BUILDING A GEO-PORTAL FOR ENHANCING COLLABORATIVE SOCIO-ECONOMIC RESEARCH IN WALES USING OPEN-SOURCE TECHNOLOGY

Robert Berry, Richard Fry and Gary Higgs University of Glamorgan

Scott Orford Cardiff University

ROBERT BERRY is a post-doctoral researcher in Geographical Information Systems (GIS) (based at the University of Glamorgan) working for the Wales Institute of Social and Economic Research Data and Methods (WISERD) Data Integration Team. Rob holds a Bachelors degree in Physical Geography from the University of Leicester and a Masters and PhD in Geographical Information Systems (GIS) from the University of Glamorgan.

RICHARD FRY is part of the Data Integration Team for the Wales Institute of Social and Economic Research, Data and Methods (WISERD) and currently in the process of submitting his PhD. Richard graduated from the Department of Geography at Swansea University, and has been awarded a Masters in GIS from the University of Glamorgan.

GARY HIGGS is currently Director of the GIS Research Centre in the Faculty of Advanced Technology, University of Glamorgan. Gary previously held posts in Cardiff University and the University of Leeds. Over-arching research interests are in the application of GIS in social and environmental studies. He is currently a co-Director of the Wales Institute of Social and Economic Research, Data and Methods (WISERD) at Glamorgan.

SCOTT ORFORD is a Senior Lecturer in Spatial Analysis and GIS in the School of City and Regional Planning, Cardiff University and the co-ordinator of the WISERD Data Team. Scott's research interests include the analysis and modelling of socio-economic processes, innovation in data integration and data construction, and the mapping and spatial analysis of qualitative data.

Correspondence to:
Rob Berry
WISERD, GIS Research Centre
Faculty of Advanced Technology
University of Glamorgan
Pontypridd
Wales, UK
CF37 1DL
raberry@glam.ac.uk

Journal of Applied Research in Higher Education
Volume 2 • Number 1 • pp77–92
January 2010

ISSN: 1758-1184

Journal website: http://jarhe.research.glam.ac.uk
Journal correspondence to: jarhe@glam.ac.uk

University of Glamorgan

Cardiff • Pontypridd • Caerdydd

BUILDING A GEO-PORTAL FOR ENHANCING COLLABORATIVE SOCIO-ECONOMIC RESEARCH IN WALES USING OPEN-SOURCE TECHNOLOGY

Robert Berry, Richard Fry and Gary Higgs
University of Glamorgan

Scott Orford
Cardiff University

Abstract

The Wales Institute of Social and Economic Research, Data and Methods (Wiserd) is a major new collaborative socio-economic research programme involving five higher education institutions in Wales. This paper introduces the work of the Wiserd data integration team and describes their plans for the development of an online geo-portal. Their aim is to support Wiserd researchers by providing a framework for integrating, managing and disseminating quantitative and qualitative socio-economic data in Wales. This paper outlines the goals of this major project, discusses the concept of the Wiserd geo-portal and reports on initial investigations into geo-portal development using free and open-source (Foss) software. The paper concludes with a brief summary of the future work of the Wiserd data integration team.

Key words: Collaborative research, geo-portal, data integration, open-source software, spatial data infrastructure (SDI), metadata.

Introduction

Many of the most detailed and complex problems facing society in the 21st Century call for investigation and analysis based upon the interdisciplinary social sciences. This was one of the main conclusions of the Rhind Commission's report into the state of social sciences research in the UK (CSS, 2003). The report concluded that whilst research was generally organised on a small-scale cottage-industry basis, social science disciplines were seen to have a major role to play in addressing the problems associated with contemporary social and economic change.

The Commission took the view that universities need to develop and support a 'critical mass' of social sciences researchers, providing the basis for developing large, coherent, interdisciplinary teams capable of bringing new approaches to these pressing social issues. This view found support with UK research funding councils including the Economic and Social Research Council (ESRC) which has formulated new funding strategies to support the promotion of such research groups (ESRC, 2009).

In Wales, such moves have been particularly welcome given the wide-ranging socio-economic problems that require immediate social scientific investigation, and the growing need for a strong and independent research base to support and inform policy-makers in the Welsh Assembly Government (Wag). This has been reinforced by the findings of the Rhind Commission which identified deficiencies in the social sciences research capacity of universities in Wales (CSS, 2003). In response, the universities of Aberystwyth, Bangor,

Cardiff, Glamorgan and Swansea have collaborated to establish the Wales Institute for Social and Economic Research, Data and Methods (WISERD)—a venture funded by the ESRC and the Higher Education Funding Council for Wales (HEFCW).

A particular area of concern for social sciences in Wales is in the area of quantitative research methods. This was the focus of a recent scoping study commissioned by the ESRC and HEFCW (Moore *et al*, 2007) which highlighted the need for the creation of a Welsh 'centre' with responsibility for delivering social science research capacity development. It was recommended that such a centre should link in strongly with, and not duplicate, national initiatives in research methods, and should break down barriers to collaboration across disciplines and between institutions. Emerging strongly from the scoping study was the notion that the advancement of quantitative methods in Wales would be greatly facilitated by breaking down barriers between quantitative and qualitative methods, building on strengths and excellence in qualitative methods, and cutting edge development of mixed methods and methodologies. WISERD enables such researchers to come together with the clear aim of building upon and developing existing expertise in quantitative and qualitative research methods. This will create a centre of international excellence in mixed methods and methodologies related to the study of key social and economic problems and their resolution.

This paper introduces the WISERD initiative, highlights the aims of the WISERD data integration theme and describes one key element of this work programme—the development of a web-based geo-portal permitting users to search, view, combine, query and visualise data over the Internet. This will act as a gateway to spatial information "by channelling all searches through a single geo-portal containing a single catalog" (Goodchild *et al*, 2007, p255). This in turn will support social science researchers by providing a framework for integrating, managing and disseminating quantitative and qualitative socio-economic data in Wales. Since it is well recognised that a considerable amount of data collected by social scientists in Wales is sparingly used at best, the purpose of the data integration theme is to bring together, integrate and make more usable the wide range of existing data sets that relate to Wales. This will be the hub around which the other activities develop and a primary focus of preliminary research efforts has concerned initial investigations into geo-portal development using free and open-source software (FOSS). This has involved a review of previous approaches to the development of such tools and their application in environmental, economic and social environments. Many of these initiatives have gone beyond the prototype stage and are being used by central and local government organisations to improve access to public data. This review has formed the basis of preliminary ideas on the construction of open-source geo-portal tools to be adopted within WISERD that should be of relevance to researchers in other contexts where the focus is on interdisciplinary research. In the remainder of the paper, the background to WISERD and the main objectives of the WISERD data integration team are described, followed by an overview of the geo-portal concept and a discussion on the use of open-source software for geo-portal development. The concept and planned development of the WISERD geo-portal is briefly discussed before concluding with a summary of the research conducted to date and an outline of WISERD's future research plans regarding provision of such tools for the benefit of the social science community, both within Wales and further afield.

The WISERD initiative

Background

WISERD brings together expertise from participating higher education institutions across a range of social science disciplines, including criminology, economics and finance, geography and geographical information systems (GIS), public health, urban studies, political science, sociology and social policy. The overarching aim is to improve the quality of social sciences research in each of the institutions and in Wales as a whole, and to provide a strong and sustainable platform of collaborative support and interdisciplinary working across the principality. The work of WISERD is organised around four main sets of thematic activities: (1) data integration and data management, (2) in depth mixed-method studies of selected 'localities', (3) policy analysis and evaluation and (4) training and capacity building within WISERD with a strong emphasis on quantitative, qualitative and mixed methods. This paper concentrates on the data integration and data management theme of WISERD. Work currently being conducted by the WISERD data integration team involves the development of a web-based geo-portal to enable the management, integration and dissemination of socio-economic metadata (i.e. information that describes data characteristics) and data in Wales, within a collaborative, multi-disciplinary research environment.

