approach.
- Formalized description of the maps: Reasons of necessity of this description, examples of existing solutions.
- Our solution: Description of our application called Jigsaw maps, source and output files and structure of control file, examples of generated maps.
- Conclusion: The relationships of this solution and other technologies, standards etc., next steps of development.

XML solution – principles and reasons

When does anybody make decision to present thematic maps on the Internet, there is the question: "Which vector formats will be used for these maps?". Many vector formats are connected with concrete desktop application (e.g. Shapefile, DXF or DWG) and their distribution on in the Internet is very complicated or impossible. For a designing of accessible cartographic application we must choose the format according to following criteria:

1. The format must support the 2D (or 2,5D or 3D) vector graphics. Reasons are mentioned in previous sections.
2. The format must be accessible through the web page browsers, because this software tools represent the most common and used clients on the Internet. The majority of users works with this software, they do not need any special education. Web page browsers are very often free and they are the component of every computer.

The majority of formats selected in accordance with previous items has the one common attribute. Main vector formats on the web (except Adobe Flash) are based on XML. Main official standardized geodata format, Geography Markup Language (GML) are based on the XML too. The source data (geoinformation) and target data (maps) are

based on the common platform. It would be very useful to manage the transformation (or creation of maps) between source and target data by XML too. Currently there is at disposal very powerful tool – the style language called XSLT (Extensible Stylesheet Language Transformations) 2.0 and related technologies (e.g. XPath, XML Information Set, Document Object Model etc.). The following sentence published by Miloslav Nic (the author of the portal ZVON.org – the Guide of the XML Galaxy) in the publication XSLT 2.0 Tutorial [7] characterizes the power and possibilities of style languages, especially the second generation of XSLT – "I have started to use XSLT 2.0 as my primary programming language (in combination with Python) and I am amazed by its power". [7]

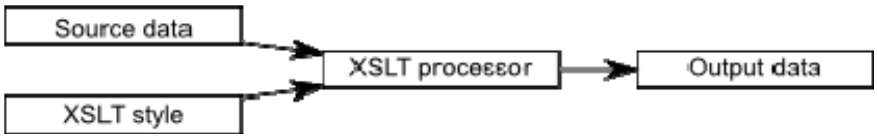

Figure 1. The basic principle of processing XML files through XSLT.

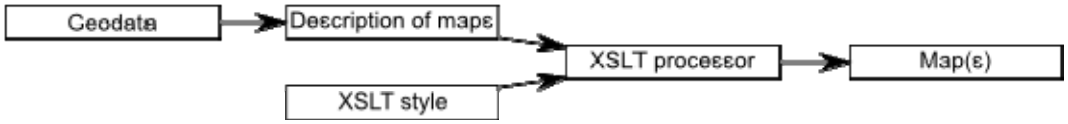

Figure 2. The modified principle of processing XML files through XSLT focused on maps nad geospatial data.

For described application focused on generating of thematic maps and based on XML formats and languages the suitable formats must be selected. Some formats were mentioned above. The following list based on the publications [2] and [3] shows formats used in following application:

1. Formats for describing and coding of geospatial data – GML, JML (JUMP GML)
2. Formats for describing and coding of vector graphical data – SVG (Scalable Vector Graphics), KML (Keyhole Markup Language)
3. Transformation and style languages (XML Transducers) – CSS (Cascading Style Sheets), XSL (Extensible Stylesheet Language) – this language is composed from languages XSLT and XSL-FO (Extensible Stylesheet Language - Formatting Objects)

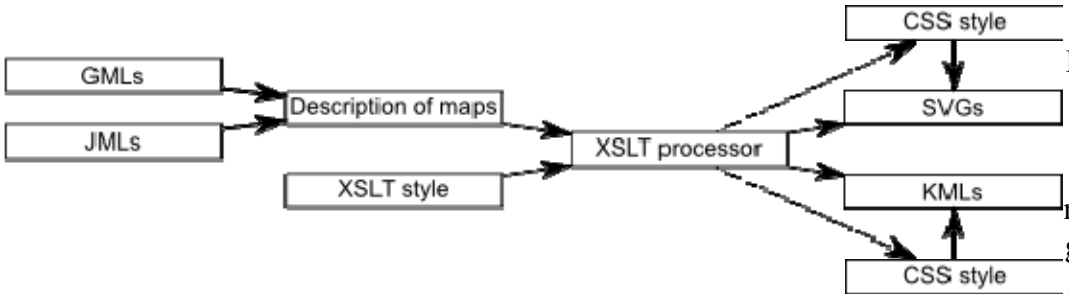

Figure 3. The schema of generating thematic maps through XSLT templates.

