

Accessing Distributed Geographically-Referenced Population Health Information via the Web.

Alan Willmore¹, Tim Churches¹, James Murty² and Kerry Taylor²

¹Centre for Epidemiology and Research
New South Wales Department of Health. NSW, Australia
Phone: +61 2 9391 9226 Fax: +61 2 9391 9232
Email: awill@doh.health.nsw.gov.au

²Mathematical and Information Sciences
Commonwealth Scientific & Industrial Research Organisation. ACT, Australia
Phone: +61 2 6216 7038 Fax: +61 2 6216 7111
Email: kerry.taylor@cmis.csiro.au

ABSTRACT

The utility of geographically-referenced information for informing population health interventions and policy development is well accepted. However, there are a number of challenges in making such information widely available in electronic form. These include: a) how to provide access cheaply to many hundreds or even thousands of staff while still allowing user-driven exploration of the data and the serendipitous discovery of correlations and associations; b) how to integrate constantly changing data from many, disparate sources without the overhead and expense of replicating data into a centralised data warehouse; c) how to restrict access to confidential data on a role-based, need-to-know basis?

This paper describes a Web pilot facility developed to address some of these challenges within the New South Wales public health system (NSW Health). NSW Health is a large organisation with approximately 100,000 employees spread across 17 administrative regions. It provides publicly-funded health services to the 6 million residents of the state of New South Wales, Australia through approximately 300 hospitals, 400 community health centres, 17 public health units and a number of administrative and research centres.

The pilot facility is based on Internet Marketplace (IMP) technology developed by CSIRO, the Australian government-owned scientific and industrial research organisation. In the IMP framework, each data source publishes an XML-based description of the data it is able to provide, what constraints are placed on that data (such as who can access it), and in what formats the data can be provided (eg as map vectors, as a raster image or as a table of attributes). A wrapper, which can be written in any programming language, is provided to enable each data source to service data requests which conform to its published service interface. For example, a Java wrapper is available to allow locally maintained Microsoft Access databases to act as data sources. End users interact with the system via a Java-based thin-client interface which runs as an applet within their Web browser. This interface allows the selection, navigation and querying of map layers and data points. User actions are translated into XML-based service requests, which are relayed to the appropriate data provider (data source) via an integration server. Results are transmitted back to the client in Scalable Vector Graphic (SVG), raster and tabular (HTML) formats. All exchanges between elements of the system are wrapped in the Hypertext Transfer Protocol (HTTP) to allow easy negotiation of firewalls and other inter-organisation barriers.

The initial pilot facility will provide staff in public health units in NSW with access to a range of information including demographic and socio-economic indices, measures of maternal and child health, death, cancer and communicable diseases incidence rates, and ambulance attendance data at a range of geographic scales ranging from NSW-wide through local government area and suburb level right down to Census Collector Districts (about 200 households). Access to Landsat images of rural areas and 1 metre and 40 cm resolution aerial photographs of urban areas is also provided to allow rapid assessment of aspects of the physical environment such as shade cover in school playgrounds. The performance of the underlying technology, training requirements and initial user reactions and acceptance will be discussed.

INTRODUCTION

In September 2000, the Australian National Public Health Partnership issued a statement of what it collectively considers to be the core functions of public health practice. [1] These are:

- Assess, analyse and communicate population health needs and community expectations
- Prevent and control communicable and non-communicable diseases and injuries through risk factor reduction, education, screening, immunisation and other interventions
- Promote and support healthy lifestyles and behaviours through action with individuals, families, communities and wider society
- Promote, develop and support healthy public policy, including legislation, regulation and fiscal measures
- Plan, fund, manage and evaluate health gain and capacity building programmes designed to achieve measurable improvements in health status, and to strengthen skills, competencies, systems and infrastructure
- Strengthen communities and build social capital through consultation, participation and empowerment
- Promote, develop, support and initiate actions which ensure safe and healthy environments
- Promote, develop and support healthy growth and development throughout all life stages
- Promote, develop and support actions to improve the health status of Aboriginal and Torres Strait Islander people and other vulnerable groups.

All of these functions involve analysis and modification of the complex relationships between populations (or sub-populations) and their social, economic, built and natural environments. Not surprising, it is now widely acknowledged that geographic information systems (GIS) are essential to modern public health practice.