Data integration team research strands

Geo-portal development

The WISERD geo-portal will provide the central GIS framework for achieving the overall research objectives and will facilitate the integration, management, analysis and dissemination of quantitative and qualitative data relevant to the WISERD programme. The geo-portal will be developed by a team working at the GIS Research Centre, University of Glamorgan, where it will be hosted on a dedicated web server. The tools developed as part of the geo-portal will allow metadata discovery

(which will include the analysis and mapping of metadata) and data download where applicable. Access to the geo-portal and the data it encompasses will vary depending upon the user and their affiliation with WISERD.

Metadatabase construction

The construction of a metadatabase will make an important contribution to substantive research and policy in Wales by providing a foundation for helping to describe how quantitative and qualitative data can be integrated across different dimensions and analysed to determine the strengths and weaknesses of these data. The dimensions include the spatial units of analysis, the temporal units of analysis (describing the continuity and currency of the data), the thematic units of analysis (describing the substantive questions, and/or variables and content), and the categorical units of analysis (describing the coding and breaks and keywords used in classifying the data). The metadatabase will provide the main framework for the collection, integration, management and dissemination of WISERD data. It will also highlight any data deficit across the dimensions which could be used to inform any subsequent large-scale social surveys in Wales.

The main benefit of the metadatabase is that it will bring together the key official data sets for Wales into a single framework, adding value to the individual data sets and allowing easier access by students and researchers alike. The metadatabase will be designed to allow researchers to search (via the geo-portal) for official data based on their own research themes (for example, ethnicity, income, migration) or for specific spatial locations and time periods. It will provide a summary of the relevant variables in the different data sets, their coding/classification schemes, details of survey methods and sample sizes, where the data is held and how it can be accessed—for example, through the Office for National Statistics (ONS), through the UK Data Archive (UKDA), through WISERD etc. The metadatabase will provide a framework through which other data can be integrated, including the data collected under the WISERD initiative. In conducting this research, there exists an opportunity to develop a new metadata standard for use in multi-disciplinary collaborative research projects based on existing robust metadata standards. This phase of the research will also tackle important issues, and develop policies and standards, related to data access, data confidentiality and data disclosure.

'Integrating' quantitative data

When discussing data integration, it is important to make the distinction between data-linking and data-pooling. Data-linking occurs when two or more data records are matched, either directly or probabilistically, to create a single record. The usual case is when data records for individual people or individual households are matched from different data sources. This type of data integration is quite rare, mainly due to the confidential nature of micro-level data which leads to the data being either suppressed or undergoing disclosure measures. These make data-linking very difficult although not necessarily impossible and WISERD is currently developing a research agenda with the Centre for Health Information, Research and evALuation (CHIRAL) at the School of Medicine's Institute of Life Science, Swansea University regarding CHIRAL's unique and innovative data-linking methodology using health record data. Data-pooling occurs when two or more data records are grouped together based upon a common attribute and the data is then used as a collective whole. This is different from data-linking as the number of data-records stays the same before and after the pooling process i.e. they are not linked to form a single data-record. This type of data integration is very common and underpins many data integration functions in GIS (such as some overlay analysis functions) and on the web (such as mash-ups). Data pooling tends to be based upon spatial units such as postcodes, rather than individual people, and is thus very well suited to a geo-portal and GIS approach to data integration.

Building directly on the metadatabase research, the WISERD geo-portal will provide a framework not only to search and access official and WISERD-generated survey data across Wales, but also to have a much clearer understanding of how datasets vary in their detail across different spatial scales, such as electoral divisions, travel-to-work areas and unitary authorities, and different time periods. The metadata will allow researchers to select one or more datasets which provide the necessary variables and crucially, provide some indication as to whether data from these different datasets can be pooled to create a larger dataset. Pooling data may be beneficial if data from individual surveys provide small sample sizes for the location or time period of interest. Even with the Welsh boosts to national government surveys, it can be difficult to analyse data below the national scale and certainly below the unitary authority scale without confronting statistical issues relating to small numbers of observations. Pooling data could be a potential solution to this and could also encourage researchers to increase their use of the diverse and rich datasets collected for Wales.

'Integrating' qualitative data

This strand of the research will address the potential innovative integration of qualitative methodological approaches, data types and analytical strategies. This includes developing, evaluating and disseminating multimodal and mobile approaches to qualitative research practice and generating new approaches for the recording, display and dissemination of complex qualitative datasets, for instance through the development of a qualitative GIS (QGIS) component within the WISERD geo-portal. This is an emerging field and the nas-

cent literature points to ways of using GIS technology to capture local knowledge and provide a better understanding of people's everyday lives (Cope and Elwood, 2009). QGIS accomplishes this by incorporating non-numerical data into GIS which may include digital images, video, sound, sketches and text (for example, from transcripts). Thus, QGIS is inherently a mixed method approach and the analysis of this data within a spatial framework can provide insights into people's perceptions, preferences and values, and how these play out and interact locally.

Although QGIS is relatively simple to conceptualise, its implementation in GIS has proven more difficult. Hence, part of this strand aims to investigate methodological approaches, particularly by drawing upon the community planning and public participation GIS (PPGIS) literature in which stories, drawings and transcripts are the typical sorts of data that are generated and analysed. This attempt to 'map' the various modes and forms of qualitative data within an explicitly spatial framework will help to locate qualitative data in physical space. It will also link, through the geo-portal, to quantitative data collected for the same locations during the locality studies survey, and to official secondary data from Government surveys. This will allow multi-modal qualitative data, such as digital images, sound and text, to be contextualized to the place it was collected. By linking it more strongly to statistical data, it will move towards a framework that gives the data a place in everyday, official, and expert discourses, such as those used in community planning practice.

It is hypothesised that the official datasets, disseminated at the various levels of spatial and temporal aggregation, will paint 'broad brush' statistical generalisations of Wales that miss the spatial detail within and between different places but which will be captured by the locality study data. By moving towards a more nuanced understanding of how general statistical measures vary by local context, this aspect of the research programme will provide a substantive insight into the use of official data in local policy making (by identifying possible deficits and gaps in official data sources, for example), as well as providing an environment for evaluating and promoting mixed methods research, teaching and learning.

Archiving quantitative and qualitative data

One long term goal of WISERD is to provide an archive of social and economic quantitative and qualitative data sets relating to Wales. Principally these will be the data collected as part of the WISERD programme of activities, but it could feasibly also include other data sets to which WISERD researchers have been given access. Long term data archiving is an increasingly important part of the funding council's (for example, the ESRC's) remit and WISERD is working with the UKDA, the Economic and Social Data Service (ESDS) and the ESRC-funded 'Timescapes' project in developing archiving standards and tools that will be incorporated into the geo-portal and the metadatabase.

The purpose of developing the geo-portal is to facilitate effective and innovative collaborative working within a multi-disciplinary research group. The essence of the geo-portal is to permit a distributed group of researchers access to integrated geo-spatial (and non-geospatial) socio-economic metadata and data for Wales. The WISERD data integration team is therefore uniquely positioned for delivering a novel type of geo-portal—the 'collaborative research geo-portal'.