For the transformation from geodata files to the maps there is necessary to use the software tool called transformation of XSLT processors. They transcribe source data to output data. XSLT processors are divided into external processors (e.g. Saxon) or processors in-built in web page browsers. Unfortunately just the in-built processor have very poor support of second version of XSLT, which contains the new attributes and elements necessary for generating of the maps (e.g. mathematical components).

The described approach has many benefits, these benefits are based on using common XML, style languages, vector formats (above-mentioned), concrete applied formats (GML, SVG, XSLT etc.). There are several advantages [4], [6], [9]:

- 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.
- The capability to convert XML documents to different media: the Web, print, and more.
- The emphasis on structure means that XML documents are better equipped to withstand the test of time (structure is more stable than formatting).
- Maximum error checking.

The main advantages of using of XML a related technologies are the simplicity [1] and the using of one type of technology and one principle. With respect to above-mentioned characteristics it is possible to say, that XML eliminates the different heterogeneities in digital or web cartography. We can paraphrase the book “Lord of Rings” and write “One format rules them all”.

Formalized description of the maps

In the previous chapter there were briefly described all components except the control file. This file is the most important, because it defines the structure of the map and attributes and parameters of the map as well. For purpose of the design of the control file the language of format intended for formalized description of the maps should be

suitable. But presently there is not any expanded format describing the structure of the maps. Therefore there are three possibilities to create the control file:

1. Creation of own schema or own structure of control file.
2. Exploitation existing proposals of formats, e.g. Map Description Markup Language (MDML) or Map Overlay Ontology (http://www.daml.org/ontologies/177).
3. Combination of own schema, existing proposals and formats describing parts of maps, e.g. Diagram Markup Language (DiaML; http://www.carto.net/schnabel/mapsymbolbrewer/schemas/diaml.xsd) – XML based language for the description of map symbols.

The first variant was chosen for our purposes. But we expect the transition to third variant or the implementation of newly-formed formalized description of the maps or their components. We suppose that there are the following aspects important to the design of formalized description of the maps:

- Technology (format or language): We used the language RELAX NG (Regular Language for XML Next Generation) as the technology of concrete (physically) data model. The RELAX NG format is very clear and brief. It uses data types and makes possible to transform to other schema languages – the program Trang makes possible the transformation of rnc file to the other form of RELAX NG or to the other schema format (W3C XML Schema and Document Type Definition). Last but not least there is another important fact – RELAX NG is the part of standard DSDL (Document Schema Definition Languages; www.dsdl.org) designed by ISO (International Organization for Standardization).
- Some ontology format (e.g. OWL – Web Ontology Language) should be used for the building of conceptual data model.
- Classification scheme: In our purpose project the classification is based on the works of Jan Pravda [8].
- Users (including the international organizations and software producers) using and implementing the proposed structure. These subjects can managed, control and promote the contingent standard.

The design of formalized description of the maps could produce following benefits:

- Automatic generating of the data visualization.
- Sharing and exchange of the map (between different systems, software, countries etc.).
- Automatic (or self) control – validation.

Our solution

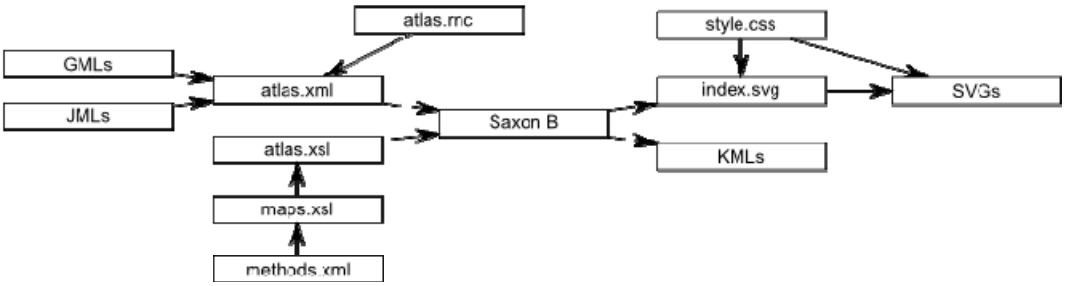

Figure 4. File structure of proposed application.

