Geoinformatics 2004
Proc. 12th Int. Conf. on Geoinformatics − Geospatial Information Research: Bridging the Pacific and Atlantic
University of Gävle, Sweden, 7-9 June 2004

MAP INTERFACE VALID COVERAGE ANALYSIS BASED ON XML METADATA

Hsiung-Peng Liao and Jung-Hong Hong
Dept. of Geomatics National Cheng-Kung University #1, University Road, Tainan, Taiwan, Republic of China
darkness@sv.ncku.edu.tw, Tel: 886(6)2757575 ext 63837, Fax: 886(6)2375764

Abstract

May overlap is one common GIS operation available from almost any GIS software nowadays. While users may add map layers of different resources into map window literally, we have argued it is rather dangerous for naive users to treat the content of the map window as a map without careful investigation. In our previous research, we suggested a "valid coverage" analysis mechanism that can automatically interpret the valid coverage of a map window when multiple map layers are selected. Since the valid coverage analysis must depend on the precise description of spatial coverage of every map layer, the best way is to take advantage of current metadata standards. By doing so, the valid coverage analysis of any selected map layers can be automatically executed as long as their metadata complying with current metadata standards. To further expand the possibility of data interoperability in the internet, we have extended our previous prototype to incorporate current XML-based metadata and introduce related geographic object descriptions from GML for representing spatial coverage. Since every aspect of data follows international standards, the proposed mechanism has a potential to be directly applied in the future geographic data environment.

INTRODUCTION

With the vastly growing volume of digital geographic data in a variety of domains and the widely accessibility of internet GIS technology, it is much easier to query and access geographic data around the world nowadays (Bishr, 1998). Sharing and correct use of acquired geographic data from others therefore becomes a critical challenge to any participants in GIS. One aspect of these challenges is the interoperability of geographic data (Zhang et al., 2000). When combining or integrating acquire geographic data, the most frequently used operation is map overlay (Dangermond, 1990) that allow users to superimpose one map layer to another in accordance with correct relative location relationship. GIS users then treat this interface as a "map" based on their metaphor (Kuhn, 1993) to traditional paper maps. While this type of overlaid map layers may indeed look like a map, we argued it can hardly be regarded as a "map" if appropriate cartographic knowledge has not been taken into consideration (Hong, 1998). The compilation of traditional maps follows rather strict design rules to ensure correct representation of the described spatial phenomena, without appropriate aids, naive users may make wrong decisions while he or she may never notice (Vitek et al., 1996). In our pervious research (Hong and Liao, 2001), we proposed a valid coverage analysis concept towards the layer-based map interface, the valid coverage concept means all data inside the spatial coverage are to be "completeness". For example, a Tainan-City map illustrate that there are five Mcdonalds in Tainan. This map will be regarded as a valid coverage map while the spatial coverage is the whole Tainan city and the recording is complete. After adding map layers into the interface, the proposed mechanism automatically extracts metadata items about spatial coverage (Medyckyj-Scott et al., 1996) and evaluate the valid coverage of map interface. The analyzed results are recorded in another thematic map layer and superimposed onto selected thematic may layers to provide a visual aid of the valid coverage. This can avoid some possible mistake and exaggerated interpretation from inexperienced users.

In our earlier works, the analysis of valid coverage was first based on metadata elements we design, we then replaced these metadata elements with bounding coordinates from current metadata standards (e.g., FGDC, ISO). Unfortunately these two alternatives are either restricted to limited use (not following standard) or only provide approximate coverage description (rectangle corner coordinates). In this continuous work, we attempt to further introduce XML-based metadata for valid coverage analysis and expand the description of spatial coverage to include irregular shape of representation. XML documents are text-based, making them more readable, easier to document (Goldfarb and Precod, 1998). Besides, since XML is an international standard (W3C, 2004), XML-based approach not only can extract and transform data by standard tools, such as XML parses (XML, 1998) and XSLT processors (XSL, 1998), but also allow users to easily design their own program to automatically parse the information they need. If all metadata files precisely follow metadata standards, the parsing and analysis can be directly completed and the whole procedure has a potential to be fully integrated with the current map interface. Bounding coordinates use the corner coordinates of a rectangle to describe spatial coverage. However, the valid coverage boundary of a geographic data file may be a point, a line, a polygon, or even multi-polygons dependent how data is collected, a simple rectangle certainly cannot suffice all needs. To be able to effectively model these types of dimensionality variation, Geography Markup Language (GML, 2003) proposed by OpenGIS Consortium is chosen. Since both metadata standards and GML are XML-based, they can be easily integrated together as long as its schema can be appropriately defined.