Geo-portals

Definitions and early developments

Tait (2005) defined a geo-portal as:

> *...a Web site that presents an entry point to geographic content on the Web or, more simply, a Web site where geographic content can be discovered.*
>
> **Tait, 2005, p34**

Tait went on to refine this definition when he considered that many modern web sites contain geographic information (e.g. business addresses) and geographic functions (static and/or dynamic maps), adding that a geo-portal should be thought of as a Web site "...where the discovery of geographic information is the prime focus" (Tait, 2005 p34). Geo-portals commonly incorporate dynamic web-mapping interfaces and traditional textual search components in order to allow users to find, and in some cases, download and process geographic information (GI) via the Internet.

Maguire and Longley (2005) traced the origins and growth of geo-portals alongside spatial data infrastructures (SDI) and discussed their importance in simplifying access to geographic information (GI) and GI services. According to the Global Spatial Data Infrastructure Association (GSDI), the term SDI is often used to denote:

> *...the relevant base collection of technologies, policies and institutional arrangements that facilitate the availability of and access to spatial data.*
>
> **GSDI, 2009, p8**

SDIs are often associated with national and international governments as they attempt to maximise the use of GI for the benefit of their citizens, environment and economies. The single most important development for European GI users in this

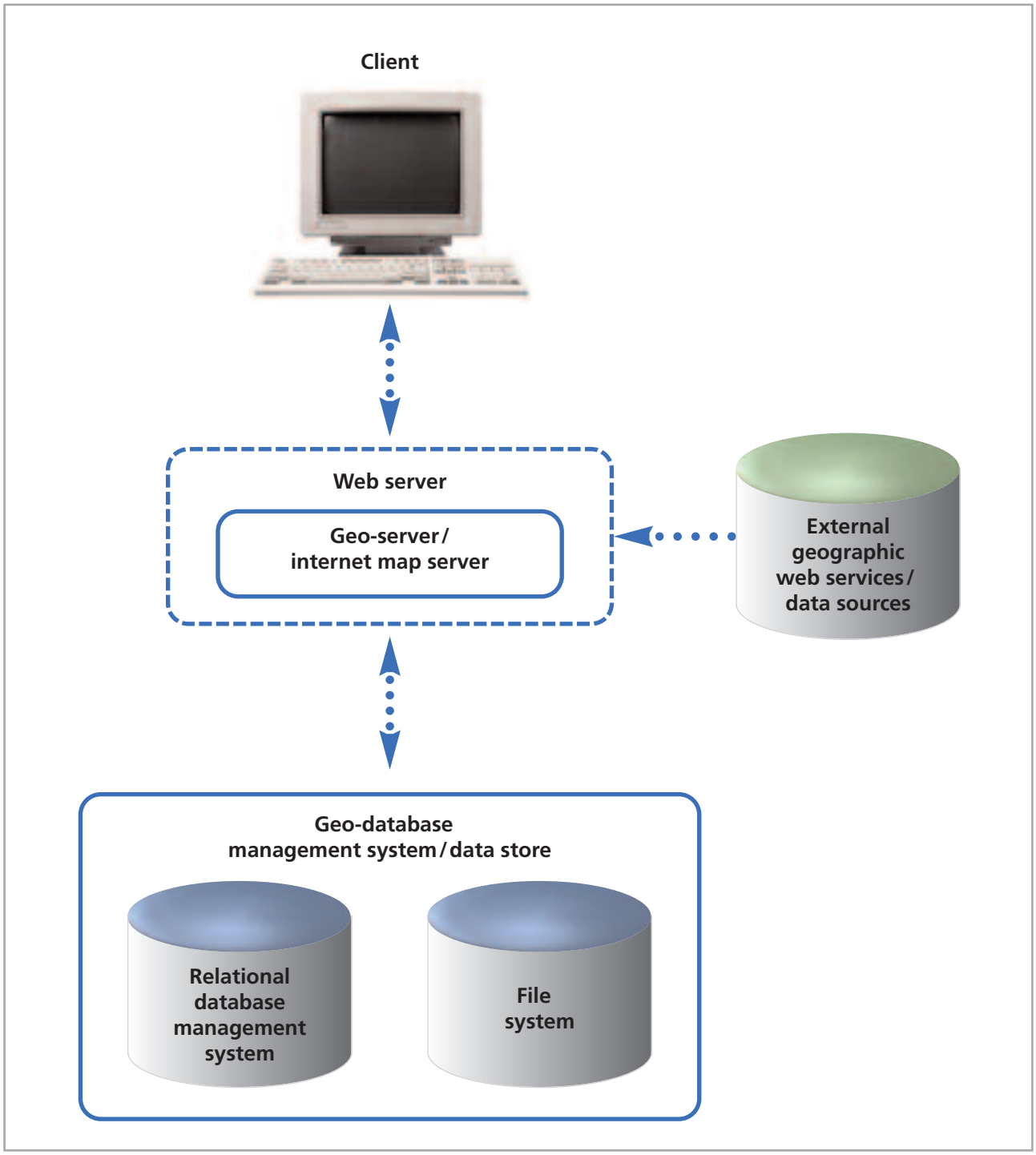

Figure 1: Simplified overview of geo-portal software and data architecture

regard is the ongoing effort to establish an EU-wide SDI or Infrastructure for Spatial Data in Europe (INSPIRE) (EU, 2009). A key component of any SDI is the metadata database/catalogue. The provision of a comprehensive metadata catalogue based on robust metadata standards is crucial for ensuring effective data management and optimising metadata discovery (using spatial, temporal or thematic queries) via online geo-portals. In addition to metadata and GI data discovery, geo-portals may offer users more advanced services such as GI data viewing, downloading, editing and processing (for example, spatial analysis and GI data transformation).

Geo-portals generally fall into one of two groups: 'national' or 'regional' geo-portals and 'domain' or 'sector-specific' geo-portals (Beaumont *et al*, 2005; Tang and Selwood, 2005). Examples of national and regional geo-portals, designed to allow varying levels of access to a broad range of GI data, include examples from the USA (the Geospatial One-Stop

website[1]), Canada (GeoConnections[2]), Norway (geoNorge[3]) and The Netherlands (the Nationaal Register). Examples of domain-specific geo-portals include the UK Government's Multi-Agency Geographic Information for the Countryside (MAGIC) environmental geo-portal[5], the British Geological Survey's GeoIndex[6], natural disasters geo-portals such as Pacific Disaster Center[7] and the Great Britain Historical GIS[8]. The MAGIC geo-portal, for example, is specifically designed to provide data related to the countryside in England through a partnership of environmental agencies (Askew *et al*, 2005). As well as the technical challenges involved in creating such a geo-portal, this has also drawn attention to a whole host of institutional problems related to factors such as data security and standards, licensing arrangements and the need for inter-agency agreements. This paper is primarily concerned with some of the technical challenges related to formats, standards, and the need for metadata.

Maguire and Longley (2005) proposed that geo-portals may also be divided into groups based on the nature of their service provision. Whilst 'catalog' geo-portals such as the United States Geological Survey's Earth Explorer[9] concentrate on providing access to GI, 'application' geo-portals offer more dynamic GI services. Examples of application geo-portals include online transport portals that incorporate sophisticated route planning components, such as Journey Planner[10] (Transport for London, 2009). Furthermore, geo-portals can also be subdivided into those which are 'commercial' and 'non-commercial'.

A burgeoning GI industry is currently supporting a large number of e-commerce-based geo-portals which allow users to discover, buy and download spatial data (for example, the Cities Revealed geo-portal[11]). The software and data components of such tools are described in the next section.