This poses a number of specific challenges in the provision of public health GIS facilities. These include:

- a) How to provide access cheaply to many hundreds or even thousands of public health workers, while still allowing user-driven exploration of the data and the serendipitous discovery of correlations and associations? Not all of these prospective users will be full-time public health practitioners – increasingly clinicians involved in the delivery of health care services also have (or ought to have) a population health perspective.
- b) How to integrate constantly changing data from many, disparate sources without the overhead, expense and “ownership” concerns associated with the replication of data into a centralised data warehouse? Modern public health practice tends to be rather post-modern in its approach, and therefore has quite eclectic information needs. An abbreviated list of areas of public health responsibility and concern appears in Appendix 1. Very often essential information exists outside the boundaries of the traditional “health system”, and lightweight mechanisms for accessing these data across organisational borders are required.
- c) How to restrict access to confidential data on a role-based, need-to-know basis? Public health practice is, ultimately, about intervention in order to improve peoples’ health, and

interventions often have to be tailored to local conditions. As a result, there is growing demand amongst public health practitioners for ever more detailed data, with correspondingly increased risk of invasion of individual privacy or loss of confidentiality. Hence the need for secure authorisation and fine-grained access controls in public health GIS.

The New South Wales Department of Health (NSW Health) is a large organisation with approximately 100,000 employees spread across 17 administrative regions. It provides publicly-funded health services to the 6 million residents of the state of New South Wales, Australia through approximately 300 hospitals, 400 community health centres, 17 public health units and a number of administrative and research centres. In the following sections we examine how the GIS architecture which we selected for a pilot public health GIS for the NSW Health system meets the challenges listed above, and we discuss areas in which further development and evaluation is required.

METHODS

System Design and Architecture

PHIMP (the working title of the pilot system) is based on Internet Marketplaces (IMP) technology which has been developed by the Mathematical and Information Sciences Division of the Commonwealth Scientific and Industrial Research Organisation (CSIRO). IMP technology provides internet/intranet access to dispersed heterogeneous information sources, delivering geographically-referenced point, vector and raster data to the end user [2].

IMP is implemented as a community of request service providers. Services establish their own externally visible domain model by registering an external schema (using XML) with a registry service. This registry service provides details of the external schema, structural transformation options and output presentation formats. The NSW Health registry service (portal) is responsible for receiving requests from the client applet, inspecting the set of IMP services available to determine which can provide data to satisfy the client's request, performing queries on these IMP services to obtain data and (optionally) merging multiple result documents. IMP services also require settings that determine how many service instances are to be pooled (i.e. how many queries may be executed simultaneously) and how long a client's request may wait for a service to become available if they are all busy.

An IMP service is a wrapper that interfaces between the IMP architecture and an existing third-party data source or application. Wrappers are custom-written, simple software modules that sit between specific data source applications and the rest of the IMP system. They translate queries from the generic Request Specification Language (RSL), which is a declarative and extensible query language, into the syntax required by the underlying third-party software (data source). Results are sent back through the wrapper and returned to the client application in Scalar Vector Graphics (SVG), raster and tabular (HTML) formats. The client application renders the SVG data, and displays it in conjunction with raster data.

Third-Party Software Wrappers

Wrappers have been written for a number of data provider applications, including a PointRDB service for serving point, thematic and attribute data from Microsoft Access and MySQL relational databases, a PostGIS service for PostgreSQL with the PostGIS extension (point, thematic, attribute, vector and polygon data), and an ImageMap service for tiled TIFF and JPEG raster formats.

MySQL is open source relational database software used to implement the NSW Health Project Portal (registry service) as well as to serve point data. A database of thematic data layer information is maintained and referenced by the registry service to construct and execute queries. Point datasets currently include locations of hospitals, attendances by ambulances at drug overdoses, and needle/syringe exchange outlets.

PostGIS adds support for geographic objects to the open source PostgreSQL object-relational database. In effect, PostGIS spatially enables the PostgreSQL server, allowing it to be used as a back-end spatial database for geographic information systems. Data is conveniently imported through batch processing

of ESRI shape files. Datasets currently served by PostGIS include demographic and socio-economic indices, schools, measures of maternal and child health, death, cancer and communicable diseases incident rates.

The ImageMap service serves tiled image data, combining tiles as required to match query boundaries. The images to be tiled may be in jpeg or tiff format. Landsat-7 Thematic Mapper scenes with 30 metre resolution are included for all of NSW to provide a backdrop for rural areas. Aerial photographs with 1 metre and 38 centimetre resolution of urban areas (including country towns) are provided to allow rapid assessment of aspects in the built urban environment, such as locations of air-conditioning cooling towers, assessment of green space for exercise, and assessment of shade cover in school playgrounds.