By choosing XML as recording media in this paper, the proposed approach can not only deal with the recording of metadata and take advantage of GML representation to precisely describe spatial coverage, it also enables the possibility for building future interoperable geographic data environment in internet. The remaining of this paper is organized as follows, the next section first reviews fundamental theory about XML-based metadata, the following section continues to discuss we extend current metadata standards to include GML-based spatial coverage. In the fourth section, some real tests for interpreting valid coverage based on available metadata are illustrated and discussed. Finally we summarize our major findings and discuss future works in the final section.

XML-BASED METADATA

Metadata was either recorded as text or database file in earlier days, but most of them have been switched to XML format in recent years (FGDC, 2001; Abiteboul et al., 2000; Hong and Liao, 2003). XML is a new standard adopted by the World Wide Web Consortium (W3C, 2004), which aims to complement HTML for data exchange on the web. The basic component in XML is called element, which is required to be delimited by a start- and an end-tag, nest properly within each other. Below is an example of contact information metadata in XML format:

<cntinfo>
<cntperp>*Liao, Hsiung-Peng*<cntper/>
<address>*#1,University Road, Tainan, Taiwan, Republic of China*</address>
<postal>*701*</postal>
<cntvoice>*886 (6) 2757575-851*</cntvoice>
<cntemail>*darkness@.sv.ncku.edu.tw*</cntemail>
<hours>*08001700*</hours>
</cntinfo>

An expression such as <address> is called a start-tag, while </address> is an end-tag. Start- and end-tags are also called markups. Unlike HTML, the merit of XML is markups can be defined and extended by domain users to meet their particular needs. Data providers can therefore record all required data in a predefined XML format, and other users or programmers can easily interpret or write programs to process XML documents. To display its content, Extensible Stylesheet Language (XSL, 1998) can be used to convert XML into HTML file. Figure 1 shows the transformation result of a metadata file displayed with a browser.

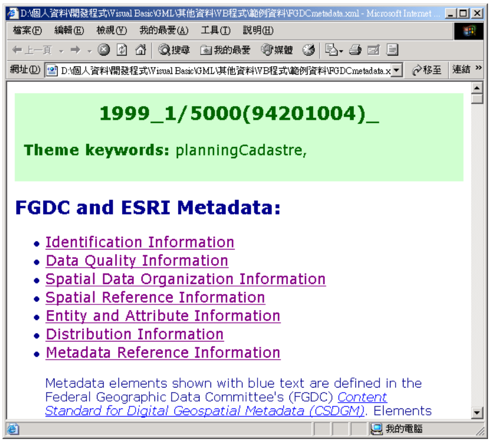


Figure 1: International metadata displayed with a browser.

For any metadata standard, the spatial coverage of a geographic data file is always a necessary metadata element because it enable the searching and correct interpretation of acquired data. The most frequently used way for representation is bounding coordinates based on the well known Minimum Boundary Rectangle (MBR) concept (FGDC, 2001; ANZLIC, 2001). With the aid of DOM (DOM, 1999), we can compile the entire document directly and construct a tree representation for it. DOM is an API for XML documents. It offers an object-oriented view of the XML document. That is, each document element is defined as an object, so programs can access the data via path-query automatically. Table 1 shows the name of the four tags used to describe bounding coordinates. All data was hypothesis to record in latitude and longitude in our test. For example, if we want to get the coordinate of west boundary, we have to first search tags <westbc> and </westbc>, then retrieve the value between these two tags (“23.40°”) and assign it as the west boundary. As long as all of the acquired metadata files follows the XML schema of the metadata standards, we can easily retrieve their corresponding spatial coverage data. If an international standards like ISO TC211 is chosen, it is even possible to analyze metadata around the world.

Table 1. Spatial Boundary.