Software and Data Components

Figure 1 shows a simplified overview of software and data architecture of a typical geo-portal which has four main components. The geo-spatially enabled relational database management system is commonly used to store, manage and integrate an organisation's GI data archive, although in some cases file system storage may be used, particularly for the management of raster data (for example, satellite images).

The website element of a geo-portal comprises two main software components: a web server, which provides the supporting framework for publishing content to the Web; and geographic software tools—Internet Map Server (IMS)/geo-server—that publish GI contained in the geo-database and facilitate access to GIS functionality (for example, map rendering, querying, editing, spatial analysis, data transformation, geo-coding and metadata management) via geographic Web services (GWSs).

The use of open and international standards, such as International Standards Organisation (ISO) and Open Geospatial Consortium (OGC) for creating and deploying GWS, is designed to increase the interoperability and sharing of GI data between individuals and organisations. Therefore, as shown in Figure 1, geo-portals can link to external sources of GI and GIS services provided by other individuals or organisations in the form of GWSs, in addition to publishing 'in-house' geographic content on the Web for consumption by external users. The fourth and final component of the geo-portal architecture is the client. The role of the client is to communicate with the geo-server and render map data in a user's Web browser. Clients can be divided into two main groups: 'thick' clients and 'thin' clients. A thick client performs much of the data processing operations independently of the geo-server. This could be through the use of a Java-based web browser plug-in, for example. A thin client, however, simply renders map graphics while the geo-server conducts the main data processing. Client side scripting languages such as JavaScript/AJAX are commonly used when developing thin clients, which are typically embedded in the web browser.

'Going Open Source'

ACCORDING to the Open Source Initiative (OSI, 2009), open source:

> *...is a development method for software that harnesses the power of distributed peer review and transparency of process. The promise of open source is better quality, higher reliability, more flexibility, lower cost, and an end to predatory vendor lock-in.*
>
> Open Source Initiative, 2009

1. Geospatial One-Stop: http://gos2.geodata.gov
2. GeoConnections: http://www.cgdi.ca
3. geoNorge: http://www.geonorge.no
4. Nationaal Register: http://www.nationaalgeoregister.nl
5. MAGIC environmental geo-portal: http://www.magic.gov.uk
6. GeoIndex: http://www.bgs.ac.uk/GeoIndex
7. Pacific Disaster Center: http://www.pdc.org
8. Great Britain Historical GIS: http://www.visionofbritain.org.uk
9. Earth Explorer: http://edcsns17.cr.usgs.gov/EarthExplorer
10. Journey Planner: http://journeyplanner.tfl.gov.uk
11. Cities Revealed geo-portal: http://www.citiesrevealed.com

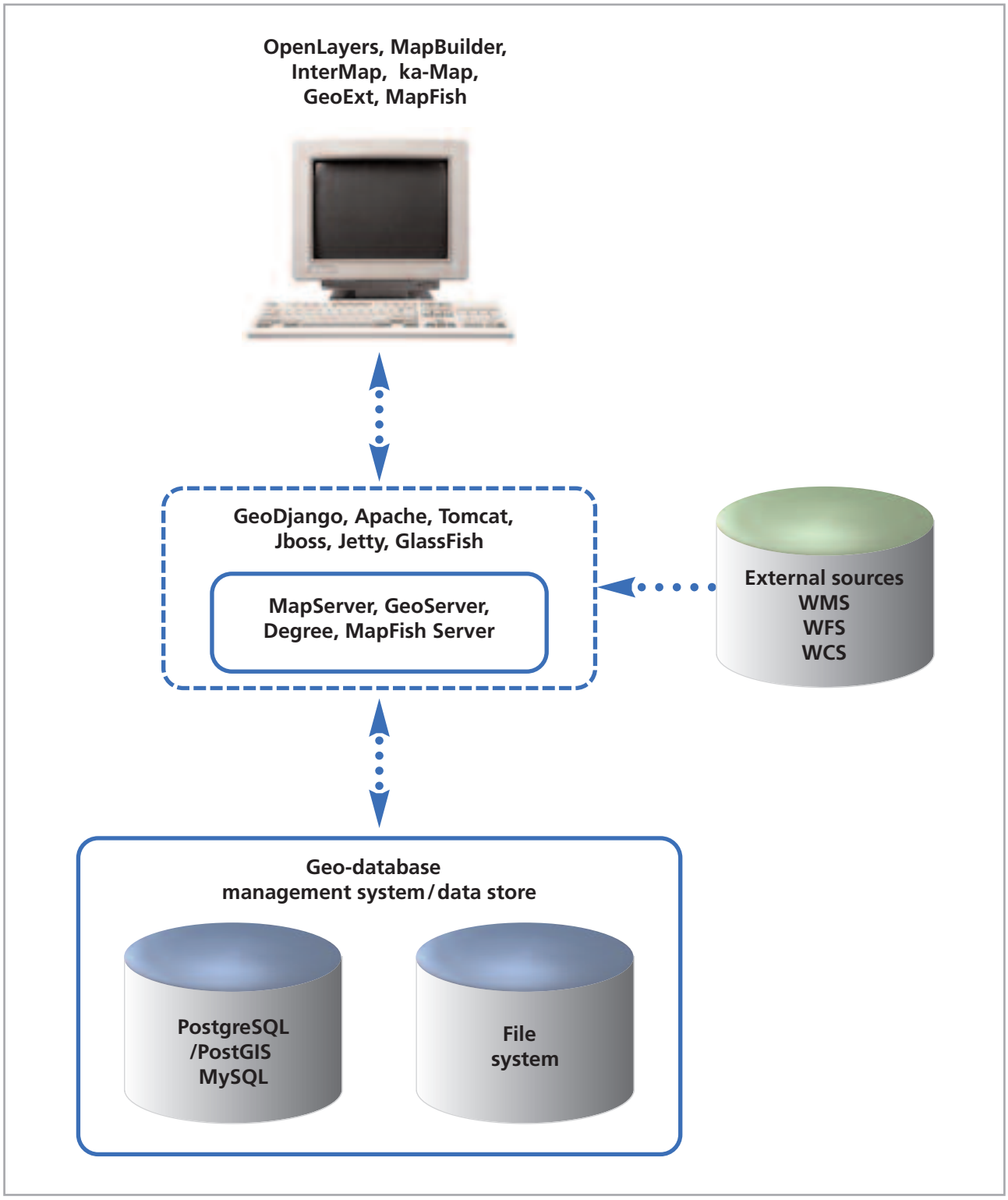

Figure 2: Geo-portal architecture showing selected GeoFOSS software components

Whereas the source code of proprietary software programmes is closed, the coding of open source software is open to modification and freely distributable. Thus, the term free and open source software (Foss) is commonly used when referring to software of this type. Within the geo-spatial realm, there is a steadily growing interest in Foss, reflected in a high-profile international annual conference series—Free and Open Source Software for Geospatial (Foss4G) supported by increasingly prominent organisations such as the Open Source Geospatial Foundation (OSGeo) and the Open Geospatial Consortium (OGC). A recent announcement by the UK government outlining an action plan to promote the

use of Foss at all levels of government, is set to add extra impetus and credibility to the open source movement in Britain (Cioc, 2009).

Building on our review of Foss, we aim to develop a research geo-portal using open source geo-spatial software (GeoFoss). There are a number of potential advantages of GeoFoss. Firstly, building on the expertise in Web development and programming within the Wiserd integration team, the use of software with an open code base permits a high level of flexibility and customisation during development. Proprietary software, as well as being beyond the reach of resource-constrained agencies, is generally much less flexible in this regard and users often have to wait for the vendor to publish software patches and updates to resolve any issues with the source code. With open source technologies, the open nature of the code, the sheer number of developers and the fact that such technologies continue to evolve at a rapid pace, means that software bugs and coding problems within GeoFoss projects, that involve the use of web-based information systems for collecting, storing and analysis of socio-economic or environmental data, are often resolved much more quickly than those that utilise proprietary software.

