

# Workflow issues for Health mapping “mashups” of OGC Web Services

Didier G. Leibovici \*, Suchith Anand \*, Jerry Swan \*,  
James Goulding \*, Gobe Hobona \*, Lucy Bastin #, Sergiusz Pawlowicz \*,  
Mike Jackson \* and Richard James §

\* *Centre for Geospatial Science, University of Nottingham, UK* # *Aston University, Birmingham, UK*  
§ *Centre for Health Care Associated Infections, University of Nottingham, UK*

## ABSTRACT

This paper explores potential uses of OGC web services in the context of health mapping and epidemiological studies. The situations or scenarios cover (i) different user perspectives, e.g. public and community, research, health professionals and (ii) different interaction levels, e.g. simple data “mashups” (overlay), use of web processing services (WPS) and participating GIS. Some particular aspects of the domain such as privacy for the individual and computational flexibility/performance raise issues in chaining web services. We propose some solutions at both the architecture and geocomputational level, as well as modifications to the standards involved.

## INTRODUCTION

Historical perspectives about health and epidemic studies using spatial information can be very informative about the benefits these disciplines can get from Geographical Information System – Science- (GIS). Using only pen and paper, combining cholera death and water pump locations, Dr John Snow demonstrated, 150 years ago, that spatial studies could be of great value in understanding and predicting infectious disease outbreaks. Yesterday’s desktop GIS provided more computational power, making the most of the visualisation display, spatial properties and statistical spatial analysis methodologies.

The internet and web services can now transfer and share all these capabilities with everyone. This evolution concerned three major methodological areas: (i) data acquisition, storage and exchange, (ii) data analysis, conflation and geocomputation and (iii) user-driven/contextual functionalities, *i.e.* for researchers, practitioners, decision makers and the general public. The experience of the user and the demanding flexibility he/she expects from these three methodological components are the drivers of geospatial science (Chang et al. 2009, Gao *et al.* 2008).

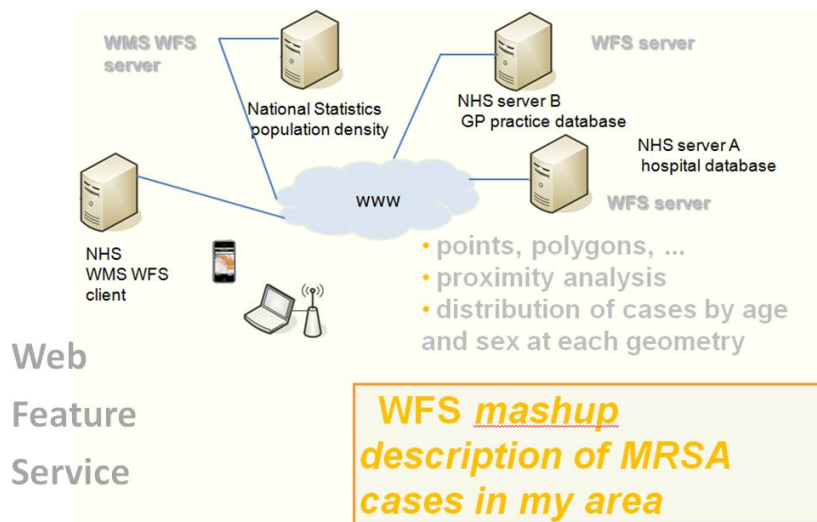
Web technologies and infrastructures, particularly the interoperability of web services, is just beginning to address the challenges of new and future health GIS services for people’s well being, epidemic monitoring, health and the environment, derived from spatial “mashups” and pertinent spatial statistics. Recent developments in web 2.0, positioning technology, location aware pervasive computing, as well as ubiquitous positioning, allowing the creation of contextual foot-prints or “scent-trails” (path and activities) provide additional means towards achieving pertinent spatial modelling. These new means can express their full potential only when embedded within interoperable web services infrastructures.

In the *International Journal of Health Geographics*, Boulos et al., 2008, give, within a long series of articles entitled “Web GIS in practice”, an interesting list of potential use of some modern technologies, particularly addressing non IT specialists, including health practitioners. These authors also use the metaphor “mashups” to explain the principle of overlaying of maps or weaving of web services in order to serve, or conflate different information sources. These geospatial mashups can be conceptually expressed and stored as workflows: the chaining of data access (WSF, WCS) and processing transformations/algorithms/models (WPS), represented as a graph for example using

BPMN (see further). Instead of focusing on particular tools, this paper investigates the role and use of the interoperability of web services *via* the standards from the Open Geospatial Consortium (OGC) for the domain of public health and epidemiology. Some typical scenarios are described using OGC web services (OGC, 2010), then particular important issues for the domain are explored with solutions that may lead to modify or extend the specification of the standards involved.