Tags Name	Example Data	Memo
Westbc	23.40°	The coordinate of west boundary
Eastbc	24.21°	The coordinate of east boundary
Northbc	120.19°	The coordinate of north boundary
Southbc	120.45°	The coordinate of south boundary

EXTEND METADATA

Although MBR is easy to use and easy to understand, it nevertheless can only record approximate boundary of geographic data. When the surveyed area is a point, a line, an irregular shape of polygon, or even multi-polygons, it is obvious a rectangle or a polygon is no longer an optimal way for spatial coverage description and it is necessary to seek other alternatives. For describing object with geographic nature, GML emerges as a possible solution. First of all, GML is also an international XML standard based on the abstract model of geography. Secondly, GML is an XML grammar written in XML Schema for modeling, transport, and storage of geographic information (GML, 2003), it provides a variety of kinds of objects for describing geographic phenomena, such as point, line, polygon, topology, coordinate reference systems, etc. Since Geography Markup Language (GML) proposed by OpenGIS Consortium can provide description of geometric objects in a variety of dimensionality, it is chosen for modeling the varied nature of spatial coverage in this paper. Following is a multi-polygon GML example:

```
<gml:featureMember>
  <gml:multipolygonProperty>
    <gml:polygonMember><gml:Polygon srsName="">
        <gml:outerBoundaryIs>
            <gml:LinearRing>
             <gml:coordinates>23.40,120.19 23.40,120.45 24.21,120.19 24.23,119.88
             </gml:coordinates>
          </gml:LinearRing>                                                        A
        </gml:outerBoundaryIs>
      </gml:Polygon>
    </gml:polygonMember>
     <gml:polygonMember>
      <gml:Polygon srsName="">
             ··· (Abbreviate) ···
      </gml:Polygon>
    </gml:polygonMember>              B
  </gml:multipolygonProperty>
</gml:featureMember>
```

All recordings between tags *<gml:coordinates>* represents the precise coordinate of a polygon (Part A). This polygon comprise of five points, every coordinate is recorded as a pair of text (e.g, “23.40,120.19”), representing the E (Horizontal) coordinate and the N (Vertical) coordinate respectively. The spatial coverage and shape of this polygon can be then be easily retrieved and reconstructed. Note the second tag has already indicated this feature is a multipolygon, so we can repeatedly add <polygonMember> to increase the number of polygons.

XML Schema defined the grammar of conformant XML data instances. To extend GML-based boundary description into current metadata standard, we have to modify the predefined XML Schema for metadata. Figure2 illustrates the major steps the XML extensible ability, we design our tag “<Valid Coverage>” to record the valid coverage. Using this tag, we can append GML-based valid coverage into standard XML-based metadata. Furthermore, to ensure all created XML documents follow fixed format, the W3C XML schema language can be further introduced to validate XML document. Next section will explain how to use this information help us in the map overlay process.

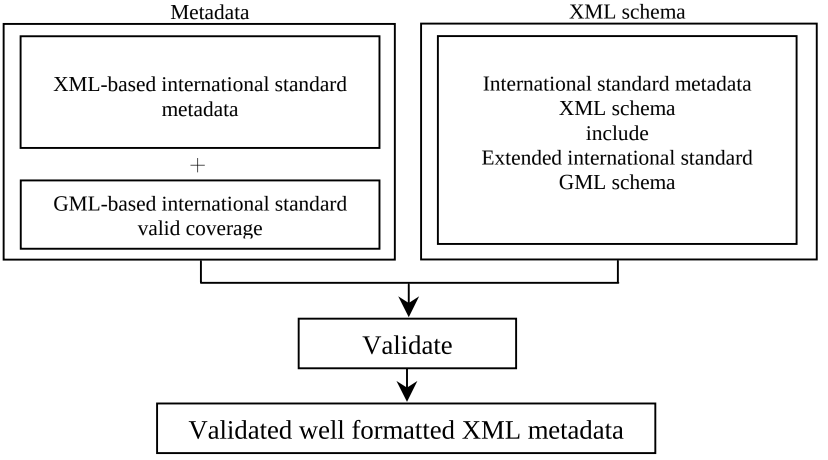


Figure 2: XML-based metadata.

MECHANISM

Hong et al. (2002) suggest that a correct interpretation of valid coverage towards map interface must be based on the analysis of the valid coverage of all corresponding data. To do so, the valid coverage of each data must be explicitly recorded and provided to users along with the requested metadata. Our mechanism was programmed with Map Objects (MO, 2003) and Visual Basic. Window 2 (refer to Figure 3) is the index map where the relative location of the work area (window 1). Window 3 enables users to control which thematic will be visible in the work area, besides, provides users to change the legends of each thematic data shown in the work area by their own.