Another potential advantage of using Foss, particularly on long-term projects, is the freedom to change and share any of the software components at any stage. If new software is developed that is more suited to a particular task, or a certain program is rendered obsolete, using Foss means that there is no pressure from software vendor contracts or financial concerns surrounding licensing that may restrict any decision-making. Proprietary software vendors may also choose to stop supporting and developing older versions of their software at some point. With Foss, there is no single entity on which the future of the software is dependent. Should a particular group should stop development of the software, then anyone can continue the maintenance and improvement of the source code.

An undeniable attraction of 'going open' was the potential financial savings that could be made by using Foss to develop the Wiserd geo-portal. Thus, a considerable reduction in costs could be made in the long term by using Foss over proprietary Internet mapping and relational database management software solutions through minimal development and support costs. From a research point of view, the current groundswell of interest in GeoFoss within the academic community was another important factor in deciding to opt for open source. The Wiserd team felt that there was a real opportunity to contribute to the knowledge in the area by developing a novel type of geo-portal based on the innovative use of GeoFoss components. However, there are those that view Foss with negativity and suspicion. There is a perhaps understandable opinion amongst some that programs containing source code with a free and open licence permitting unrestricted modification, distribution and usage, are simply 'too good to be true' and/or not in the spirit of commercial endeavor. For financial controllers and information technology (IT) managers of organisations used to routinely budgeting for proprietary software licence fees, this may be disconcerting. Even if a good awareness of Foss exists within an organisation, it may be the case that the right Foss tool for a particular task does not exist, or is insufficiently powerful or robust. Generally, Foss are less mature than their proprietary equivalents and may often not be the best choice, particularly in organisations that lack the expertise to realise the potential advantages of 'going open'.

Another concern voiced by some, relates to the security of open source software. There are those that argue that software with open code is exposed to attack by viruses and other malicious code, but the counter-argument from the open source community is that any irregularities in the code are picked up very quickly by the developers, and do not represent a danger to the future releases/updates of a particular program. In view of the various potential benefits and drawbacks highlighted previously, we do not propose to adopt a GeoFoss versus proprietary software stance and regard the different approaches as complimentary and not confrontational paradigms. For this particular project, the use of GeoFoss components for constructing the geo-portal offers the best solution for flexible, bespoke development and also has a greater merit in terms of academic research priorities.

There are already a large number of GeoFoss components available for building Internet mapping sites and geo-portals. Figure 2 shows some of the more popular GeoFoss packages in relation to the geo-portal architecture described previously although this is by no means an exhaustive list. Researchers and commercial developers are now able to build geo-portals combining advanced GIS functionality with slick user interfaces using various combinations of such GeoFoss components. One current example is the MarineMap geo-portal being developed by a consortium led by the University of California, Santa Barbara (MarineMap, 2009). MarineMap (Figure 3) utilises a blend of Post-GIS, GeoServer, Geo-Django, GeoExt and OpenLayers software to produce an online decision support tool for marine spatial planning that probably represents the state-of-the-art in GeoFoss-based geo-portal development. Some initial thoughts on choices of GeoFoss components for the Wiserd geo-portal are discussed later in the paper.

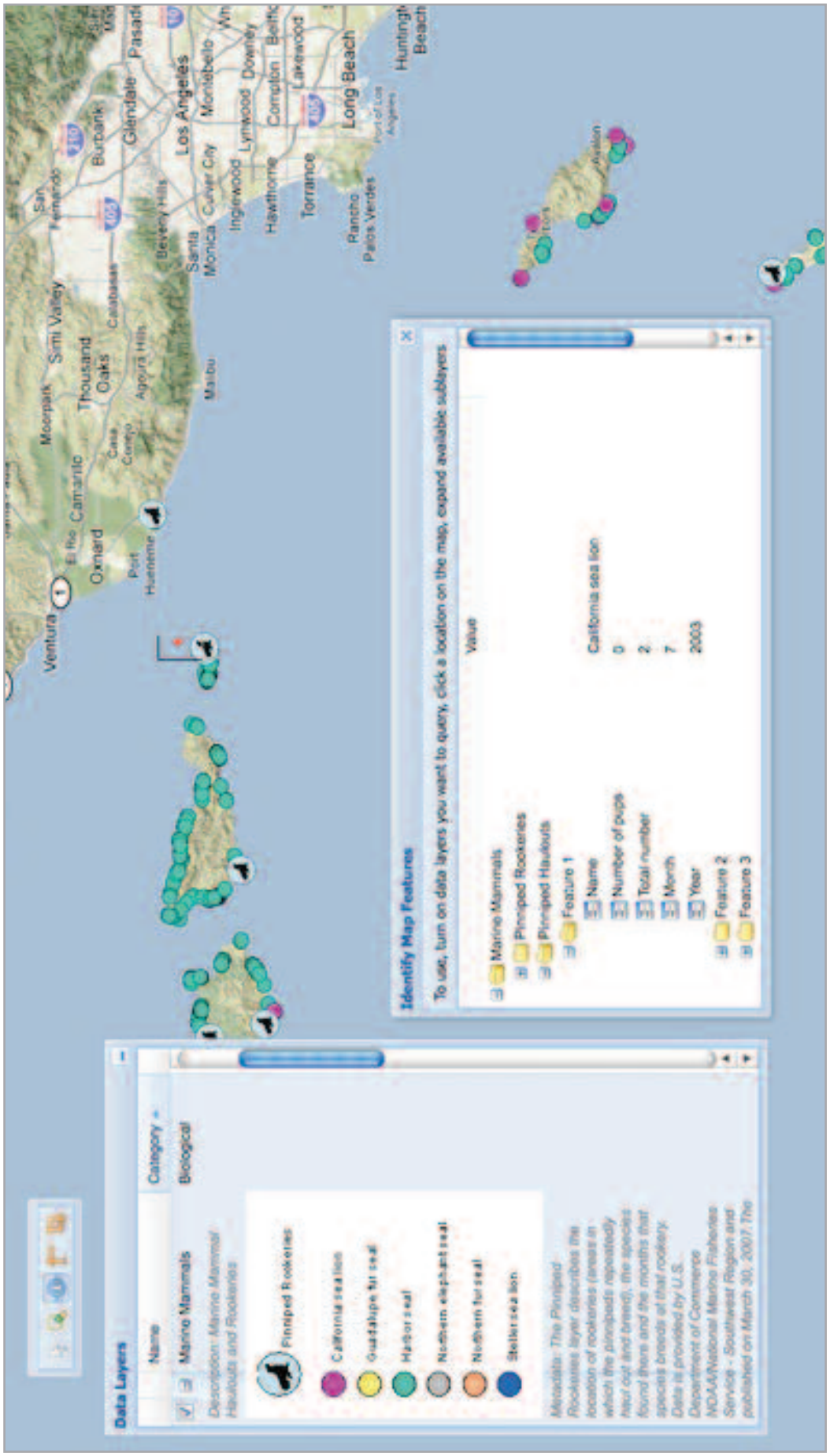

Figure 3: MarineMap: Example screen shot from the state-of-the-art GeoFOSS-based geo-portal development (http://marinemap.org/marinemapl)

The team's open source approach to the development of the WISERD geo-portal is not only restricted to software. Topographic base mapping data will be an essential component of the geo-portal for allowing users to find, visualise and analyse data and metadata via a dynamic cartographic web interface. However, as base mapping is required for the whole of Wales, the licensing costs of using commercially-available data such as Ordnance Survey (OS) will likely prove prohibitively expensive. Even if a research licence for OS data was to be negotiated, WISERD envisages making elements of the geo-portal available to external non-academic users at some point in the future. The use of freely-available open source mapping data in the form of OpenStreetMap (OSM) (OSMF, 2009) could offer a possible solution to these problems. OpenStreetMap data is crowd-sourced or volunteered geographic information (VGI) meaning that anyone can contribute to its creation and anyone can download, use and distribute it under an open licence agreement.