## HEALTH MAPPING SERVICES

In order to place emphasis on the interoperability for web services for geospatial use cases, the standards from OGC and ISO organisations prevail. For any disease, a simple data conflation can be conducted by using Web Map Services (WMS, see OGC, 2010), where an analysis could be, for example, a visual appreciation of the relative density of MRSA cases (Methicillin Resistant Staphilococcus Aureus, see Lowy, 2010) for a certain area (selection of a bounding box): overlay for a given area of a density map of cases and a population density map. A more interesting query (as suggested by figure 1) provide, for example, a distribution of cases by age and gender for a selected hospital or area delineated. This can be performed using WFS servers (Web Feature Services, see OGC, 2010). Within the domain of spatial epidemiology of MRSA (Grundmann et al. 2010), such functionality can be seen on the website [www.spatialepidemiology.net/SRL-Maps/maps/](http://www.spatialepidemiology.net/SRL-Maps/maps/), which uses Google maps API for the web client (using a WFS or not to access the GML files).

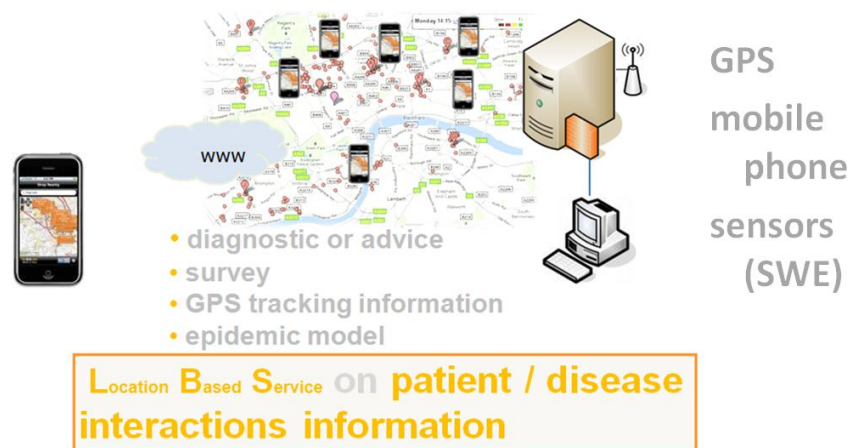


**Figure 1: Hypothetical architecture using Web Feature Service to provide MRSA cases information per General Practice and/or per hospital located in the requested map area.**

Notice that we mentioned the query is done for a selected area, but this depends in fact on the ability of the client to gather the attribute values of each selected geometry, and evaluate/display the aggregated distribution. This can be either implemented as such within the client or the client calls a WPS (Web Processing Service, see OGC 2010).

So an “ideal” health mapping client must be able to look for existing data services (WMS, WFS, WCS), then query the one selected, either to build up a data mashup and then run a selected WPS to provide a result from an analysis (a statistic, a statistical map), or, build up a more complex workflow involving more than one WPS also harvested. One can already interpret the selection of the mashup and the use of a WPS, as a workflow.

Apart from accessing and analysing data, web services is also used to perform surveys and collect data, which is also an important part of health and epidemic studies (if not the most time-consuming and expensive). Sensor Web Enablement (SWE, see OGC 2010) and location based services (LBS) provided, for example, by a mobile phone enable relevant ways of conducting a population survey and acquiring data. As seen in figure 2, an LBS using a mobile phone can deliver a public service in refining a standard diagnostic questionnaire with local information regarding the risk of a specific disease, and personal information coming may be from wearable devices (e.g., temperature) or from medical history either stored on the phone or remotely accessed (may be encrypted).



*Figure 2: Web2.0 and Participating GIS principles provided by mobile phone.*

A whole survey can also use mobile devices and GPS tracking of consenting patients. The latter acts as a LBS giving feedbacks to patients to get better assessment of their risks due to the environment (including other known cases), and as a basis for simulation studies, besides the current study of risk factors.

