# The WISERD Geoportal: An innovative tool for discovering social science research data

Berry, R. [1], Fry, R. [1], Higgs, G. [1], Orford, S. [2], Jones, S. [2]

[1]Wales Institute for Social and Economic Research, Data and Methods (WISERD), GIS Research Centre, University of Glamorgan, Pontypridd, Wales, UK, CF37 1DL
Tel. +44 (0) 1443 483377 Fax +44 (0) 1443 483715
raberry@glam.ac.uk, www.wiserd.ac.uk
[2] Wales Institute for Social and Economic Research, Data and Methods (WISERD), Cardiff University, 46 Park Place, Cardiff, Wales, UK, CF10 3BB

**Summary:** This paper describes the development of the WISERD Geoportal (WGP) – an innovative web-based GIS for improving the discovery of social science research data in Wales. The research focuses on the key methods used in the building of the WGP, including the creation of a rich database containing metadata for socio-economic data in Wales and the development of the application using an open source software stack. In the conference presentation we will offer our thoughts on the future developments of the WGP, its ongoing sustainability and its potential importance in the social science research landscape.

**KEYWORDS:** Geoportal; metadata; GIS; open source; social science

## 1. Introduction and Background

The Wales Institute for Social and Economic Research Data and Methods (WISERD) is an interdisciplinary, cross-institutional academic research group based in Wales, UK. WISERD was formed in 2008 with aim of boosting the strength of social science research in Wales by creating a large group of researchers based at the five main Welsh universities (Aberystwyth, Bangor, Cardiff, Glamorgan and Swansea). The aim of the WISERD is to draw upon collaborative inter-institutional working and wide-ranging expertise in order to improve the sustainability of social science research in Wales and produce major innovations in research methods.

One of WISERD's key research objectives is to develop innovative technological approaches to facilitate the discovery, visualisation and analysis of social science data. WISERD recognises that, while there are a large and diverse number of datasets in Wales (and the UK) that are potentially available to social scientists, economists and analysts, there are few effective mechanisms for discovering these data (UK Data Archive excepted). As a result, potentially useful data that has been collected are often under-used or not used at all, for research and analysis. Given the current economic climate, it is important that the value of data held in various repositories throughout Wales is optimised, in line with current UK Government thinking on the reuse and repurposing of existing data sources (National Data Strategy, 2009). In response to this, WISERD are developing a web portal in order to help researchers and analysts discover socio-economic research data more efficiently and, in doing so, encourage the re-use of existing data and help stimulate collaborative research between researchers and organisations.

The web portal has been (temporarily at least) assigned the name 'WISERD Geoportal' (WGP), since much of the metadata contained within its database is geographically referenced and is accessed/queried via GIS-style cartographic user interface. While there is no universal agreement on the precise definition of a *geoportal* (unsurprising given the variation in scope and scale of online mapping/GIS applications that have been labelled with the term), many tend to agree with the principle that a geoportal is some kind of web-based application used to discover and/or access

geographic data and geographic information services (see for example Armstrong et al., 2011; Beaumont et al., 2005; Goodchild at el., 2007; Maguire and Longley, 2005; Tait, 2005; Tang and Selwood, 2005). Many geoportals are effectively front ends to spatial data infrastructures (SDIs) permitting various degrees of access to data at local, regional and national levels or in the case of the European Union's INSPIRE geoportal (http://www.inspire-geoportal.eu/) at the pan-national level (see Masser (2011) for a recent overview). In addition, some geoportals have been developed with narrower thematic remits in mind; for example for the management of subsurface data in the Netherlands (Lance et al., 2011) or to access data on risk to natural hazards at the global scale (Giuliani and Peduzzi, 2011).

Based on an on-going review of worldwide geoportals, we believe that the WGP is innovative in that it focuses on the discovery of socio-economic research data, supporting both quantitative and qualitative social science research and analysis. Another noteworthy aspect of the WGP is that it has been produced almost exclusively using Free and Open Source (FOSS) software in favour of proprietary software, mainly due to the flexibility that FOSS affords in terms of customisation and flexibility during development. Also, FOSS is currently a 'hot' topic in GIS research internationally and this, coupled with the current UK Government drive to promote the use of open source software at all levels of government (CIOC, 2009), means that there was is opportunity to add to the knowledge in this research area and contribute a novel FOSS4G (Free and Open Source Software for Geospatial) 'case study' application.

