GEOSPATIAL SERVICES FOR DECISION SUPPORT ON PUBLIC HEALTH

S. Gao [a, *], D. Mioc [a], X. Yi [b], F. Anton [c], E. Oldfield [b]

[a] Department of Geodesy and Geomatics Engineering, University of New Brunswick, Canada - (sheng.gao, dmioc)@unb.ca
[b] New Brunswick Lung Association, Canada – xiaolun.yi@gmail.com, eddie.oldfield@nb.lung.ca
[c] Department of Informatics and Mathematical Modeling, Technical University of Denmark, Denmark – fa@imm.dtu.dk

KEY WORDS: GIS, Internet/Web, Mapping, Interoperability, Decision Support, Public Health

ABSTRACT:

Sharing of health information is critical for preventing disease, responding to emergencies, educating the public and policy makers. Protecting privacy and confidentiality of health information remain important cornerstones of the public health system, however many health professionals and authorities do not have the ability to visualize health information to make time-sensitive decisions, since they do not have the time, money, or skills to statistically analyze vast amounts of distributed data and render aggregated results into a geographic interface for quick interpretation. The technology to do so, web based geographic information systems and related standards, has matured yet confidence in such technology to visualize or share health information is only beginning to emerge. Currently, four major problems still exist in health geographic applications. They are related to health mapping methods, mapping variables, reusability of health applications, and interoperability issues. To handle these problems, we designed a Health Representation XML schema and SOA based architecture to support health data sharing and representation. The schema makes it possible to exchange the statistical results of health data as well as representation through XML and GML. The OGC services such as WMS, WFS, and WPS enable the statistical exploration and representation of health information. A Web-portal is developed to support the integration of different services for visualization of health maps, hypothesis generation, and decision making. This architecture provides quick access to spatial and health data for understanding the trends in diseases, and promotes the growth and enrichment of the SDI in the public health sector.

1. INTRODUCTION

Currently, many determinants such as booming population, environmental pollution, rapid urbanization, convenient transportation, and global warming are improving the conditions for disease outbreaks. To prevent and mitigate the risk of diseases, it is important to build a robust health system to support evidence based decision making. The sharing of essential health information is one of the most feasible routes to achieve global public security (WHO, 2007). There is strong need for access to maps on disease prevalence, mortalities, health determinants, transmission patterns, and components of healthcare responses.

As health phenomena have revealed strong spatial aspects, maps can show geographic distributions and spatial patterns of diseases. Analyzing and mapping the spatial aspects of disease can improve our understanding of disease etiology, facilitate work with therapists to educate the public, and augment decision-making on programs that aim to prevent illnesses. The health applications using spatial components of diseases can be traced back to 1854 when Dr. John Snow combined geospatial information to analyze the cholera deaths and found clusters around water pumps (McLeod, 2000). There are three important functions of Geographical Information System (GIS) in health research and policy analysis: spatial database management, visualization and mapping, and spatial analysis (Cromley and McLafferty, 2002). Database management include linking, integrating and editing many kinds of data that are located on the Earth's surface, such as health, social, environment data. Visualization and mapping can explore the spatial patterns and correlations of diseases and many factors such as census and environment. Spatial analysis utilizes the spatial relationship to generate new health patterns. For instance, Kriging is a popular method used for interoperating health data. When a disease appears, GIS can represent disease information rapidly and analyze the spread of disease dynamically.

Meanwhile, the rapid development of the Internet promotes the popularity of Web-based GIS, which itself shows great potential for the sharing of health information through distributed networks. Distributing and sharing health maps via the Web helps decision makers across health jurisdictions and authorities collaborate in preventing, controlling, and responding to a specific disease outbreak. Users without tools or without necessary skills can make use of GIS functions through Web-based GIS (Wright et al., 2003). The Documented applications are already making health information accessible through the Web (Benneyan et al., 2000; Edberg, 2005). Custom online interactive health maps can be implemented using Google Maps API, Google Earth KML or MSN Virtual Earth Map Control (Boulos, 2005). Several Web-based GIS applications can generate disease maps dynamically from the server side (Blanton et al. 2006;Inoue et al. 2003), while other applications employ Java Applets and Scalable Vector Graphics (SVG) approaches to visualize health information (Kamadjeu and Tolentino, 2006; Qian et al., 2004). Boulos and Honda (2006)

* Corresponding author.

proposed to publish health maps through Web Map Service with Open Source Web-based GIS software.