## **SOME ISSUES AND SOLUTIONS**

Confidentiality and privacy are the first constraints that are often discussed in epidemiological studies. Obviously if the data from the survey are accessible via the internet, this becomes even more crucial. Data access must be controlled depending on the information requested and the user profile. GeoXACML the extension of XACML from OASIS (XACML, 2005) and current developments of GeoRM (OGC, 2010) enable definition and control mechanisms for a web service to be accessed.

In the previous paragraph we already mentioned the workflow functionality that is desirable not only in the disciplinary domains discussed in this paper, but also in environmental modelling in a broad sense. An important feature for monitoring situations is to be able, to store a temporary result issued from chaining web services, to define and store its workflow along with its lineage. In public health and epidemiological studies, the semantic of the queries stored in the metadata of the workflow is of particular importance when for example, the progress of the knowledge about a disease leads to refining of selection criteria. Some standards, such as BPMN (Business Process Model Notation, BPMN 2009) from the OMG and XPD (XML Process Definition Language, XPD 2008) from the

WfMC or the more well known BPEL (Business Process Execution Language, BPEL 2007) are discussed within an OGC Domain Working Group to approach the issues for geospatial workflow.

At the intersection of the two previous issues is the concern about obtaining the most pertinent (or appropriate) results based on the available data at the finest (or appropriate) scale. To solve the antagonism between privacy and accessing fine scale data, we propose extending the web services standard in order to allow full access to data at computational level (from the WPS) but for the results to be output at a coarser scale, (Figure 3.). The access right of the WMS, WFS or WCS contains a link to an upscaling (generalisation) processing service. The scale is maintained within a chain of WPS until the output from the last WPS. The WPS standard has also to be modified to allow the final chaining to perform the upscaling or generalisation.

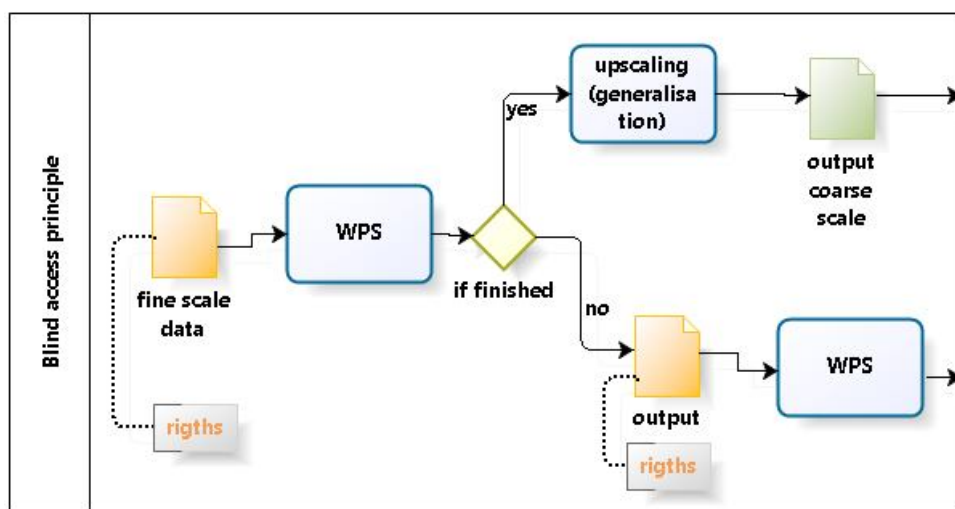


Figure 3: Principle of blind access to fine scale data for WPS preserving privacy.

At a geocomputational level, some heavy computing processing can diminish the performance; a lot of statistical methodologies can be expressed within an updating paradigm where a current statistical map or map associated with a current statistic is updated according to an incoming dataset. The generic signature of such “updating” WPS then contains: *datasets, current result, new datasets*. Some data assimilation methodologies proceed in the same way. Disease clustering is a very time consuming geoprocessing statistical method (Kulldorff 1995, Lawson et al. 2006). Following this “updating” approach we are modifying some spatial statistical algorithms exploring collocated events, (Leibovici et al. 2008, 2009 and 2010), to allow better performance for the WPS encapsulating them. These WPS are working with the R package (R Development Core Team, 2007) as back-end (Williams et al. 2010), which we believe to provide added flexibility about implementing classical, modified and new statistical (spatial) methods in epidemiology and public health.

## DISCUSSION

This is a position paper describing some necessary developments of the OGC standard (WPS, and workflow in development) in order to promote its use in the domain of public health and epidemiology for different users. Even though geospatial “mashups” raise common issues for many fields, some particular aspects for public health and epidemiology need attention. Confidentiality and privacy are amongst the main sources of these issues. We described a solution regarding the antagonism between obtaining pertinent fine scale results from a workflow and preserving the privacy

of citizens: the “blind access” principle. We also presented the principle of “updating WPS” to increase performance at computational level.