This paper concentrates on the technical development of the development of the WGP, building upon an earlier paper that provided a conceptual background and a more detailed discussion of its place within WISERD and the wider social science research landscape (Berry et al., 2009). The remainder of this paper focuses on the key methods used in the building of the WGP, including the creation of the 'WISERD Metadatabase' (WMDB) which underpins the WGP and the open source 'software stack' used to develop the application. We then outline further proposed developments of the WGP before offering our conclusions.


## 2. Methodology

### 2.1 Database

The WMDB contains detailed metadata for a diverse range of quantitative and qualitative socio-economic research data related to Wales including government surveys (down to the question and response level), primary WISERD interview transcripts, 'grey' data (e.g. academic publications) and administrative data (e.g. school-level education data). By using a spatially-enabled relational database (PostGIS-PostgreSQL), metadata records, where possible, have been linked to real-world locations allowing the database to be searched using GIS-based spatial search tools. In order to ensure quality and consistency the database has been designed so that metadata conform to both international and national non-spatial and spatial metadata standards respectively (i.e. Dublin Core, DDI 3.0, INSPIRE/GEMINI 2.1). A general overview of the contents of the WMDB is shown in Table 1. Bespoke data management tools (VB.NET Windows application; PostgreSQL database functions) are being developed in order to semi-automate the task of maintaining the WMDB and ensure on-going adherence to the relevant metadata standards.

Table 1. Summary of the contents of the WISERD Metadatabase

| Data type | Data source | Statistics | Geographic data link |
|---|---|---|---|
| *Quantitative* | Surveys (Government and Academic) | 96 Surveys (9760 questions) | Official UK administrative boundary data |
| | Administrative data | 7 Years of Education Data | Official UK administrative boundary data |
| *Qualitative* | WISERD interview transcripts | 47 Interview Transcripts | Gazzeteer based geo-referenced locations (x,y coordinates) |
| | 'Grey' data | 9 PhD Theses | Gazzeteer based geo-referenced locations (x,y coordinates) |

## 2.2 Technology Stack

The basic system architecture of the WGP is shown in Figure 1. It shows how two bespoke software programs have been developed in order to populate the WMDB. The first of these is a web-based qualitative metadata entry tool that allows users to automatically generate semantically important keywords and geographic place names from text documents using the OpenCalais web service (OpenCalais, 2011) and the EDINA Unlock web service (EDINA, 2011) respectively. The second application is a desktop program which is designed to streamline manual data entry of survey questionnaires whilst generating standards-compliant metadata down to the question-response level. Both of these programs upload data to a database server running PostGIS where it is accessed by the end user via a secure web server (running Windows Server 2008 OS) based at the University of Glamorgan.
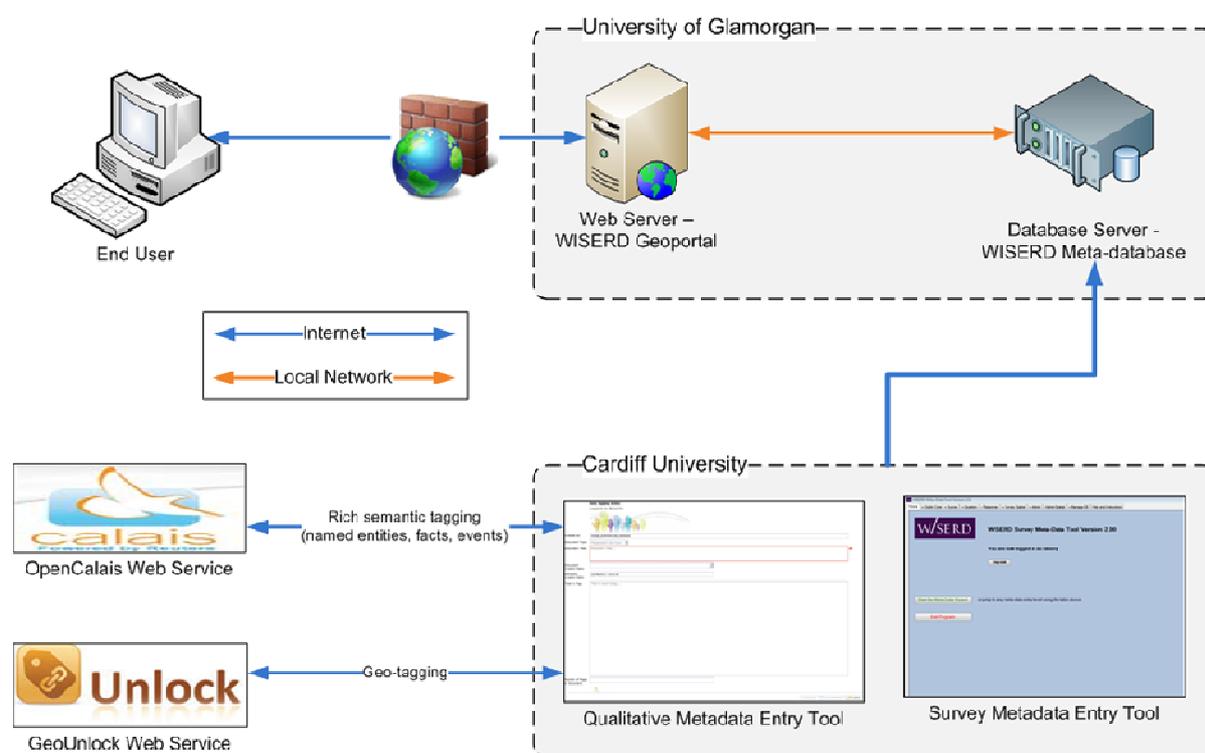