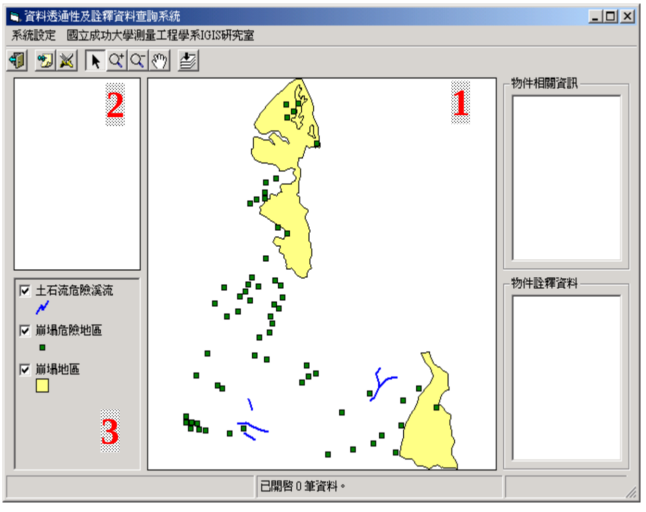
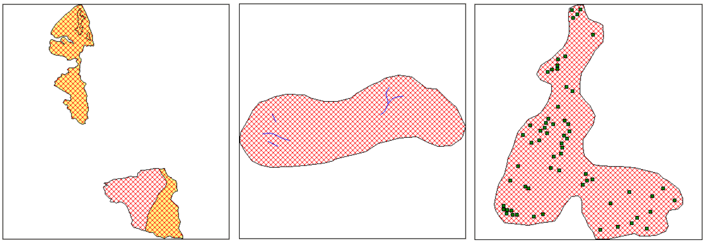


Figure 3: Valid coverage interpretation mechanism.

When users add a map to this mechanism, the system will automatically deal the following four steps:

1. Read the corresponded metadata.
2. Extract the recorded GML-based inventory boundary (discussed in previous section) of each map from XML-based metadata.
3. Use the boundary to create multi-polygons shape, and provide users a visual aid of the boundary on the window directly (the diagonal cross fill area in figure 4).
4. Overlay every boundary shape of all visible thematic maps to determine the valid coverage. Figure 5 is the interpretation result of all visible thematic maps.



a. polygons b. lines c. points

Figure 4: Visual aid of each boundary.

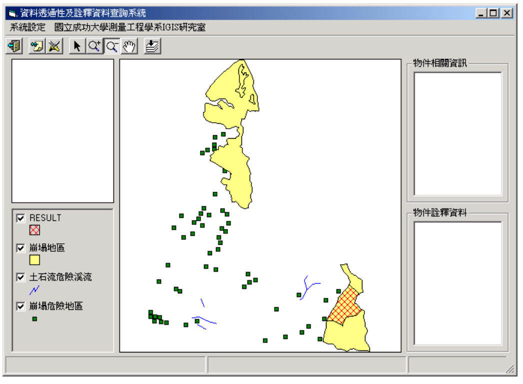


Figure 5: Interpretation result of all visible thematic maps.

With these steps a general form to extract the boundary of valid coverage is the following ("{}" denoting required elements, "*[]*" are examples, and "UPPER CASE" denoting language elements, "lowercase" denoting variables):

OPEN thematic map *[land.shp]*
READ METADATA *[land_metadata.xml]*
FIND TAGS {<Valid Coverage>}
EXTRACT BOUNDARY *[multi-polygons]*
CREATE SHAPE {multi-polygons}
OVERLAY SHAPES
ILLUSTRATE THE INTERPRETATION RESULT

Boundaries of all coverage are subject to hypothesis tests in order to define and extract valid coverage from actual data. In Figure 4, we can find out the valid coverage (the diagonal cross fill area) differs from each thematic map. For example, Figure 4.a displays the polygon data, while the boundary of valid coverage is multi-polygons. In the past, without property metadata, users may not notice this fact.

In spite of we add three thematic maps in Figure 5, the window illustrate that only a small area can be regarded as valid coverage. Thus only this area can be treated as map, and only the map inside this area can provide us a correctly decision. Other area exterior the valid coverage is only a reference map. If users want to use those area outside the valid coverage, fist of all, they have to think about what is lack of the data. Otherwise, they may make a wrong decision but they never notice.