Privacy regarding tracking capabilities was not fully discussed here, but in the first instance it could appear to pose similar confidentiality principles as in traditional population surveys, nonetheless the spatial components (“scent-trail” or contextual foot-prints) can contain even more sensitive information. The principle of “blind access” is also applicable for this kind of data.

A recent review on software for disease surveillance (Robertson and Nelson, 2010) pointed out that flexibility and performance are very important features. The reviewed desktop applications but mentioned the potential of web applications, benefiting from increased computing power and architecture; we believe the principles developed in the present paper to be effective in these aspects for spatio-temporal data as well. Different experiments and scenarios implementing the above principles for enhancement of the web service standards will be demonstrated at the AGILE conference.

## BIBLIOGRAPHY

- BPMN 2009 Business Process Model Notation version 2 beta 1. OMG (Object Management Group), <http://bpmn.org/>
- BPEL 2007 Standard WS-BPEL 2.0. OASIS, <http://docs.oasis-open.org/wsbpel/2.0/wsbpel-v2.0.pdf>
- Boulos M., Scotch M., Cheung K., and Burden D., 2008 Web GIS in practice VI: a demo playlist of geo-mashups for public health neogeographers. *International Journal of Health Geographics*, 7(1), 38.
- Chang, A., Parrales M., Jimenez J., Sobieszczyk M, Hammer S, Copenhaver D, et al., 2009. Combining Google Earth and GIS mapping technologies in a dengue surveillance system for developing countries. *International Journal of Health Geographics*, 8(1), 49.
- Gao S., Mioc D., Anton F., Yi X., Coleman D., 2008 Online GIS services for mapping and sharing disease information. *International Journal of Health Geographics*, 7(1), 8.
- Grundmann H, Aanensen DM, van den Wijngaard CC, Spratt BG, Harmsen D, et al. 2010 Geographic Distribution of *Staphylococcus aureus* Causing Invasive Infections in Europe: A Molecular-Epidemiological Analysis. *PLoS Med* 7(1): e1000215.
- Kulldorff M., and Nagarwalla N., 1995 Spatial disease clusters: Detection and inference. *Statistics in Medicine*, 14(8), 799-81
- Lawson A., Gangnon R., and Wartenberg D., 2006 Developments in disease cluster detection. *Special Issue: Statistics in Medicine* 25, (5)
- Leibovici D.G., 2009 Spatio-temporal Multiway Decomposition using Principal Tensor Analysis on k-modes: the R package PTak. *Journal of Statistical Software* (accepted August 2009)
- Leibovici D.G., Bastin L., and Jackson M., 2008 Discovering Spatially Multiway Collocations. *GISRUUK Conference 2008, Manchester, UK*.
- Leibovici D.G., Bastin L., and Jackson M., 2009 Higher Order Cooccurrences in Point Pattern Analysis and Decision Tree Clustering. *Computers & Geosciences*, (submitted)
- Leibovici D.G., Bastin L., Anand S., Swan J., Hobona G., and Jackson M., 2010 Spatially Clustered Associations in Health GIS. *GISRUUK Conference 2010, London, UK*.

- Lowy F.D., 2010 Mapping the Distribution of Invasive Staphylococcus aureus across Europe. *PLoS Med* 7(1): e1000205
- OGC 2010 OpenGIS® Standards and Related OGC documents. *Open Geospatial Consortium: website*, <http://www.opengeospatial.org/standards>
- R Development Core Team 2007 R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Robertson, C and Nelson, T (2010) Review of software for space-time disease surveillance. *International Journal of Health Geographics*, vol. 9, 2010, p. 16
- XACML 2005 eXtensible Access Control Markup Language. *Organization for the Advancement of Structured Information Standards* [http://www.oasis-open.org/committees/tc\\_home.php?wg\\_abbrev=xacml](http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml)
- XPDL 2008 XPDL 2.1: [XML](#) Process Definition Language version 2.1. *WorkFlow Management Coalition*. <http://www.wfmc.org/xpdl.html>
- Williams, M\*, Cornford, D, Bastin, L, Jones, R and Parker, S (2010) Automatic processing, quality assurance and serving of real-time weather data. *Computers and Geosciences*. (accepted)