2. EXISTING PROBLEMS ON HEALTH GEOGRAPHIC APPLICATIONS

In spite of the continuous development of geographical health applications, the following four problems still need to be handled.

Firstly, the methods to generate maps from health related activities need to be considered. There are different kinds of health activities, such as hospital observation, laboratory tests and results, healthcare and medication services, and training and education for patients. Since these activities are social events and related to spatial location, a proper way to support mapping of these activities on maps is a foremost concern in health geographic applications. Many web-based health applications dynamically generate maps, but they lack data source description, and method declaration on how the maps are generated.

Secondly, many mapping dimensions for health data representation should be supported. In finding support of disease surveillance, the variables in disease type dimension, temporal dimension, gender dimension, and age dimension are valuable. For instance, disease patterns show striking differences at different representation scales, which is recognized in many health studies (Albert et al., 2000; Leitner and Curtis, 2006).

Thirdly, integrating and reusing current health applications are constrained to a large extent. Zeng et al. [15] pointed out that the isolation of existing stand-alone disease management systems leads to a data sharing problem. Most of the health information systems have a closed architecture - even the ones that use web-based technology are difficult to integrate. Typically, users can only access maps from such a health application, and it is difficult to integrate datasets from these applications.

Fourthly, different health application lacks interoperability between them. Interoperability makes it easy to communicate, execute programs, or transfer data among various systems in a unified manner. With closed and centralized legacy architecture, a web-based GIS system can not fully adapt to current distributed, heterogeneous network environments, and is unlikely to provide users with the needed data and services due to its lack of interoperability, modularity, and flexibility (He et al., 2005). In health decision making, it is important to access various kinds of data such as hospital locations and available medical resources through standard interfaces.

3. HRXML FOR HEALTH REPRESENTATION DATA EXCHANGE

In the mapping of health related activities, statistical methods are used to connect health related activities with maps. The following statistical methods are considered in this research: Crude Morbidity Rate (CMR), Normalized Morbidity Ratio (NMR), Age-Specific Morbidity Ratio (ASMR), Age-Adjusted Morbidity Ratio (AAMR), and Standardized Morbidity Ratio (SMR), Summation, Mean, Standard Deviation, Variance, Skewness and Kurtosis. These methods are concentrated on spatial, temporal, and demographic factors and their influence on health related activities, which can show the health information distribution with spatial, temporal, age, and gender differences. Other statistical methods can be introduced to support more influential factors.

XML is very popular in supporting the data exchange in heterogeneous systems. Health Level 7 (HL7) standards that are accredited by non-profit America National Standards Institute, allow representation and exchange of information from online patient records to pharmacy formularies in HL7 XML documents. The primary domain of HL7 standards is clinical and administrative data, and explicit spatial information and health data mapping are not considered. Instead of exchanging huge raw clinical data for health mapping, we focus on creating a Health Representation XML (HRXML) schema for the sharing of the statistical results of health data as well as their representation. In the design process, our intention is to make the schema as simple and extendable as possible. Three dimensions of representation are related with spatial data: semantic, geometric and graphical (Bédard and Bernier, 2002). Therefore, we include these three kinds of representations in the HRXML schema. Semantic representation describes the health related activities and the statistical methods used. Geometric dimension shows what type of geometry (point, line, and polygon) will be used to represent these health data. Graphic representation defines what styles or symbols are used to generate health maps.

As shown in Figure 1, the designed HRXML schema includes three parts: health, mapping data, and representation.