In addition to the initial cost savings over commercial map data, there are also no restrictions in place preventing the creation of secondary data (for example, spatial data created by digitising a base map) from OSM, as there would be with some commercial mapping products. Another advantage of using OSM is that unlike Internet-mapping APIs, such as Google Maps (Google, 2009), that can be used to embed free dynamic mapping in a web page, OSM data can be downloaded, meaning that developers are not reliant on external map servers for displaying their mapping. A major drawback of OSM however, is that despite its rapid expansion and increasing popularity, coverage at finer scales in the UK is currently far from complete. Whilst many of the centres of major towns and cities have been mapped in detail, smaller town and rural areas are less well charted. Haklay (2009) speculates that at present rates, detailed mapping for the whole of the UK could be completed within three years. In an investigation into the quality of OSM, Haklay (2009) found that the positional accuracy of OSM was adequate (within 6 metres of the position recorded by the OS) for many mapping requirements, but not on a par with national mapping standards.

In order to address the current deficit of detailed OSM coverage in some of the proposed WISERD 'localities', particularly in rural study areas, the data integration team plan to hold a series of mapping exercises or 'mapping parties' where WISERD researchers from different backgrounds can join together and produce their own OSM data. Besides being of direct benefit to the WISERD programme, both in terms of map resource generation and knowledge capacity-building within the WISERD group, such exercises will expand the OSM coverage for everyone and help to support the open source community in general. These exercises may also benefit the communities/'localities' being mapped in the long term.

The WISERD geo-portal

The WISERD database

The main objective of the WISERD data integration team is to acquire, integrate and make more usable (via a geo-portal), the wide range of existing quantitative and qualitative data that relate to Wales. This is designed to provide WISERD researchers with the necessary information resources with which to conduct innovative world-class research and inform government policy debates. This is an important element of the WISERD programme and is the hub around which all other activities will develop. The WISERD database will be complex and diverse, consisting of quantitative, qualitative, spatial and non-spatial data in a variety of formats and media. The sources of data will also be varied and include primary data collected as part of the WISERD 'localities' research, secondary or official data (primarily quantitative), collected by the UK Government/WAG, the ONS and the private sector, and grey data (published and unpublished routinely collected administrative data on individuals). Examples of these datasets (many of which WISERD will not be holding directly) are shown in Table 1.

In addition, a unique aspect of the database will be the inclusion of metadata and data related to the specific questions used in each of the surveys (Table 1). It is hoped that users accessing the database via the WISERD geo-portal will be able to use this resource to help them identify similar themes across the surveys, discover new data, and visualise gaps (spatial and thematic) in the existing data. In order to meet the main objective of this part of the WISERD research programme, the WISERD data integration team will tackle four distinct strands of research: (1) metadatabase construction; (2) quantitative data integration; (3) qualitative data integration; and (4) quantitative-qualitative data integration. These strands are discussed briefly in the following section.

The WISERD geo-portal

Figure 4 shows a non-technical overview of the WISERD geo-portal which places an emphasis on the spatial framework that will be used to integrate, manage and disseminate the wide range of quantitative and multi-modal quantitative data that will form the key resource for WISERD researchers. When the geo-portal first becomes operational, researchers will have access to two main user functions, namely metadata discovery and data download. The geo-portal will allow users to perform spatial, temporal and thematic queries to search for metadata using both cartographic (i.e. dynamic web map) and text-based interfaces (Figure 4). Further advanced cartographic, editing, analysis and data management functions may also be incorporated, based on a user-demand for such

Data Source	Data Type	Example data
PRIMARY	**Quantitative**	Surveys — household and business
	Qualitative	Interview transcripts
		Field notes
		Audio
		Still images
		Video
SECONDARY	**Quantitative**	UK Household Longitudinal Study
		Census of Population
		Living in Wales
		British Household Panel Survey
		British Social Attitudes Survey
		British Election Study
		Wales Health Survey
		British Crime Survey
		Labour Force Survey
		Workplace employee relations survey
		New Earnings Survey
		ONS Model-income estimates
	Qualitative	Existing archived data
'GREY'	**Published**	Electoral register
		Reported crime statistics
	Unpublished	Health records
		School records

Table 1 The WISERD database

services. In addition to helping users discover what data exists and where they can find it, an important function of the geo-portal will be the visualisation of data 'gaps' (particularly official data) in Wales so that researchers can help to address areas of 'data deficit' in future. In doing so, there is an opportunity to develop novel methods of spatial data geo-visualisation in order to identify and illustrate such gaps. Where feasible, the geo-portal will also offer a data download service. In some cases, particularly where official secondary data sources are concerned, data download may not be possible due to data access/licensing restrictions. However, at the very least, users will be supplied with the metadata containing information on where such data is held and how it can be accessed. Also, hyperlinks to external data sources will be provided in the metadata where relevant.

The full range of GeoFoss components that will be used to build the Wiserd geo-portal have yet to be decided, although PostGIS will almost certainly be handling the database duties, while GeoServer is currently the favoured software for developing the IMS. Also, there are a number of possibilities for developing client-side user mapping interfaces using a variety of JavaScript-based libraries such as GeoExt, OpenLayers and MapFish etc. The development of the geo-portal therefore presents a good opportunity to evaluate different GeoFoss components for building Internet mapping resources, research which is currently lacking in the published literature. Other 'spin-off' projects that may emerge during the main geo-portal development phase could, for example, include research into the usability of Web-based mapping/geo-portal interfaces, another area which is under-researched at present.

Metadata

Metadata is the building block of the Wiserd geo-portal. It underpins the data integration, management and dissemination research, and forms the backbone of what will effectively become the Wiserd SDI. Metadata is central to the functionality of geo-portals for facilitating efficient data discovery, data evaluation and data usage (AGI, 2009). Effective metadata will be essential for maximising the value of database searches and ensuring that geo-portal users are able to