Supporting Software

The PHIMP portal is implemented on a server running Red Hat Linux release 7.1, using the Apache open source web server. Apache is used to serve static content over the internet using the HTTP and HTTPS (secure HTTP) protocols. More specifically, in the NSW Health Pilot Project it is used to interface with the Jakarta Tomcat application servlet to run Java programs. Both the portal and IMP services are configurable to run on Unix, Linux, Windows and Apple Macintosh OS X systems.

Jakarta Tomcat is an open source application server that implements the Java Servlet (registry, IMP services) and Java Server Pages (administrative database tool) web service technologies. Tomcat handles execution of Java programs while Apache handles the HTTP and HTTPS connections.

The Java Software Development Kit (SDK) is an open source toolkit for compiling and running Java programs. The Java SDK provides the tools necessary to run Java programs (such as IMP services). It also comes with security tools for doing simple X509 security certificate management.

OpenSSL is an open source toolkit implementation of the Secure Socket Layer (SSL) and Transport Layer Security (TLS) protocols, including support for manipulation of X509 certificates. Open SSL is used to create X509 certificates that securely identify the various components of the NSW Health Pilot Project system, and to encrypt all data transmitted between elements of the system.

The software required on the client computer to run the PHIMP applet comprises the Sun Java 2 Runtime Environment (version 1.4.0 or later), and of course the PHIMP Java applet code, which runs inside the user's Web browser. The client applet incorporates an open source SVG rendering library written by the CSIRO.

Although almost all of the components used to construct and run PHIMP are available under open source licenses, the Java program code for the IMP services themselves is proprietary. However, the heavy reliance on free software for infrastructure helps to keep the overall cost of the IMP implementation down, and dramatically simplifies software licensing arrangements in organisations comprising hundreds of distinct business and organisational units, such as NSW Health.

Security

Communication between the portal (registry) and external data sources takes place over HTTPS connections, in which both ends of the communication adequately identify themselves as trusted entities in the system. X509 certificates are a standard way of storing uniquely identifying information about an entity (e.g. a computer or a user). End-points (e.g. registry, IMP services) in the system uniquely identify themselves using an X509 certificates which must be "signed" as trusted by the security administrator of the system. All data passed over an HTTPS connection is encrypted using 128 bit session keys.

The portal is the only entry point into the system for clients. Upon contacting the portal a client must provide a valid username and password. When the portal subsequently dispatches a query on behalf of a client it passes the client's username and password information to each data source (IMP service),

allowing each data provider to restrict access by particular users to particular datasets as they feel is appropriate.

Access control is configured using XML text files at each data source location, and is thus directly under the control of each data provider. The security configuration files use simple rules to allocate a particular client into roles, and to specify which role-members are to be granted access to particular data layers. A client may fit into more than one role (such as both general user and privileged user) and each role can be granted (or denied) access to any number of data layers.

System Administration

When a data layer is made available via an IMP service the portal must be made aware of the new layer. This is achieved with an administrative database tool. The tool enables an administrator to:

- register information about new data layers to be displayed by the applet
- modify and view information about existing data layers
- remove existing layers from the repository so that these are not displayed by the applet

Information required for each thematic layer includes:

- Name - a unique name
- Type - service type (PointRDB, PostGIS, ImageMap)
- URL - the address specifying the location of the data layer
- Z-Axis - the order in which the layer will be presented in the NSW Health Applet
- Map Coordinates of Region – the bounding coordinates of the region covered by the data layer, specified in latitude and longitude.
- Maximum Area of Region – the maximum area for which the layer will be displayed when queried
- Minimum Area of Region – the minimum area for which the layer will be displayed when queried

Access to layers, including point data, is controlled by the areal extent of the current map view (Minimum/Maximum Area of Region for each layer). When viewing the entire state of NSW, data compiled to Area Health Service (AHS) and Statistical Local Area (SLA) boundaries is displayed. As the zoom level increases (when viewing smaller areas) more detailed data at Census Collector District (CCD) level (about 200 households) is available and the AHS- and SLA-level data become unavailable. This mechanism greatly improves the performance of the system by reducing the amount of data that needs to be processed each time a user selects a new view (and the client application dispatches a corresponding query).

Client Interface

When the client applet is launched the user is prompted for a username and password that provides the credentials for access to the available thematic, point and raster layers.

The main applet window displays the mapped data and provides toolbar buttons that allow the user to:

- load data for a new map region
- query the attributes of individual map objects
- navigate around the map
- print the current map as it appears in the main map window
- email the system administrators for support
- access the online help

The current position of the mouse and view dimensions are shown in the status bar at the bottom of the applet window, with a radio button to toggle between decimal degrees and degrees/minutes/seconds.