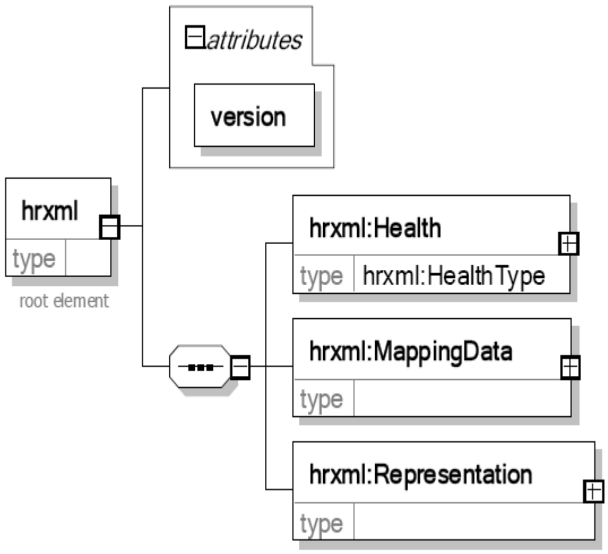

Figure 1. HRXML schema

The health part includes the basic information of the health related activities, with the name, title, description, and keywordList elements, and a type attribute. HealthType is an abstract complex type. It can be extended to support disease observation or other activities.

The mapping data part mainly records the data used for mapping. As shown in Figure 2, it includes the bounding box of the data, spatial data, relation between spatial data and mapping values, and mapping values.

--"BoundingBox" represents the spatial range of the mapping data.

--"SpatialData" could be GML from WFS services, GML records, or Xlink to GML databases. The health data are statistical values and are linked with the spatial data through the joining attribute.

-- "Relation" records the joining attribute and the matching ID value of both spatial data and mapping values.

--"Mapping values" includes the health data source description, statistical method used and mapping value lists. The statistical method part describes the name, title, description, health data source, and statistical parameters of the statistical method used. Statistical methods are used to generate classification maps and charts for health related activities. We predefined some parameters from the spatial, temporal, and demographic aspects for public health, such as AgeFrom, AgeTo, and StartTime, which can show health distributions with spatial, temporal, age and gender differences. Users can add additional parameters in the parameter group to support advanced statistical methods.

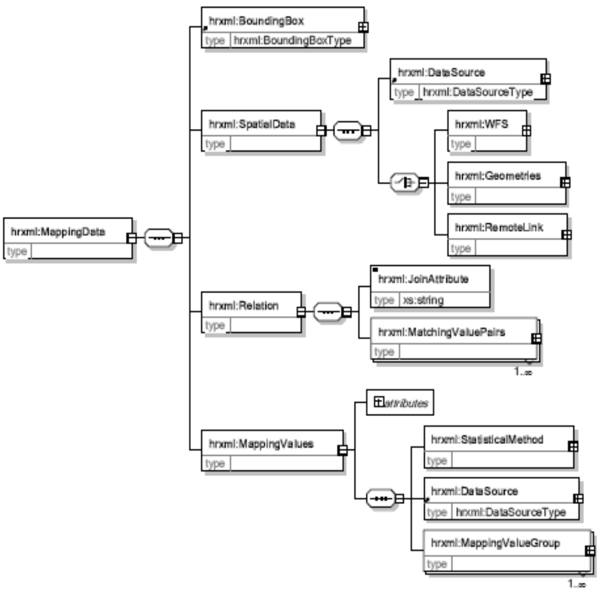

Figure 2. The mapping data part schema

The Representation part defines the kind of style used to represent health maps. It consists of the default representation bounding box and style description. Depending on the kind of representation, the StyleType is further extended to ChartStyleType, PointStyleType, LineStyleType, and PolygonStyleType. For instance, the PolygonStyleType includes the border and fill elements. The type of filling in a polygon can be gradient fill or range based fill. For the range based fill, the fill method can use colour, pattern, and texture. The border element contains the colour, line style and line weight of the border.

In the HRXML, the spatial related statistical health data are well described and able to be exchanged. With representation styles, the health data can be shown as thematic maps, which make information communication more understandable.