**Figure 1.** Basic system architecture of the WISERD Geoportal

As discussed previously, we have used exclusively FOSS components to develop the WGP, an approach that was felt offered the best flexibility in terms of being able the select and customise the best available applications and code libraries for producing a geoportal (Figure 2). A PostgresSQL database, spatially-enabled with a PostGIS database extension was selected for handling the data management duties as it offered better functionality than alternative FOSS databases such as MySQL and SpatialLite (Obe and Hsu, 2011). Importantly, PostGIS is able to output data in a variety of formats suitable for consumption by JavaScript libraries such as OpenLayers, GeoExt and ExtJS that were used to create the WGP's graphical user interface (Figure 2). The ASP.NET MVC development environment was used to create the WGP application itself on the web server (running IIS7/Apache), with the latest version of Geoserver handling the rendering of map layers generated from database queries. The open source theme is extended to the base mapping data, which is currently provided by the Ordnance Survey (OS OpenData<sup>TM</sup>).
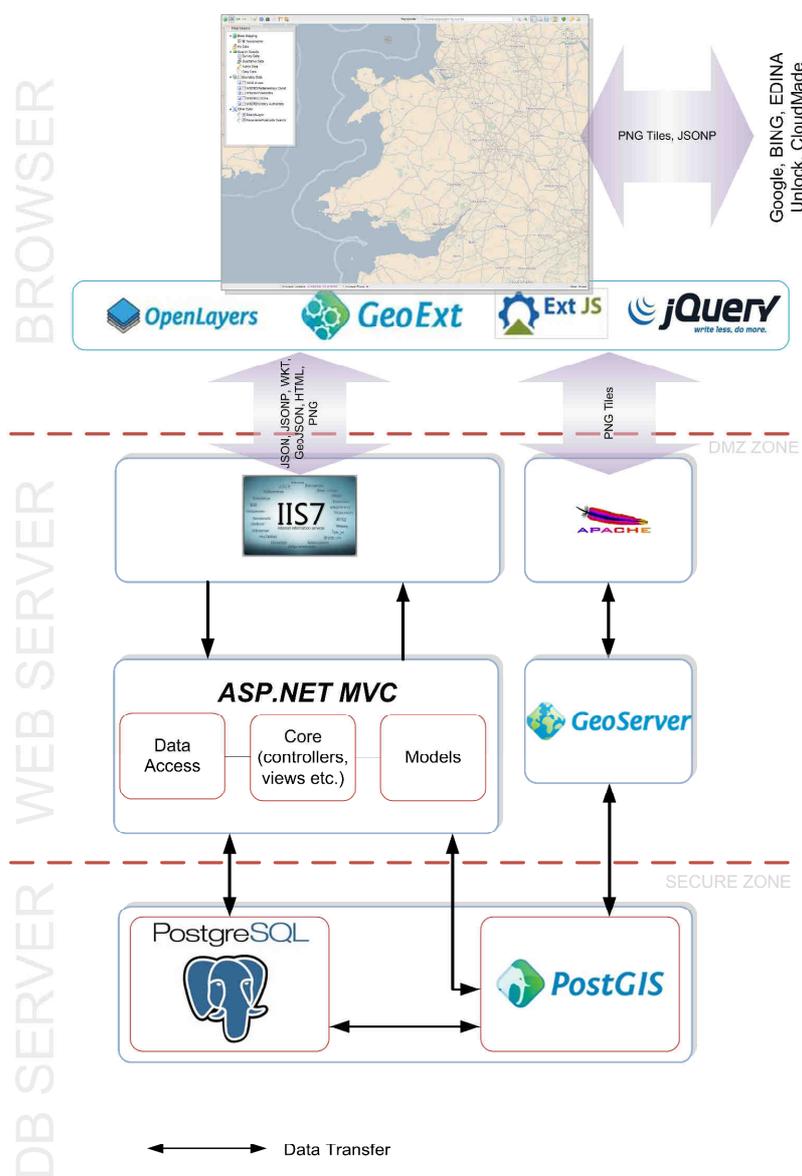