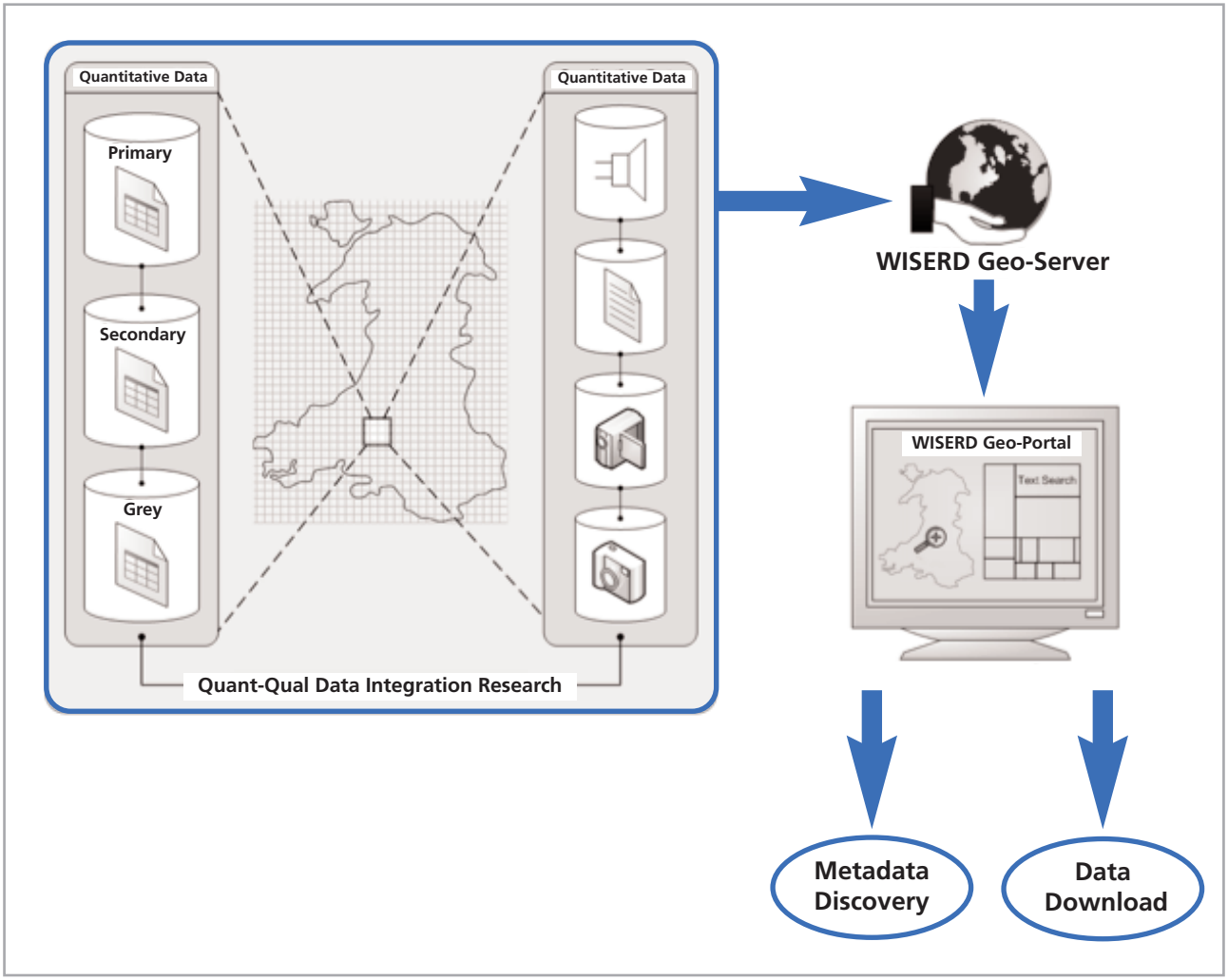

Figure 4: Non-technical overview of the WISERD Geo-portal

find datasets that are potentially able to satisfy their sets of specific requirements. Also, comprehensive metadata will allow researchers to fully evaluate a data resource in order to ascertain whether it is fit for its intended purpose and will provide the information necessary to configure a system or software in order to process it.

Another important application of metadata is in facilitating the discovery and interoperability of distributed datasets which may be an important consideration should the WISERD geo-portal be used to publish selected metadata, data or GWSs to the Internet for consumption by external parties at some point in the future. Conversely, a standards-based approach to metadata will be crucial for allowing users to search for external distributed data resources via the WISERD geo-portal search interfaces. In order to meet these requirements, a robust metadata standard, which should be compliant with the EU INSPIRE spatial metadata standard to ensure future-proofing of the geo-portal and its data resources and services, must be strictly implemented throughout the course of the project.

There are a plethora of metadata standards used worldwide for defining the content, quality, source and condition of spatial data (Moellering *et al*, 2005). At the core of many world-wide geo-spatial metadata standards is the International Organisation for Standardisation (ISO) standard 19115, adopted in 2003 (ISO, 2003). Many national standards are derivations of ISO 19115, with a minimum requirement that each nation retain at least thirteen core metadata elements in order to maintain ISO compliance. In the UK, the GEMINI Version 1.0 standard was produced in 2004 to provide a "defined element set for describing geo-spatial, discovery level metadata" (AGI, 2004, p3). Designed to supersede the GI Gateway Metadata Specification (previously the National Geo-spatial Data Framework), GEMINI Version 1.0 included the core metadata elements of eGovernment Metadata Standard (eGMS) and ISO 19115 and optional elements of those standards deemed necessary as a result of the original GEMINI consultation process (AGI, 2004). In 2006, the standard was revised by a working group of UK organisations operating under the auspices of the Association of Geographic Information (AGI). In addition to correcting the errors in GEMINI 1.0 and ensuring ongoing compatibility with eGMS, the main purpose of the revision was to meet the requirements for spatial metadata of the EU INSPIRE directive whilst maintaining conformance with ISO 19115 (AGI, 2009; Walker and Rackham, 2006). Data resources included in the WISERD geo-portal will likely be based on an extended version of this standard (GEMINI 2.3).

A limitation of current spatial metadata standards however, is that: "...they lack fields to express information beyond the technical and geometric domain" (Schuurman and Leszczynski, 2006, p719). Thus, whilst metadata and GWS standards are designed to facilitate greater data interoperability at the system, syntactic and structural level, they are less well equipped at addressing issues of semantic interoperability. Therefore, according to Athnasis *et al* (2009):

> *...usability of information created in one context is often of limited use in another context, due to insufficient means for meaningful interpretation, a problem widely known as semantic heterogeneity.*
>
> Athanasis *et al*, 2009, p302

General metadata standards for geo-spatial data fail to capture the epistemological and ontological contexts relating to the origin, nature and intended purpose of the data in question. Without a deeper understanding of data context there is a danger that semantically heterogeneous data will be misinterpreted and invalidly integrated (Schuurman and Leszczynski, 2006). Also, metadata discovery is mainly based on searching according to spatial, thematic or temporal keywords and accordin to Athanasis and co-workers:

> *... users [of geo-portals] often do not find what they are looking for, because querying in the portal catalogue depends on the specific elements of the geo-spatial metadata standards.*
>
> Athanasis *et al*, 2009, p302

Consequently, another important area of research during the development of the WISERD geo-portal will be to investigate means of addressing semantic heterogeneity issues and ways of improving the efficiency of metadata catalogue querying/searching. This work will aim to evaluate and extend ongoing research into ontology-based metadata, semantics, and novel database paradigms, such as the Resource Description Framework (RDF), for tackling such problems (see for example Schuurman and Leszczynski, 2006; Athanasis *et al*, 2009).

Discussion

THE AIMS of this paper have been to describe the overarching aims of the WISERD initiative, to describe the principal objectives of the work of the Data Integration theme within WISERD and to focus on one aspect of the delivery of such tools, namely that related to the development of a geo-portal. Askew *et al* (2005, p85), in an evaluation of one of the most widely used geo-portals in the UK (the MAGIC project), suggest that there is "the need to understand the requirement to move from developing a geoportal from a data provider's perspective to developing facilities from a data user's perspective". This is very much the approach we have adopted

in the early stages of the WISERD project with close liaison being encouraged with the other teams involved within WISERD (such as the policy and qualitative themes) as well as from meetings/correspondence with the main data user's within Wales. The geo-portal is currently in its preliminary planning stages, but will ultimately form a key element of the WISERD research programme providing a valuable tool for social scientists in Wales and beyond. It will in turn build on the strengths and weaknesses of existing frameworks such as the Geospatial One-Stop sponsored by the US Federal Government (Goodchild *et al*, 2007) and from European initiatives such as the INSPIRE project. One area where spatial web portals are perceived to be limited, for example, is in terms of more powerful analytical tools that go beyond data searching and map visualisation; although Zhang and Tsou (2009) highlight some important exceptions such as the prototype developed by Nyerges *et al* (2006) and Yang *et al* (2007). This could form the basis for follow-up research to extend the capabilities of the geoportal depending on user demand and feedback from the beta version.