4. ARCHITECTURE DESIGN FOR HEALTH APPLICATIONS

Service oriented architecture provides a flexible way to share data as well as processing functions over the Internet to reduce costs of building complex systems. The service oriented architecture supports loose coupling between components and makes services reusable. As a result, Spatial Data Infrastructure (SDI) has been evolving from data driven architecture to service oriented architecture. To address geospatial data sharing and interoperability, Open Geospatial consortium (OGC) has been developing several specifications, such as Web Map Service (WMS), Web Feature Service (WFS), and Web Processing Service (WPS). WMS publishes its ability to produce maps rather than its ability to access specific data holdings, and generates spatially referenced maps dynamically (OGC, 2001). WFS defines the interfaces for the access and manipulation of geographical features and elements through Geography Markup Language (GML) (OGC, 2005b). WPS provides standardized interfaces to facilitate publishing, discovery and binding geospatial services that enable spatial processing functions across a network (OGC, 2005a).

Accessing health information through standard interfaces is important to achieve interoperability. In this way, such access could improve the ability to intervene in health issues, and inform the public of the availability of resources and community health programs. To accomplish a web-based application for statistical exploration of health information, we take advantage of the standard OGC services including WMS, WFS, and WPS. The proposed architecture (see Figure 3) includes three tiers: a data tier, a service tier and a web portal tier.

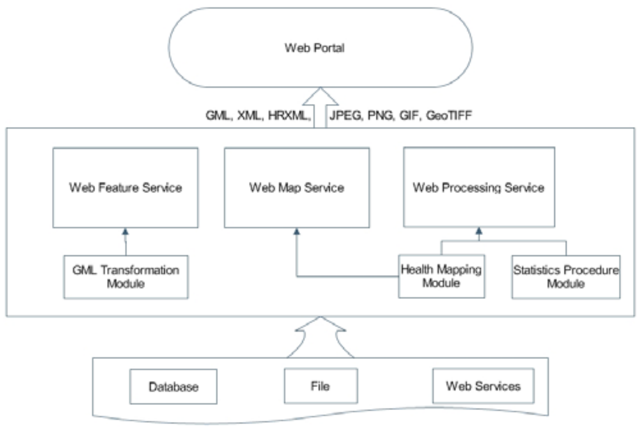

Figure 3. System architecture

The data tier stores all the health related data and spatial data for health studies. These data could be available from databases or web services.

The service tier implements WMS, WFS, and WPS for health studies.

--WMS provides standard interfaces to generate maps and charts for visualization of health information. It utilizes the health mapping module to generate maps. The health module can generate various style maps such as unique colour classification, graduate colour classification, pie charts and bar charts, to show events or facilities distribution. The input data could be obtained from HRXML, GML, WFS, WPS, Oracle DBMS, Key-Value documents, or shape files.

--WFS uses the GML transformation module to share spatial data through GML. It can be linked with the mapping values (part of HRXML) to create thematic health maps.

--WPS is used to analyze spatial-temporal health data. The health data analysis supports data rolling up from a low spatial level to a high spatial level. WPS uses the health mapping module and statistical procedure module to calculate the result, and the output could be HRXML or health maps (JPEG, PNG, GIF, and GeoTIFF). The input data of WPS could be obtained from WFS, GML Oracle DBMS, or shape files.

The web portal tier is a client for the visualization of disease data and maps. It can bring together different facets of health information into one location to improve health promotion and health care research, education, and policy-making.

5. IMPLEMENTATION

Based on the above framework, we implement a prototype for the sharing of health information. The health data we used include four kinds of respiratory disease data collected by the New Brunswick Lung Association. The disease data are geo-coded to spatial position through the postcode. The spatial data we used include the six levels of spatial boundary data that cover the entire territory of New Brunswick. The six levels are "Province", "Health Region", "Census Division", "Census Subdivision", "Forward Sortation Area", and "Dissemination Area" geo-layers. All the health data and geometrical boundary data are stored in Oracle 11g. With the spatial operation provided by Oracle, the disease data can be easily rolled up from a low spatial level to a high level, and low counts i.e. less than five observations, or false counts are not represented to further ensure privacy and accuracy. We use CARIS WMS and WFS to publish the data. The implementation of health mapping module utilizes the Geotools library. The statistical procedure model is implemented with the Oracle PL/SQL code. The client side is developed using Javascript and HTML.