When we do not include the geospatial data files linked to control file (atlas.xml), our system contains five fundamental source files:

- atlas.xsl – the main style file defining just the color schemes based on ColorBrewer (www.colorbrewer.org), parameters of output files (e.g. encoding) and link to the file maps.xsl.
- maps.xsl – this part provides the design of menu, start page (index.svg), environment of maps and link to the file methods.xsl.
- methods.xsl – these templates make possible the generating of concrete maps.
- atlas.xml (the name of this file is not strictly predefined) – this file is described in the next paragraph.
- atlas.rnc – the schema file defining the formalized description of the maps. This file contains allowed elements, attributes, interconnections, data types and other limits.

The following code shows the example of control file. This concrete file describes the map made up three overlapping components. These components are designed from three different source data files. For the map are used different methods and graphical parameters. The description of particular rows of code is written in comments (<!-- comment -->) under respective line.

```
<?xml version="1.0" encoding="UTF-8"?>
<!-- Header of XML file containing the version of XML and used coding
(in accordance with standard Internet Assigned Numbers Authority
CHARSETS).  -->

<Atlas title="Atlas (Prototype)">
<!-- Root elements of each atlas or group of maps with one compulsory
attribute – the name. -->

 <Map title="Czech Republic">
<!-- Basic element of each map with one compulsory attribute – the
```

```
name. -->

  <Method name="QualArea" fill="#1E90FF">
<!-- Description of cartographic method – its name and parameters, in
this case the qualitative areas with the the color of fill are used. -
->

   <Data file="Gen_Data/kraje1_gen.gml"/>
<!-- Link to visualised geodata file. -->

  </Method>
  <Method name="QualArea" border="navy" border-multiple="2">
<!-- Description of cartographic method – its name and parameters, in
this case the qualitative areas with the the color and width of border
are used. -->

   <Data file="Gen_Data/okresy_gen_vyber.gml"/>
  </Method>
  <Method   name="QuantArea"   parameter="ogr:OB01"   colors="Quant1"
classing="Quantils" classes="6">
<!-- Description of cartographic method – its name and parameters, in
this case the choropleth maps with six intervals divided into quantils
and the selected color schema are used. -->

   <Data file="Gen_Data/orp_gen_vyber.gml"/>
  </Method>
 </Map>
</Atlas>
```

The output in SVG form is composed of the file index.svg (the SVG page containing the menu, list of maps, links to maps and simple information – creation date, number of maps) and map files. All SVG files share the CSS file describing presentation attributes of map controls. The KML output is represented just by the separate KML files with internal or inline styles.

Except the benefits mentioned in previous chapters (goals of XML or markup languages and advantages of formalized description of the maps) there are some other goals. The using of existing standards and interconnection of standards of different organization (W3C – World Wide Web Consortium and OGC – Open Geospatial Consortium) is very important. Because the existing standards are very detailed described and they are implemented in software products, existing data set or legislative rules. In general the usage of existing standards, technologies, formats etc. eliminates the heterogeneities and redundant items. It could lead to an improvement of effectiveness and cost reduction.

These paper (with respect to required range) cannot contain the samples of output maps. These samples will be inserted into presentation and on-coming web page on the domain git.zcu.cz.

Conclusion

The application of XML and related technologies in digital cartography or in geoinformatics is common with contemporary map users. Many people use for example the web services (e.g. Web Map Service, Web Feature Service, Web Coverage Service etc.) today. Also the users of ESRI products work with XML – the language ArcXML is important for creating of map compositions. The above-mentioned format KML is associated with very popular software Google Earth and many other clients (e.g. OpenLayers). XML technologies are connected with the future of information sciences including geoinformatics and digital cartography, too. XML technologies represent the fundamental platform of next evolutionary phases of web – Web 2.0, Web 3.0 etc. are based on XML and related technologies. This application and above all the formalized description of the map is very important in the term of principles of semantic web, which represents one of the fundamental activities of W3C – W3C Semantic Web Activity [5]. It is possible to search other connections in the activities of organization International Cartographic Association (ICA), which associates cartographers in the whole world. ICA Research Agenda [11] represents the very important document of ICA describing the next development of cartography and the possibilities of cartographic research. In this document there are emphasized terms like interoperability, accessibility, web technologies, cooperation with other branches of science, data harmonization, semantics, metadata, visualization models, etc. The technologies, methods and principles mentioned in this paper are relating to this documents. Described application support mainly these aspect of ICA Research Agenda focused on web technologies and standards – e.g. geoinformation (semantics, ontologies) or usability and accessibility of the maps (user centred design). In some cases (e.g. web technologies, metadata or semantics) XML and related technologies represent the fundamental component.

