

# Health Atlas Ireland

**Howard Johnson,**

Health Intelligence, Health Service Executive,

Tel +353-1-6352040.

Email [howard.johnson@mailf.hse.ie](mailto:howard.johnson@mailf.hse.ie)

**Mel McIntyre,**

Managing Director, OpenApp,

Tel +353 1 614 5978,

Email [mel.mcintyre@openapp.ie](mailto:mel.mcintyre@openapp.ie)

## Introduction

The purpose of this paper is to introduce Health Atlas Ireland, an open-source software(OSS) mapping, database and statistical system, briefly explain what it is designed to achieve and how it is developed and maintained.

Health Atlas Ireland integrates GIS, database and statistical components. A user-friendly interface supports web-enabled access across the Irish health sector and collaborating agencies.

Grant aided by the Health Research Board, the Atlas is custom built by the public health sector in partnership with two universities. Epidemiological and service location analyses are central themes.

## Purpose and usage

Health Atlas is a 'voyage of discovery' for health service planning and health event data analysis. The purpose of the system is to **help answer questions** related to health events, emergency response, health services and demographics, initially in the Republic of Ireland and eventually worldwide as related to Irish Health Services. The methodology employed is to answer all questions in '**real time**', from raw data – there is no preprocessed data.

- Show me the population within defined distances of a hospital – say, 5 Km, 10Km, 15Km
- Analyze my point data for road traffic accidents and overlay them on a themed kernel map
- Correlate coronary disease with deprivation levels and display as bi-variate symbols by area.

## Points of Interest

Aspects of the Atlas we would like to highlight are the following:

**Presentation** - Presents data in a standard Web Browser as charts and maps without requiring plug-ins.

**Analysis** - Analyses raw data, supports small number suppression, statistically processed data (indirect and direct standardisation, Bayesian smoothing, cluster tests (Kernel, Potthill Whittinghill, Stone)), supports multivariate cartograms, allows drill down to small area and point data.

**Open Source** - Developed exclusively in open source software – R, GRASS, PostGIS, Mapserver, Zope3

○ The development and production environments also utilize many OSS components – Linux, Xen, OpenLDAP, Subversion, Trac

**Security** - Ensures data protection through an information governance framework - role-based access depending on training/skill, agency, and need to analyze or view confidential data.

**Integration** – Integrates many disparate datasets and map sets

○ Many datasets – census; hospital activity; cancer incidence; road collision locations, mortality, births, perinatal data; vaccine uptake. More being added monthly.

○ GeoDirectory (national address database) – generates point data map layers from address lists in csv or MS Excel formats.

○ Supports upload of users point or polygon data from many format.

○ Multiple map sources - Ordnance Survey Ireland raster, Ortho and Vector maps Map layers: Ordnance Survey Ireland raster (including aerial photography) and vector maps from national to house level.– Worldmaps from free web services - Geological Survey Ireland – Radiology

**Flexible Output** – Supports A0 to A4 map outputs in PDF and image formats, report templates, snapshots of queries, table export to csv or Excel.

## System Development

The system is developed by selecting best of breed open source software components and integrating them in a manner that delivers a seamless system view to the end user.

The popular open source methodology of 'release early, release often' fits well with the user interaction and uptake.

Starting in February 2006, the system has been delivered in five phases to date

- Phase 1 – Basic data selections and map rendering from census and electoral divisions.
- Phase 2 – Add health datasets, (HIPE, Cancer) and implement SMRs.
- Phase 3 – add A0 printing, manual data upload, point data analysis, statistical tests.
- Phase 4 – integrate geodirectory and add network analysis capability.
- Phase 5 – refine security, usability and performance.