CONCLUSION

It is no doubt that future GIS uses must largely rely on geographic data provided by others. The lack of professional knowledge (about acquired data) and a naive attitude about map overlay substantially increase the risks of wrong data use and decision making. Continuing our research thread of valid coverage analysis, we change metadata format from dbf to XML and introduce GML to augment the capability of describe spatial coverage in different spatial dimensionality. The fact that all data are recorded in XML simplifies the recording and interpretation process and further increase the possibility for future internet geographic distribution. For example, we may be able to acquire geographic data, along with their metadata, from a variety of data sources online and display them in our browsers. The valid coverage information can be automatically extracted and analyzed from metadata and provided to users for reference. Valid coverage is just one aspect of data interoperability, though current research based on international standard already improves the data sharing via internet. Users may access data from different site, view the overlay map on their browser, while they may never know where the site is. The browser can automatically provide the user properly warring about the retrieved map, not only valid coverage, but also other information, like accuracy, temporal and so on. Thus, in the future research, we will take our focus on taking more influenced element in the metadata into consideration. To ensure all sharing data can be correctly used for users.

REFERENCES

Abiteboul, S., Buneman, P. and Suciu, D., 2000: *Data on the Web: From Relations to Semistructured Data and XML*. Morgan Kaufmann Publishers, San Francisco, California. ISBN: 1-55860-622-X. pp. 27-50.

ANZLIC, 2001: AN*ZLIC Metadata Guideline*. Australian New Zealand Land Information Council. http://www.anzlic.org.au/asdi/metaelem.html.

Bishr, Y., 1998: Overcoming the semantic and other barriers to GIS interoperability. *International Journal Geographical Information Science* 12(4), 299-314.

Dangermond, J., 1990: A classification of software components commonly used in geographic information systems. *Introductory Readings in Geographic Information Systems*, Taylor Francis, pp. 30-51.

DOM, 1999: *The World Wide Web Consortium (W3C)'s DOM (Document Object Model) Web Page*. http://www.w3.org/DOM/.

FGDC, 2001: *Contents Standards for Digital Geospatial Metadata*. http://geology.usgs.gov/tools/metadata/standard/metadata.html.

GML, 2003: *OpenGIS ® Geographic Markup Language (GML) Implementation Specification, Version 3.0*. http://www.opengis.org/docs/02-023r4.pdf

Goldfarb, C.F. and Precod, P., 1998: *The XML Hand Book.* Prentice Hall, Englewood Cliffs, NJ.

Hong, J.H., 1998: More than map overlay – integration of geographic data from the perspective of data. *Proceedings of the 4th Symposium on Geographic Information System*, Taipei (CD-ROM).

Hong, J.H. and Liao, H.P., 2001: Map metaphor consideration toward overlaying geographic data: a coverage metadata perspective. *Asia GIS 2001*, Tokyo JAPAN. pp. 4-5.

Hong, J.H., Liao, H.P. and Lee, B.L., 2002: The valid spatial coverage analysis mechanism for overlaying geographic data from different sources. *Journal of Surveying Engineering* 44(4), 97-114.

Hong, J.H. and Liao, H.P., 2003: A geographic data interoperability perspective towards the valid coverage of map interface. *1th Digital Earth International Conference*, Taipei, (CD-ROM).

Kuhn, W., 1993: Metaphors create theories for users. *Proceedings of European Conference, COSIT'93*, Lecture Notes in Computer Science, No. 716, pp. 366-376.

Medyckyj-Scott, D., Cuthbertson, M. and Newman, I., 1996: Discovering environmental data: metadatabases, network information resource tools and the GENIE system, *International Journal Geographical Information Systems* 10(1), 65-84.

MO, 2003: *Map Objects*. ESRI web page. http://www.esri.com/software/mapobjects/mo-index.html.

Vitek, J.D., Giardino, J.R. and Fitzgerald, J.W., 1996: Mapping geomorphology: a journey from paper maps, through computer mapping to GIS and virtual reality, *Geomorphology* 16, 231-249.

XML, 1998: The World Wide Web Consortium (W3C)'s XML web page, http://www.w3c.org/XML/.

XSL, 1998: *Extensible Stylesheet Language (XSL)*. http://www.w3c.org/TR/WD-xsl.

W3C, 2004: The World Wide Web Consortium (W3C). *Extensible Markup Language (XML) 1.0 (Third Edition)*. http://www.w3.org/TR/2004/REC-xml-20040204/.

Zhang, J., Javed, M., Shaheen, A. and Gruenwald, L., 2000: Prototype for wrapping and visualizing geo-referenced data in a distributed environment using XML technology. *Eighth International Symposium of ACMGIS*. McLean, Virginia, November 2000.