**Figure 2.** The WISERD Geoportal software stack

## 2.3 Functionality

The main window of the WGP GUI is shown in Figure 3. From this main window, users are able to access four different kinds of tools for: 1) account management (add/edit personal details, report bugs, 'bookmark' maps, print view; 2) basic GIS analysis (measure tools, 'identify features, annotate map); 3) textual database search and 4) spatial database search. Of these the latter search tools are the most significant as these are used to discover, visualise and analyse the metadata in the WMDB.
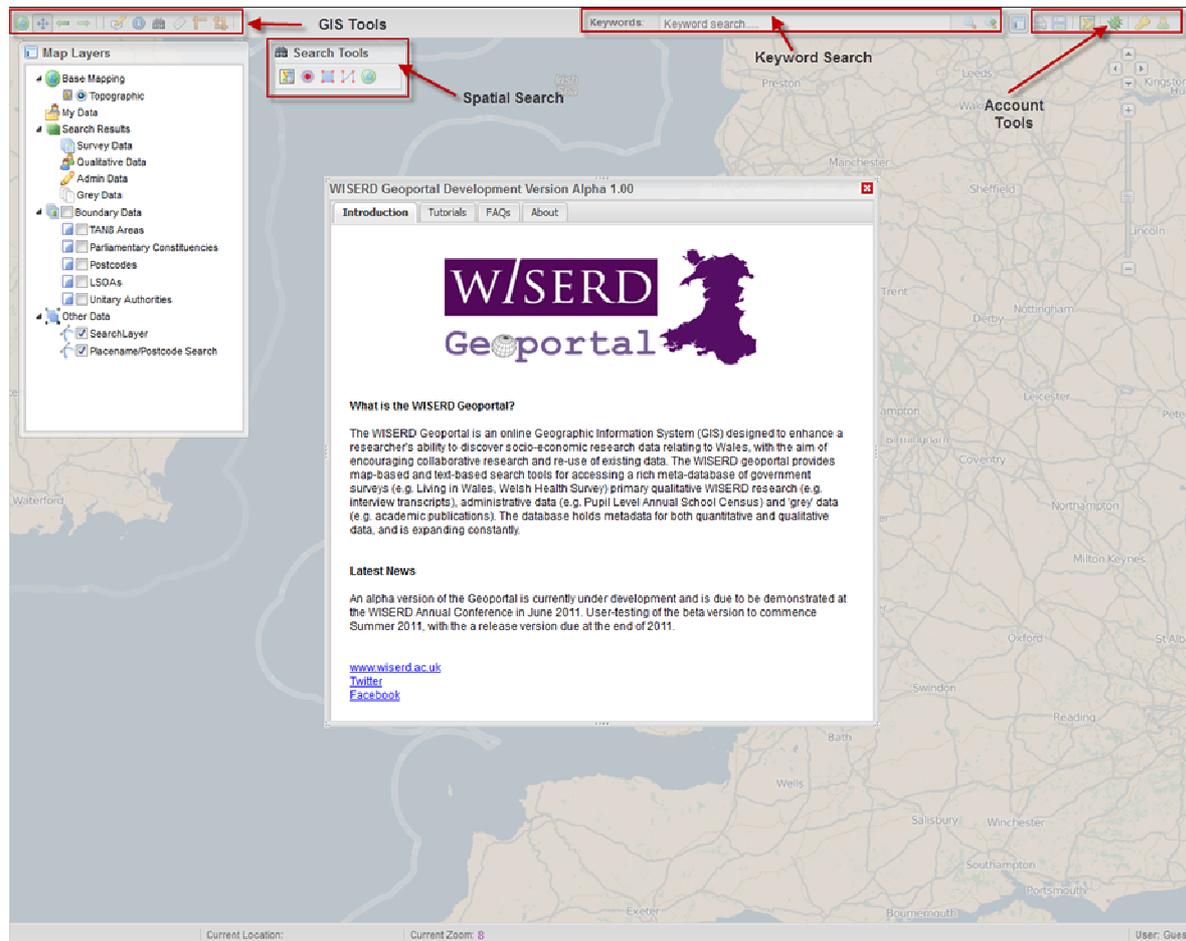