The components integrated to date are as follows:

**PostGIS** - PostgreSQL is a powerful, open source relational database system. It has more than 15 years of active development and a proven architecture that has earned it a strong reputation for reliability, data integrity, and correctness. PostGIS adds support for

geographic objects to the PostgreSQL object-relational database. In effect, PostGIS "spatially enables" the PostgreSQL server, allowing it to be used as a backend spatial database for geographic information systems (GIS), much like ESRI's SDE or Oracle's Spatial extension. PostGIS follows the OpenGIS "Simple Features Specification for SQL" and has been certified as compliant with the "Types and Functions" profile.

**The R Project for Statistical Computing** - R is a language and environment for statistical computing and graphics. It is similar to the S language and environment which was developed at Bell Laboratories. R provides a wide variety of statistical (linear and nonlinear modelling, classical statistical tests, time-series analysis, classification, clustering, ...) and graphical techniques, and is highly extensible. The S language is often the vehicle of choice for research in statistical methodology, and R provides an Open Source route to participation in that activity.

**MapServer** - a development environment for building spatially-enabled internet applications. We use MapServer primarily for at rendering spatial data (maps, images, and vector data), and it can facilitate multiple projections (through Proj4)and many data formats. Mapserver is supported by many open source libraries such as GD, GDAL,

**GRASS** - Commonly referred to as GRASS, this is a Geographic Information System (GIS) used for geospatial data management and analysis, image processing, graphics/maps production, spatial modeling, and visualization. This is primarily a desktop system utilized as a server process to write it's output to PostGIS.

**Zope3** – providing the web interface, web server, application logic and many content management functions such as workflow, site membership, role based security and integration of all application components. The user interfaces uses many Ajax and Javascript libraries, PrimaGIS, Nuxeo CPS and integration and logic customizations are written in Python and Python Cartographic Libraries(PCL.

One of the features of the open source development methodology is that experts are often available to provide training and support. In the case of Health Atlas Ireland, Sean Gilles (US – PCL author) and Kai Hainnenan (Finland - PrimaGIS) provided invaluable support for user interface and Mapserver integration while Otto Dassau (Germany ) of Nature Consult has provided training and coaching on GRASS.

Project code management is achieved with **Subversion** whereby each developer checks in code enhancements from their own development system, allowing us to maintain multiple system images at any one time. Each check in triggers **BuildBot** which runs a comprehensive suite of tests which minimizes regression testing at system level. Issues and errors are reported and tracked in **Trac** and notified by email to all users..

Going forward, the system will evolve in three broad areas

1. Addition of new health and environmental datasets, update of all existing datasets and correlation across datasets.
2. Addition of new functionality for statistical processing further utilizing the available R libraries and further exploitation of available GRASS functionality.
3. Improvements in usability and provision of training to support broad usage throughout the Health Service Executive (HSE) and related agencies.

