University of Glasgow | Experimental Particle Physics

GLAS-PPE/2009-21
19th November 2009

Department of Physics and Astronomy
Experimental Particle Physics Group
Kelvin Building, University of Glasgow
Glasgow, G12 8QQ, Scotland
Telephone: +44 (0)141 330 2000 Fax: +44 (0)141 330 5881

Supporting Security-oriented, Inter-disciplinary Research: Crossing the Social, Clinical and Geospatial Domains

R.O. Sinnott (1), T. Doherty (1), C. Higgins (4), P.S. Lambert (3), S. McCafferty (1), A.Stell (1), K.L.L. Tan (2), K. J. Turner (2), J.P. Watt (1)

1 National e-Science Centre, University of Glasgow, Glasgow, G12 8QQ
2 Dept. of Computing Science & Mathematics, University of Stirling, Stirling, FK9 4LA
3 Dept. of Applied Social Science, University of Stirling, Stirling, FK9 4LA
4 EDINA, University of Edinburgh, Edinburgh, EH9 1PR

Email: r.sinnott@nesc.gla.ac.uk or t.doherty@physics.gla.ac.uk

Abstract

How many people have had a chronic disease for longer than 5-years in Scotland? How has this impacted upon their employment? Are there any geographical clusters in Scotland where a high-incidence of patients with such long-term illness can be found? How does the life expectancy of such individuals compare with the national averages? Such questions are important to understand the health of nations and the best ways in which health care should be delivered and measured for their impact and success. In tackling such research questions, e-Infrastructures need to provide tailored, secure access to an extensible range of distributed resources including, amongst others, primary and secondary e-Health clinical data; social science data; and geospatial data. In this paper we describe the security models underlying these e-Infrastructures and demonstrate their implementation in supporting secure, federated access to a variety of distributed and heterogeneous data sets, exploiting the results of a variety of projects at the National e-Science Centre (NeSC) at the University of Glasgow.

1. Introduction

Much scientific research now crosses the boundaries of individual research disciplines. For example, if one considers studies in particular chronic diseases and the response of specialised study-specific treatments then this can require interplay between the clinical sciences, the biological sciences, the social sciences and the geospatial sciences amongst others. To accommodate such inter-disciplinary research, e-Research infrastructures need to support the seamless and transparent linkage across disciplines and the resources they offer. Furthermore, this has to be aligned with the way in which the researchers themselves wish to work, and satisfy all concerns associated with the numerous stakeholders in this space, e.g. on security, access control and individual privacy. This is made especially challenging given the evolving nature of science and the associated resources available and the often dynamic nature of collaborations themselves.

For many disciplines that are primarily computationally-bound, i.e. where access to large high performance computing (HPC) facilities is the primary handicap restricting scientific progress, resources such as the UK e-Science National Grid Service (www.ngs.ac.uk) and ScotGrid (www.scotgrid.ac.uk) are available. However for many would be users, these resources are still largely established for the computationally savvy who are proficient with, or at least happy to deal with the nuances and intricacies of complex Grid middleware. Indeed the initial step in gaining access to such facilities is the requirement to obtain a UK e-Science X.509 certificate. This step is off-putting for many users especially since it often requires them to refer to registration authorities (which may not exist at their particular institution) and once allocated, convert these certificates to formats that are suitable for accessing and using Grid resources. These issues are described in detail in (Sinnott 2006, Watt 2007). Furthermore, these HPC facilities are primarily targeted to user communities to compile and run their own simulations although we note that progress in supporting application portfolios - albeit HPC-oriented applications - is now underway through the NGS job submission portal for example (https://portal.ngs.ac.uk/). We claim that the vast majority of researchers that could potentially benefit from research infrastructures are not limited by lack of access to HPC resources. Rather, it is the lack of targeted services and research environments that limit their research and, especially with regard to inter-disciplinary research, the interconnectedness of these research environments.

The service-oriented architecture paradigm where Grid services provide access to resources is potentially more aligned with the actual needs of research communities. In this model, researchers should, at least in principle, be able to make use of distributed resources without necessarily being savvy computational scientists. Instead, tailored environments should allow the coupling of collections of these services together for specific research purposes, e.g. through portals or workflow environments, as part of a user-oriented research framework. A fundamental requirement in the realisation of this for many disciplines is ensuring that these services and the resources they make available are only accessible to legitimate individuals. The framework of rules and regulation by which the legitimacy of an individual or collaboration more generally can be defined will vary from project to project and more generally from domain to domain. In e-Research and Grid parlance, such a framework is typically expressed through virtual organisations (VO) which identify the resources to be shared and the terms and agreements by which these resources can be used by researchers. Key to the success of any service-oriented

architecture-based VO model is the delivery of the environments aligned with, and driven by, research needs and working practices. These environments should allow the various stakeholders in this space, e.g. data providers, to define and enforce their own policies on access control for example.