The Control Panel window provides information about each layer displayed in the applet window. From here the user can view metadata about a selected layer, modify the appearance of the map and save map data as an SVG document. A graphic area at the bottom of the window displays a context map showing the boundary of the study area (region being viewed in the main applet window) relative to the NSW state border. The appearance of the map may be altered by making layers visible/invisible and by making layers partially transparent.

DISCUSSION

Performance

Currently the PHIMP pilot system is running on a single PC with a Pentium III 866MHz processor and 512 Megabytes of RAM. Queries on the system at a state wide (regional) level and at high resolution (street) level respond in 15 – 30 seconds. Performance degrades when performing queries at the district level (a view with extents covering 2 – 5 SLAs). Delays of up to two minutes can be experienced, which is unacceptable for staff used to working with high speed internet connections. Delays can be minimised by reducing the Maximum Area of Region for CD compiled layers so that smaller areas need to be viewed to for these layers to be queried. Reducing the Maximum Area of Region for particular layers has the trade-off of not allowing users to see high resolution datasets at the district or regional scale.

Improvements

The processing demands of the query planner part of the portal have been identified as the main bottleneck. Network bandwidth appears to be only a secondary constraint. Prior to commencement of a full trial the PHIMP system is to be installed on two Pentium IV personal computers each with a 1.8 GHz processor and 1 Gigabyte of RAM. Increased processing power, and sharing IMP services across two computers is expected to improve the response times of the system. This will allow the Maximum Area of Region values to be increased for viewing of high resolution datasets at the district or regional scale.

Configuration of the system required significant initial effort with much painstaking editing of XML files in a text editor. Similarly, addition of IMP services and data layers to the system is not a simple task. Further development of the IMP architecture is needed to allow new data sources to be more easily incorporated and to provide graphical user interfaces for other administrative tasks.

Additional functionality in the client Java applet has been identified as an area for further development. Tools to allow spatial analysis, distance calculations and selection will ultimately be needed as users grow in sophistication.

Perhaps the biggest challenge will be developing methods of allowing users to search and navigate through, potentially, many hundreds of data layers, and to choose a smaller “working set” of layers of immediate interest to them. Development of a taxonomy of public health practice suitable for such use is an essential first step, and this work is currently underway.

CONCLUSIONS

The IMP architecture appears to meet most of the requirements which we have identified for a public health GIS in a large, distributed organisation such as NSW Health, at least in theory. IMP should also be relatively easy to enhance due to its open architecture and the wide availability of the software tools and components from which it is constructed.

How well this architecture performs in practice remains to be seen. A roll-out of the pilot system to Public Health Units in NSW has commenced, and we hope to be able to report on the performance, utility and user acceptance of the system within the next six months.

REFERENCES

[1] National Public Health Partnership: **Public Health Practice in Australia Today: A Statement of Core Functions**. Melbourne: National Public Health Partnership Secretariat 2000.

[2] MA Cameron, KL Taylor, DJ Abel: **The Internet Marketplace Template: An Architecture Template For Inter-enterprise Information Systems**, *Sixth International Conference on Cooperative Information Systems (CoopIS 2001)*, Trento, Italy, September 2001.

[3] National Public Health Partnership: **National Delphi Study on Public Health Functions in Australia**. Melbourne: National Public Health Partnership Secretariat 2000.

APPENDICES

Appendix 1 – Selected Public Health Functions [3].