**Figure 3.** The main window of the WISERD Geoportal's graphical user interface (GUI)

The text search utilises the PostgreSQL text searching functionality to search for matching keywords in the quantitative and qualitative metadata records, although a more advanced search form is currently being developed that will allow users to refine their searches using a number of parameters. Results from textual database searches are then returned as a table within a JavaScript (ExtJS) window. Spatial searches can be conducted in a number of ways. Users can: 1) enter a place name and specify a distance (circular buffer) around that location in which to conduct a search; 2) drop a pin anywhere on the map and specify a buffer distance; 3) click and drag a rectangular search area or; 4) draw a custom polygon in which to perform the search. The results of a spatial search are returned as map layers (users can specify which ones to render) in the WGP's table of contents (Figure 4).
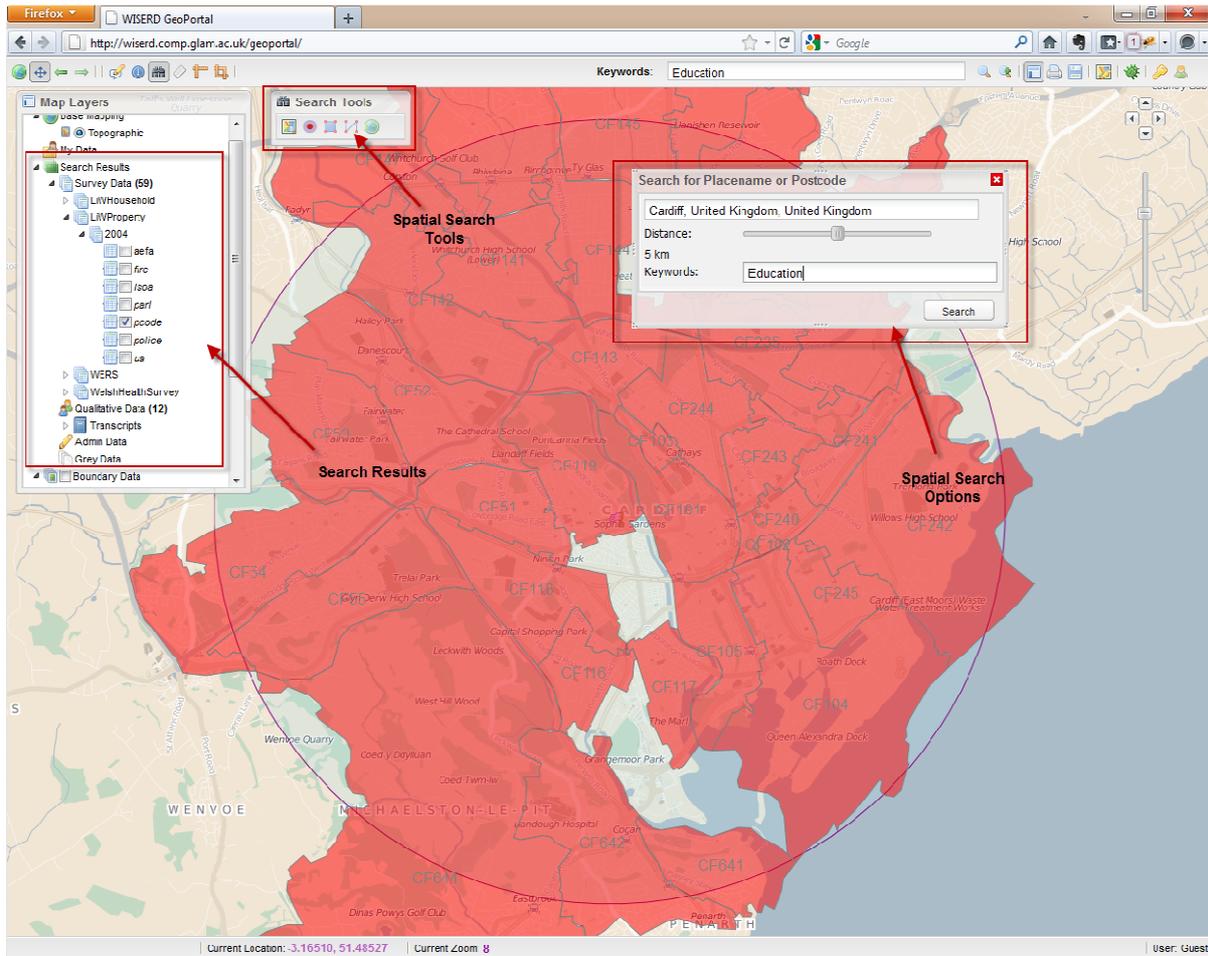
**Figure 4.** Mapping the results of a spatial search using the WISERD Geoportal