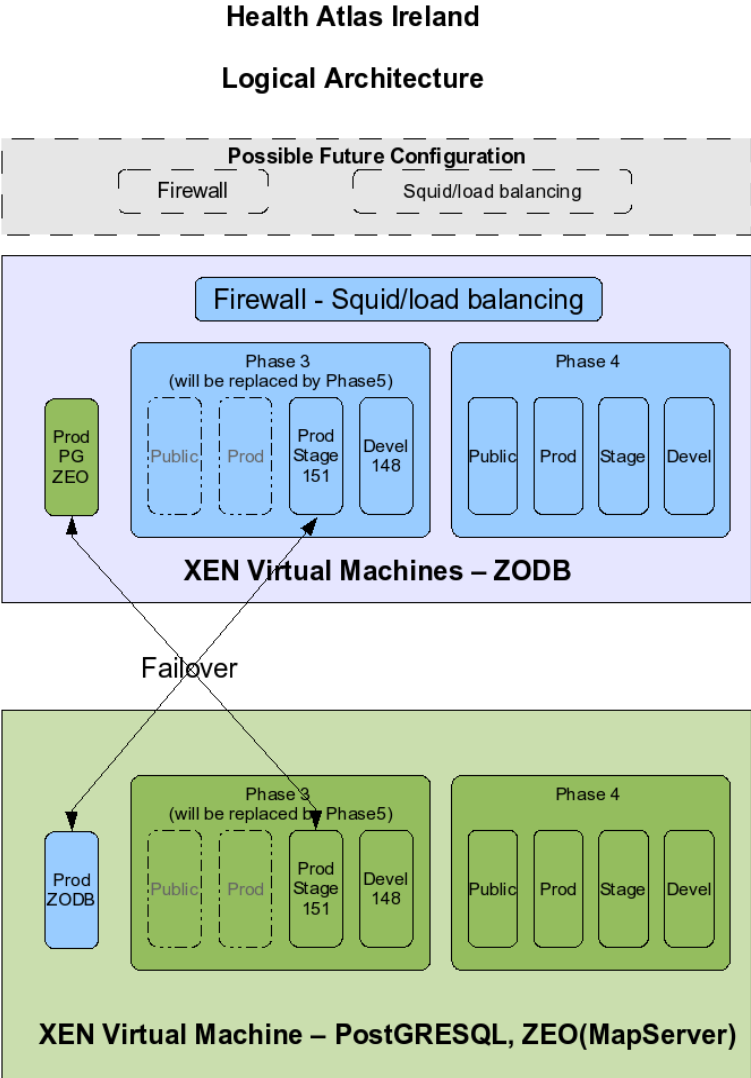
## **System Deployment**

Due to the requirement to collaborate with multiple agencies within and outside the HSE,

the system is deployed at Higher Education Authority Network (HEANET) which provides secure access, over the public Internet, to all public sector and private agencies.

There are currently two servers deployed, each running Xen Virtualisation on Linux. This allows us to implement separate instances for Development Staging, User Acceptance Testing, Production and Public access for each phase of the project currently in use.

Illustration 1: Possible system architecture.



## Security

A number of artifacts are used to implement a stringent security model.

1. User access is by assignment to 'Local Roles' (Viewer, Researcher, Administrator) in defined Agency/Section Folders.
2. A Zope3 custom dataset utility allows the system to connect separate databases to agency folders, enabling agencies to administer their own data outside the Health Atlas.
3. User security is implemented through User Identifier and Password, maintained in an OpenLDAP database and managed through PHP Ldap Admin, enforcing strong passwords and periodic update.
4. Network security is implemented through Virtual Private Networks which ensure only named computers can access the system. HTTPS is used as the transmission protocol which encrypts data transmitted and ensures it cannot be intercepted.
5. Only port 80 is accessible from external and all Firewall rules are implemented in Iptables. Access for root via ssh is disabled and we use ssh access only to master Xen.
6. The system is monitored by OSSEC (system security through log monitoring), SNORT (network monitoring) and Nagios (system services). SNMP is used by Cacti to monitor system performance (load average, IO utilization and Network usage)<sup>b</sup>

## User Experience to date.

- Data quality is generally less than expected.
- Data integration is the real benefit
- Security of data is a constant concern
- Users seem to easily comprehend the user interface, browser interface and system outputs (maps, charts, tables)

## Project Team

Johnson H<sup>1</sup>, Cullen C<sup>1</sup>, Boyle E<sup>1</sup>, Doyle D<sup>1</sup>, Brazil J<sup>2</sup>, McKeown P<sup>2</sup>, Staines A<sup>3</sup>, Pringle D<sup>4</sup>, Beaton D<sup>1</sup>, McIntyre M<sup>5</sup>, Hennessy C<sup>5</sup>, O'Kiersey C<sup>6</sup>

(1) Health Information Unit, Health Intelligence, National Population Health Directorate, Health Service Executive (HSE), Ireland

(2) Health Protection Surveillance Centre, HSE, Ireland

(3) School of Public Health and Population Sciences, UCD, Ireland

(4) Department of Geography, National University of Ireland, Maynooth, Ireland

(5) OpenApp, Dublin, Ireland

(6) Siemens Ireland