Figure 4 shows an example of an HRXML document generated by WPS.

```
<?xml version="1.0" encoding="UTF-8" ?>
- <hrxml:hrxml version="" xmlns:hrxml="http://nblung.ca" xmlns:xsi="http://www.w3.org/2001/XMLSchema-
  instance" xsi:schemaLocation="http://nblung.ca HRXML.xsd">
 + <hrxml:Health xsi:type="hrxml:DiseaseObservationType">
 - <hrxml:MappingData>
  + <hrxml:BoundingBox srsName="EPSG:4326">
  - <hrxml:SpatialData>
   + <hrxml:DataSource>
   - <hrxml:WFS>
      <hrxml:URL>http://131.202.98.45:8090/wfsservlet</hrxml:URL>
      <hrxml:LayerName>healthregion</hrxml:LayerName>
     </hrxml:WFS>
    </hrxml:SpatialData>
  + <hrxml:Relation>
  - <hrxml:MappingValues>
   - <hrxml:StatisticalMethod>
      <hrxml:Name>CMR</hrxml:Name>
      <hrxml:Title>CMR</hrxml:Title>
      <hrxml:Description>Crude Morbidity Rate</hrxml:Description>
    - <hrxml:ParameterGroup>
       <hrxml:StartTime>990101</hrxml:StartTime>
       <hrxml:EndTime>991231</hrxml:EndTime>
       <hrxml:AgeFrom>0</hrxml:AgeFrom>
       <hrxml:AgeTo>24</hrxml:AgeTo>
       <hrxml:Gender>Male</hrxml:Gender>
       <hrxml:Geolayer>HR</hrxml:Geolayer>
      </hrxml:ParameterGroup>
     </hrxml:StatisticalMethod>
   + <hrxml:DataSource>
   + <hrxml:MappingValueGroup>
    </hrxml:MappingValues>
   </hrxml:MappingData>
 + <hrxml:Representation>
</hrxml:hrxml>
```

Figure 4. An HRXML example

A portal has been implemented in the integration of these OGC services for decision making. The public HTML portal provides access to aggregate health maps for the general public. The HTML portal is developed to allow easy and quick access to WMS/WPS services for visualization purposes. As shown in Figure 5, a CMR distribution map from WPS and some facility distribution maps from WMS are integrated. It provides users (researchers, health officials, practitioners, policy makers, and epidemiologists) with access to GIS functionality for visualizing health data, and evidence-based decision-making on disease outbreaks.

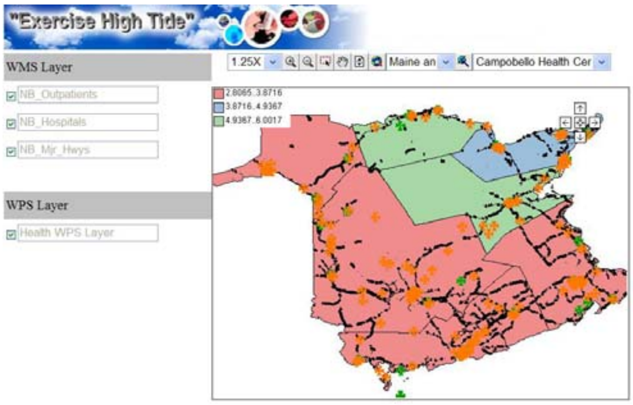

Figure 5. The web portal interface

6. CONCLUSION

In this research, we have applied the service oriented architecture in health geographic applications, with many standard WMS, WFS, and WPS services. This architecture will facilitate the data sharing as well as the reusability and interoperability of health services. We develop the HRXML schema to exchange statistical health data as well their representation based on XML and GML specifications. A user-friendly web application has been built for the exploratory and descriptive analysis of health information, hypothesis-generation, and decision-making. It provides quicker access to spatial and health data in understanding the trends in disease, and promotes the growth and enrichment of the SDI in the public health sector. Sharing of health information can improve the ability to intervene in health issues, and inform the public of the availability of resources, reduce the number of people affected by illness, and therefore reduce costs to the health-care system. Our future work will be on the implementation of the OGC Web Catalogue Service and accomplish the semantic query and access of health geographic services.

ACKNOWLEDGMENTS

This research work has received the financial support from GeoConnections secretariat of natural resources Canada for a project titled “Development of Web Application and Services within the CGDI framework for Community Health Programs of the New Brunswick Lung Association”.