Once records have been retrieved from the database following a text or spatial search, the user is able to explore the metadata in more detail using a number of tools. For example, detailed metadata forms can be viewed for each record. For the surveys this would include comprehensive metadata at the survey, question and response level as well as a full Dublin Core metadata record. Analysis tools are also being developed in order to extract further value from the metadata. For example, where survey response data exists, interactive charting tools can be accessed that enable a user to visualise response statistics for a particular administrative/geographical area. A number of tools also exist for visualising qualitative metadata. These include a 'word cloud' function to quickly and effectively visualise the occurrence of key words within an interview transcript for example, and a charting tool that allows users to draw graphs showing the frequency of selected place names in a text document (e.g. academic journal paper). The purpose of these tools is to provide a researcher with a feel for the thematic and geographic focus of a research interview, without disclosing any potentially sensitive information.

## 3. Future Developments

The WGP is currently in its 'Alpha' stage of development with delivery of a 'Beta' release version scheduled for the end of December 2011. Further development of the WGP will then continue until at least July 2012, with further funding opportunities being sought to sustain its future in the long term. In the meantime the team will be working on several important developments, including:

- Structured user evaluation/usability testing (early 2012).
- Re-branding of the WGP (feedback from dissemination events has indicated that the term

*geoportal* is unsuitable given the target audience of social scientists who are primarily non-GIS users).

- Continued expansion of the WMDB.
- GUI improvements (taking into account usability evaluation feedback).
- Improved cartographic rendering.
- Expansion from Wales-only to UK-wide metadata.
- Investigating the use of other semantic tagging services (e.g. GEMET Thesaurus) for generating metadata.
- Transfer of web server/database infrastructure to a cloud computing platform for greater reliability and scalability
- Automated data exchange between the WGP and external organisations (e.g. Welsh Government, UK Data Archive) and service provision to important initiatives such as the UK Location Programme and www.data.gov.uk.