Described approach was used in project of the printed publication Atlas of International Relationships [12] (presented in XXIII. ICA Conference, Moscow) which was created by external XSLT styles. Now the proposed application is used and supported by the project VisualHealth (original title in Czech – Vizualizace zdravotnich dat pro podporu interdisciplinirniho vzdelavani a vztahu s verejnosti; complete title in English – The Visualization of Health Data for the Support of Interdisciplinary Education and Relation with Public; the project code 2E08028). This project is developed within the framework of Programme 2E – Human sources (2006-2011) of Ministry of Education, Youth and Sports in the Czech Republic. In this project there are cooperating three partners from the Czech Republic – the University of West Bohemia in Pilsen, Masaryk University in Brno (the leader of the project) and Faculty Hospital Brno. The project VisualHealth is focused mostly on cartographic visualization, on creating of different types of thematic maps of public health data for education, their presentation and popularization. Simple insight through easy comprehensible maps to this very complicated data could lead to a higher level of prevention and protection of health. Within the framework of this project there are processed different cartographic methods of geospatial data visualization on

the Internet. One type of data visualization in the VisualHealth project was based on XML source data, XSLT templates and SVG or KML graphics.

The author of this paper suppose that with respect to above-mentioned reasons the further development of this application will continue after finishing of the project VisualHealth. The following list offers the possible steps of next development:

- The building and designing of user interface making for the education and modification of control file describing the maps.
- The background research of classification systems.
- The formation of new cartographic interpretation methods.
- The optimization of actual XSLT templates.
- The finalization and optimization of data model written in RELAX NG.
- The conversion of desktop application to the server application making possible online map generating.

References

1. CERBA, O. 2006. Cartographic e-documents & SGML/XML. In GIS Ostrava 2006 Informatics for geoinformatics. Ostrava: VSB-TUO, 2006. ISSN 1213-239X. Online: <http://gis.vsb.cz/GISEngl/Conferences/GIS_Ova/GIS_Ova_2006/Proceedings/Referaty/cerba.html>.
2. CERBA, O. 2008. Application of Markup Languages in Cartography. GIS... Ostrava 2008, 15th year of international symposium. ISBN 978-80-254-1340-1. Online: <http://gis.vsb.cz/GIS_Ostrava/GIS_Ova_2008/sbornik/Lists/Papers/012.pdf>.
3. CERBA, O. 2009. Thematic Maps in Browsers. In GIS Ostrava 2009. Ostrava: Tanger spol. s r. o, ISBN 978-80-87294-00-0.
4. HAROLD, E. R. 2001. XML Bible. Second Edition. Hungry Minds, 2001. 1206 p. ISBN 0-7645-4760-7.
5. HERMAN, I. 1994-2009. W3C Semantic Web Activity [online]. W3C. Online: <http://www.w3.org/2001/sw/>.
6. MARCHAL, B. 2000. XML by Example. Oue, 2000. 505 p. ISBN 0-7897-2242-9.
7. NIC, M. 2005. XSLT 2.0 Tutorial. Zvon.org, Online: <http://www.zvon.org/xxl/XSL-Ref/Tutorials/index.html>.
8. PRAVDA, J. 2006. Metody mapoveho vyjadrovania: klasifikacia a ukazky. Bratislava: Geograficky ustav SAV, 2006.
9. RAY, E. T. 2003. Learning XML. Sebastopol: O'Reilly & Associates Inc., 2003. ISBN: 0-596-00420-6.
10. ROBINSON, A. H. et al. 1995. Elements of Cartography. 6th ed. John Wiley & Sons. 674 p. ISBN 0-471-55579-7.
11. VIRRANTAUS, K., FAIRBAIRN, D. 2007. ICA Research Agenda on Cartography and

GI Science. In XXIII International Cartographic Conference, 4.-10. August 2007, Moscow, Russia. International Cartographic Association.

12. WAISOVA, S. et al. 2007. Atlas mezinarodnich vztahu. Plzen: Vydavatelstvi a nakladatelstvi Ales Cenek, s.r.o., 2007. ISBN 978-80-7380-015-4.