REFERENCES

Albert, D.P., Gesler, W.M., Levergood, B., 2000. *Spatial analysis, GIS and remote sensing : applications in the health sciences.* Ann Arbor Press, Chelsea, MI.

Bedard, Y., Bernier, E., 2002. Supporting multiple representations with spatial view management and the concept of "VUEL". *Proceedings of Joint Workshop on Multi-Scale Representations of Spatial Data, ISPRS WG IV/3, ICA Commission on Map Generalisation,* Ottawa, Canada, July 7-8.

Benneyan, J.C., Satz, D., Flowers, S.H., 2000. Development of a web-based multifacility healthcare surveillance information system. *Journal of healthcare information management,* 14(3), pp. 19-26.

Blanton, J.D., Manangan, A., Manangan, J., Hanlon, C.A., Slate, D., Rupprecht, C.E., 2006. Development of a GIS-based, real-time Internet mapping tool for rabies surveillance. *International Journal of Health Geographics,* 5.

Boulos, M.N., 2005. Web GIS in practice III: Creating a simple interactive map of England's Strategic Health Authorities using Google Maps API, Google Earth KML, and MSN Virtual Earth Map Control. *International Journal of Health Geographics,* 4.

Boulos, M.N., Honda, K., 2006. Web GIS in practice IV: Publishing your health maps and connecting to remote WMS sources using the Open Source UMN MapServer and DM Solutions MapLab. *International Journal of Health Geographics,* 5.

Cromley, E.K., McLafferty, S., 2002. *GIS and public health.* New York: Guilford Press.

Edberg, S.C., 2005. Global Infectious Diseases and Epidemiology Network (GIDEON): a world wide Web-based program for diagnosis and informatics in infectious diseases. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America,* 40(1), pp. 123-126.

He, L.L., Yang, J.G., Deng, C., Qi, H.N., 2005. Multi-agent framework for service-oriented geospatial computing. *2005 International Conference on Machine Learning and Cybernetics, ICMLC 2005,* pp. 143-148.

Inoue, M., Hasegawa, S., Suyama, A., Meshitsuka, S., 2003. Automated graphic image generation system for effective representation of infectious disease surveillance data. *Computer Methods and Programs in Biomedicine,* 72(3), pp. 251-256.

Kamadjeu, R., Tolentino, H., 2006. Web-based public health geographic information systems for resources-constrained environment using scalable vector graphics technology: A proof of concept applied to the expanded program on immunization data. *International Journal of Health Geographics,* 5.

Leitner, M., Curtis, A., 2006. A first step towards a framework for presenting the location of confidential point data on maps-results of an empirical perceptual study. *International Journal of Geographical Information Science,* 20(7), pp. 813-822.

McLeod, K.S., 2000. Our sense of Snow: The myth of John Snow in medical geography. *Social Science and Medicine,* 50(7-8), pp. 923-935.

OGC, 2001. Web Map Service Implementation Specification. Available at: http://portal.opengeospatial.org/files/?artifact_id=1058.

OGC, 2005a. OpenGIS Web Processing Service. Available at: http://portal.opengeospatial.org/files/?artifact_id=13149.

OGC, 2005b. Web Feature Service Implementation Specification. Available at: http://portal.opengeospatial.org/files/?artifact_id=8339.

Qian, Z., Zhang, L., Yang, J., Yang, C., 2004. Global SARS information WebGIS design and development. *International Geoscience and Remote Sensing Symposium (IGARSS),* pp. 2861-2863.

WHO, 2007. The world health report 2007: a safer future: global public health security in the 21st century.

Wright, D.J., O'Dea, E., Cushing, J.B., Cuny, J.E., Toomey, D. R., 2003. Why web GIS may not be enough: A case study with the virtual research vessel. *Marine Geodesy* 26(1-2), pp. 73-86.

Zeng, D., Chen, H., Tseng, C., Larson, C.A., Eidson, M., Gotham, I., Lynch, C., Ascher, M., 2004. Towards a national infectious disease information infrastructure: a case study in West Nile virus and botulism. *Proceedings of Proceedings of the 2004 annual national conference on Digital government research,* Seattle, Washington.