## 4. Conclusions

The WGP has been specifically developed to improve the discoverability, accessibility and usage of social science data amongst academics and policy makers in Wales. In our view it is a unique geoportal in that it is dedicated to supporting both quantitative and qualitative social science research, pioneering a new application area for web-based GIS. Having both spatial and non-spatial data of this nature integrated into a single web-based framework is a positive step forward from the current status quo - the WISERD Geoportal now offers the opportunity for a 'one stop shop' where the need for a desktop GIS to conduct basic spatial analysis and visualisation of social science data could be potentially dispensed with. Also, the wide range of disparate metadata contained in its database is proving to be attractive to potential end-users, providing the impetus for continued development of the WISERD Geoportal in order to establish it as an indispensible tool for researchers and analysts working with social science research data.

## 5. Acknowledgements

## 6. References

Armstrong M, Nyerges L, Wang S and Wright D (2011). Connecting geospatial information to society through cyberinfrastructure, in Nyerges TL. and Couclelis, H. and McMaster, R. (Eds), *The SAGE Handbook of GIS and Society*, Sage Publications Ltd., London, 109-122.

Beaumont P, Longley P and Maguire D (2005). Geographic information portals—a UK perspective. *Computers, Environment and Urban Systems*, 29(1), 49–69.

Berry R, Fry R, Higgs G, and Orford S (2010). A Geo-portal for Enhancing Collaborative Socio-Economic Research in Wales Using Open Source Technology. *The Journal of Applied Research in Higher Education*, 2(1), 77-92.

CIOC (2009). Open source, open standards and re-use: Government action plan. Accessed 21 November 2011: http://www.cabinetoffice.gov.uk/media/123372/090224opensource.pdf

EDINA (2011). Unlock the potential in your data with our simple web services. Accessed 21 November 2011: http://unlock.edina.ac.uk/

Goodchild M, Fu P and Rich P (2007). Geographic information sharing: The case of the Geospatial One-Stop portal. *Annals of the Association of American Geographers*, 97(2), 250-266.

Giuliani G and Peduzzi P (2011). The PREVIEW Global Risk Data Platform: a geoportal to serve and share global data on risk to natural hazards. *Natural Hazards and Earth System Sciences*, 11, 53-66.

Lance KT, Georgiadou YP, and Bregt AK (2011). Evaluation of the Dutch subsurface geoportal: What lies beneath? *Computers, Environment and Urban Systems*, 35 (1), 150-158.

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

Masser I (2011). Emerging frameworks in the Information Age: The Spatial Data Infrastructure (SDI) Phenomenon, in TL Nyerges, H Couclelis and R McMaster (Eds.) *The SAGE Handbook of GIS and Society*, Sage Publications Ltd., London, 271-286.

National Data Strategy (2009). UK Strategy for data resources for social and economic research 2009-2012, ESRC, Polaris House, Swindon, UK.

Obe R and Hsu S (2011). PostGIS in Action. Manning Publications. Greenwich, Connecticut. USA.

OpenCalais (2011). How Does Calais Work? Accessed 21 November 2011: http://www.opencalais.com/about

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

Tang W and Selwood J (2005). Spatial Portals: Gateways to Geographic Information. ESRI Press, Redlands, CA.

**Biographies**

Dr Robert Berry is a post-doctoral researcher at the GIS Research Centre, University Glamorgan, and is part of the WISERD Data Team. His research interests include public participation GIS, landscape visualisation, web GIS, open source GIS, and the application of GIS for social science research and hazard and disaster management.

Dr Richard Fry is a post-doctoral researcher at the GIS Research Centre, University Glamorgan, and is a part of the WISERD Data Team. His research interests include web and open source GIS and the application of GIS in transport and health research.

Prof Gary Higgs is currently Director of the GIS Research Centre in the Faculty of Advanced Technology, University of Glamorgan and a co-Director of WISERD at Glamorgan. Over-arching research interests are in social and environmental applications of GIS, most recently in the areas of health geography, the emergency services and environmental justice.

Dr 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. His research interests include the analysis and modelling of socio-economic process, innovation in data integration and data construction and the mapping and spatial analysis of qualitative data.

Samuel Jones is a research assistant at WISERD based at Cardiff University and is part of the WISERD Data Team. He has responsibility for providing technical assistance and data management skills to the WISERD team and assisting with the data integration and management programme.