The project is ambitious, bringing together several interrelated strands of innovative research related to data integration (quantitative, qualitative and quantitative-qualitative; spatial and non-spatial), geo-visualisation, GeoFOSS evaluation and web development, interface usability research, SDI and metadata development and ontological and semantic approaches to geo-portal database development. A prototype version of the geoportal is scheduled to come online by December 2011, but research, development and evaluation will continue thereafter. In terms of data integration, the initial research focus has been on quantitative data integration, particularly of official secondary data in Wales, while investigations into qualitative GIS and qualitative-quantitative data integration will be conducted during the more advanced stages of the project. The immediate concern, however, will be the creation of a WISERD SDI which will underpin the development of the WISERD geo-portal. Production of the SDI will encompass the development a WISERD database, metadatabase and metadata standard, and the formulation of data policies and data management strategies. The data integration team anticipates that a successfully implemented WISERD SDI may be used as a model for future multi-disciplinary collaborative research projects. Similarly, through the creation of the WISERD geo-portal, the team has an opportunity to develop a model Web-based GIS framework for supporting large-scale multi-disciplinary academic research programmes. The project, involving five universities within Wales, is therefore in a unique position to draw on expertise across the social sciences in Wales, to address some of the crucial gaps highlighted in the findings of the Rhind Commission and to build on such research efforts through capacity building initiatives in mixed methods approaches to address some of the key social issues concerning Wales.

References

AGI (2004) UK GEMINI standard version 1.0: A geo-spatial metadata interoperability initiative. London, The Cabinet Office. Available online at: http://www.gigateway.org.uk/metadata/pdf/UK_GEMINI_v1.pdf (accessed 1 June 2009).

AGI (2009) UK GEMINI standard: A UK metadata standard for discovery of geographic data resources. Available online at: http://www.gigateway.org.uk/metadata/pdf/GEMINIV2_3c.pdf. (accessed 1 June 2009).

Askew D, Evans S, Matthews R & Swanton P (2005) MAGIC: A geoportal for the English countryside. *Computers, Environment and Urban Systems* **29**, 71-85.

Athanasis N, Kalabokidis K, Vaitis M & Soulakellis N (2009) Towards a semantics-based approach in the development of geographic portals. *Computers and Geosciences* **35**(2), 301-308.

Beaumont P, Longley PA & Maguire DJ (2005) Geographic information portals—a UK perspective. *Computers, Environment and Urban Systems* **29**, 49-69.

British Geological Survey (2010) Geoindex. Online at: http://www.bgs.ac.uk/GeoIndex (accessed 15 January 2010).

CIOC (2009) Open source, open standards and re-use: Government action plan. London, The Cabinet Office. Available online at: http://www.cabinetoffice.gov.uk/media/123372/090224opensource.pdf (accessed 15 January 2010).

Cope S & Elwood S (2009) *Qualitative GIS: A mixed methods approach*. London, Sage.

CSS (2003) Great expectations: the social sciences in Britain. London, Commission on the Social Sciences.

Economic and Social Research Council (2009) Welcome to ESRC: society today. Online at: http://www.esrc.ac.uk (accessed 3 June 2009).

EU (2009) INSPIRE directive. Online at: http://inspire.jrc.ec.europa.eu/index.cfm (accessed 30 June 2009).

GeoConnections. Online at http://www.cgdi.ca (accessed 15 January 2010).

GeoNorge (2007). Online at http://www.geonorge.no (accessed 15 January 2010).

Geospatial One-Stop. Online at: http://gos2.geodata.gov (accessed 15 January 2010).

Goodchild MF, Fu P & Rich P (2007) Sharing geographic information: An assessment of the geospatial one-stop. *Annals of the Association of American Geographers* **97**(2), 250-266.

Google (2009) Google Maps API. Online at: http://code.google.com/apis/maps/ (accessed 6 July 2009).

Great Britain Historical GIS Project (2009) A vision of Britain through time. Online at: http://www.visionofbritain.org.uk (accessed 30 June 2009).

GSDI (2009) Global spatial data infrastructures association cookbook index. Online at: http://www.gsdi.org/gsdicookbookindex.php (accessed 30 June).

Haklay M (2009) Summary of OpenStreetMap quality assessment paper (currently under Review in Environment and Planning B). *Centre for Advanced Spatial Analysis S4 Event*, 8 January, London, UCL.

ISO (2003) *ISO Standard 19115: Geographic information - metadata*. International Organization for Standardization. Available online at: http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=26020 (accessed 1 June 2009).

Maguire D & Longley P (2005) The emergence of geoportals and their role in spatial data infrastructures. *Computers, Environment and Urban Systems* **29**(1), 3-14.

MarineMap (2009) MarineMap: Decision support for marine spatial planning. Online at: http://www.marinemap.org/ (accessed 3 July 2009).

Moellering H, Aalders H & Crane A (2005) *World spatial metadata standards*. Oxford, Elsevier.

Moore L, Lynch R, Maio G, Moore G, Orford S, Robinson A, Taylor C & Whitfield K (2007) Social science quantitative methods capacity building in Wales: ESRC/HEFCW scoping study. *Cardiff University School of Social Sciences Working Paper 91*.

MAGIC. Online at http://www.magic.gov.uk (accessed 15 January 2010).

Nationaal Georegister.nl. Online at: http://www.nationaalgeoregister.nl/ (accessed 15 January 2010).

Nyerges T, Ramsay K & Wilson M (2006) Design considerations for an Internet portal to support public participation in transportation improvement decision-making. In: Balram S and Dragicevic S (eds.) *Collaborative geographic information systems*. Hershey, Idea Publishing.

OSI (2009) Open Source Initiative (OSI). Online at: http://www.opensource.org/ (accessed 2 July 2009).

OSMF (2009) OpenStreetMap. Online at: http://www.openstreetmap.org/ (accessed 6 July).

Pacific Disaster Center geoportal. Online at http://www.pdc.org (accessed 15 January 2010).

Schuurman N & Leszczynski A (2006) Ontology-based metadata. *Transactions in GIS* **10**(5), 709-726.

Tait M (2005) Implementing geoportals: Applications of distributed GIS. *Computers, Environment and Urban Systems* **29**(1), 33-47.

Tang W & Selwood J (2005) *Spatial portals: Gateways to geographic information*. Redlands, ESRI Press.

Transport for London (2009) Journey Planner. Online at: http://journeyplanner.tfl.gov.uk (accessed 15 January 2009).

United States Geological Survey (2009) Earth Explorer. Online at: http://edcsns17.cr.usgs.gov/EarthExplorer/ (accessed 15 January 2010).

Walker R & Rackham L (2006) Metadata guidelines for geospatial datasets in the UK (part 1). Department for Communities and Local Government. Available online at: http://www.gigateway.org.uk/pdf/guidelines/MetadataGuidelines1.pdf (accessed 16 June 2009).

Yang P, Evans J, Cole M, Alameh N, Marley S & Bambacus M (2007) The emerging concepts and applications of the spatial web portal. *Photogrammetric Engineering and Remote Sensing* **73**, 691-698.

Zhang T & Tsou M-H (2009) Developing a grid-enabled spatial Web portal for Internet GIServices and geospatial cyberinfrastructure. *International Journal of Geographical Information Science* **23**(5), 605-630.