| | |
|---|---|
| <p>1 Research, monitoring and assessment</p> <p>1.1 Identifying health research priorities</p> <p>1.2 Undertaking research on health issues in populations</p> <p>1.3 Undertaking research on health services and programs</p> <p>1.4 Assessment of population needs and risks to determine which groups require services</p> <p>1.5 Monitoring the delivery of health services and programs</p> <p>1.6 Monitoring morbidity and mortality</p> <p>1.7 Monitoring the determinants of health</p> <p>1.8 Evaluation of health policies</p> <p>1.9 Evaluation of health care services and programs</p> | <p>7 Prevention, surveillance and control of communicable diseases</p> <p>7.1 Screening for selected communicable diseases</p> <p>7.2 Immunization provision</p> <p>7.3 Disease surveillance</p> <p>7.4 Disease outbreak control</p> <p>7.5 Treatment of infectious disease cases</p> <p>7.6 Veterinary public health services</p> <p>7.7 Risk assessment and management in relation to environmental hazards</p> <p>7.8 Public health laboratories</p> |
| <p>2 Ensuring healthy and safe environments</p> <p>2.1 Town planning and land use</p> <p>2.2 Controlling housing standards and provision</p> <p>2.3 Ensuring drainage, sewerage and solid waste disposal</p> <p>2.4 The production and protection of safe water</p> <p>2.5 Controlling land degradation and soil loss (e.g. by erosion)</p> <p>2.6 Ensuring access to public transport and educational opportunities</p> <p>2.7 Ensuring access to physical activity and other recreation facilities</p> <p>2.8 Ensuring access to facilities for social interaction</p> <p>2.9 Vector control</p> <p>2.10 Controlling hazardous substances and wastes</p> <p>2.11 Controlling atmospheric pollution</p> <p>2.12 Controlling noise</p> <p>2.13 Controlling radiation</p> <p>2.14 Controlling food access, quality and safety</p> <p>2.15 Controlling the quality of therapeutic goods and appliances</p> <p>2.16 Ensuring safe workplaces</p> | <p>8 Prevention, surveillance and control on non-communicable diseases</p> <p>8.1 Behavioural risk factor surveillance</p> <p>8.2 Screening for selected non-communicable diseases (e.g. cervical cancer, diabetes)</p> <p>8.3 Chronic disease self management</p> <p>8.4 Provision of cholesterol lowering and anti-hypertensive drugs</p> <p>8.5 Provision of anti-depressant drugs</p> <p>8.6 Drug treatment and rehabilitation services (e.g. methadone)</p> <p>8.5a Lifestyle programs</p> |

| | |
|---|--|
| <p>3 Health education and community development</p> <p>3.1 Developing individuals' health knowledge, attitudes, practices and skills</p> <p>3.2 Providing health information and education to individuals and groups</p> <p>3.3 Developing community capacity to participate in health issues</p> <p>3.4 Community mobilisation & advocacy for health issues</p> <p>3.5 Developing community capacity to protect and promote health</p> <p>3.6 Building social networks and social support in communities</p> <p>3.7 Social marketing of health information</p> <p>3.8 Providing health information and education in community languages</p> | <p>9 Prevention and surveillance of injuries</p> <p>9.1 Prevention of intentional injury</p> <p>9.2 Prevention of non-intentional injury</p> <p>9.3 Injury surveillance</p> |
| <p>4 Health policy and public policy development and implementation</p> <p>4.1 Assessing the impact of other sectors' policies on health</p> <p>4.2 Assessing the differential impact of health policies on disadvantaged communities</p> <p>4.3 Ensuring universal access to health care services</p> <p>4.4 Developing fiscal strategies that support health improvement</p> <p>4.5 Developing sustainable financing for health protection and promotion</p> <p>4.6 Developing financial incentives to encourage preventive health care</p> <p>4.7 Reviewing, formulating & enacting health legislation</p> <p>4.8 Developing and advocating for legislation and regulations that protect and promote health (e.g. occupational health & safety standards, road safety legislation)</p> <p>4.9 Enforcing health legislation and regulations</p> | <p>10 Healthy growth and development programs and services</p> <p>10.1 Family planning services</p> <p>10.2 Prenatal and neonatal screening</p> <p>10.3 Maternal and child health care (not including pregnancy)</p> <p>10.4 School health services</p> <p>10.5 Individual dental check ups</p> <p>10.6 Individual medical check ups</p> <p>10.7 Prevention based care from alternative or complementary therapists</p> <p>10.8 Mental health promotion programs</p> |
| <p>5 Public health education and training</p> <p>5.1 Providing education and training for public health practitioners</p> <p>5.2 Providing education and training for other health care workers</p> <p>5.3 Providing education and training for professionals in other sectors</p> <p>5.4 Providing education and training for consumers and community groups</p> | <p>11 Programs and services directed at specific population groups and individuals</p> <p>11.1 Generating an organised approach to dealing with the consequences of chronic illness and disability</p> <p>11.2 Health services for indigenous peoples</p> <p>11.3 Health services for refugees</p> <p>11.4 Health services for homeless people</p> <p>11.5 Health services for victims of violence and other crises</p> <p>11.6 Emergency and natural disaster services to protect health</p> <p>11.7 Mental health care services</p> <p>11.8 Occupational health services</p> |
| <p>6 Public Health management</p> <p>6.1 Building organisational structures and processes for public health within agencies</p> <p>6.2 Building organisational links between health agencies</p> <p>6.3 Building links between the health sector and other sectors</p> <p>6.4 Developing and implementing quality assurance processes for public health</p> <p>6.5 Human resource development in the public health workforce</p> | |

| | |
|---|--|
| 6.6 Developing resource allocation and priority setting systems | |
|---|--|