Our focus in this paper is to describe how security-driven e-Infrastructure solutions can be established and be aligned with the way research communities are comfortable in accessing internet resources more generally. Through exploiting the Internet2 Shibboleth technologies (http://shibboleth.internet2.edu) and the UK Access Management Federation (www.ukfederation.org.uk), and offering supporting tools that allow for a range of security models for a variety of distributed and heterogeneous services combining centralised and federated VO models (Sinnott 2007), we demonstrate how one of the fundamental tenets of the Grid model, namely single sign-on, can be supported to enable user-oriented interdisciplinary research. We emphasise that this single sign-on model goes beyond the X509 authentication-based models as typified with HPC access control through Globus grid mapfiles (http://www.globus.org) for example, to include single sign-on models with finer-grained authorisation infrastructures. To demonstrate this, we show how the seamless interplay of collaborative research environments and resources can support research into a broad range of social, clinical and related research areas. We focus in particular upon the Scottish research landscape and specific clinical, social and geospatial resources – each of which has their own particular access control policies. However we emphasise that the models and paradigms put forward are more widely applicable and have been adopted in other research projects at NeSC Glasgow.

The rest of the paper is structured as follows. Section 2 describes the data landscape in Scotland focusing in particular upon clinical data sets, social science data sets and geospatial data sets related to understanding the health and societal impacts of health on the nation. Section 3 describes the security models and tools support that support a range of distributed single sign-on models. Section 4 describes a case study in accessing and using these distributed data sets focusing on the area of depression and self harm. Finally in section 5 we draw conclusions on the work and outline future areas of work.

2. The Scottish Data Landscape

One of the major obstacles facing substantive research is in discovering, accessing and using data. This is especially challenging when inter-disciplinary research is undertaken, where a multitude of heterogeneous data sets exists. At the National e-Science Centre at the University of Glasgow (www.nesc.ac.uk) we have been involved in a wide range of projects covering the clinical, social and geospatial domain. In the clinical domain these include amongst others (see www.nesc.ac.uk/hub/projects) the MRC funded Virtual Organisations for Trials and Epidemiological Studies (VOTES – www.nesc.ac.uk/hub/projects/votes) project; in the social science domain these include the ESRC funded Data Management through e-Social Science (DAMES – www.dames.org.uk) project and in the geo-spatial domain these include the JISC funded Secure Access to Geospatial Services (SeeGEO - http://edina.ac.uk/projects/seesaw/seegeo/) project.

2.1 Clinical Data Landscape

For the VOTES project referred to above, the focus of the work was to develop a Grid-enabled framework to support the different processes involved in establishing and running clinical trials and epidemiological studies. These phases included

recruitment to clinical trials; study management of clinical trials and data collection throughout the course of a given clinical trial. The infrastructure developed within VOTES was based upon the Grid concept of a Virtual Organisation (VO) but adapted to the specific security and ethics concerns of the clinical domain to establish clinical virtual organisations (CVO). More information on the VOTES project and its application are described in (Stell 2007, Ajayi 2007, Sinnott 2008).

As part of the work of VOTES a detailed enumeration of the software systems and data sets across Scotland was undertaken. For primary care systems, the predominant software in use is the General Practice Administration System for Scotland (GPASS - www.gpass.scot.nhs.uk). This is in fact used by over 85% of GPs across Scotland. GPASS primarily serves as a way for GPs to effectively organise their data electronically, including managing patient data and administrative data associated with their individual practices. Furthermore GPASS allows uploading of a subset of their patient data to an aggregated repository managed by their local authorities.

For secondary care systems, one of the most widespread of software systems is the Scottish Care Information (SCI) Store (www.sci.scot.nhs.uk). SCI store is a regional clinical repository used by many hospitals that holds details of inpatient and outpatient data along with a wide variety of other hospital specific data sets, e.g. lab results, treatments and inpatient/outpatient summaries. SCI store also allows for upload of data from GPASS systems and deals with patient referrals for example.

Both SCI store and GPASS provide front-end software to specific clinical back-end databases. These databases follow a prescribed schema definition from the Information Services Division (ISD – www.isdscotland.org) of the Scottish NHS. GPs build up individual data-sets about their own particular practice areas including their patient data and their administration for example. Subsets of this information are then uploaded to a central repository hosted by SCI Store. The data from the numerous SCI store installations across Scotland is then placed into centralised clinical data warehouses.

One of the major clinical data resources that exist across Scotland is the Scottish Morbidity Records (SMR). SMR data sets are constructed in conjunction with the General Register Office (GRO) for Scotland. They consist of many a rich variety of clinical data often going back decades that describe the health and clinical information of the Scottish population. These data sets cover: hospital information (inpatients/outpatients/treatment etc); psychiatric admissions and discharges; cancer registrations and deaths.

Key to the integration of the Scottish primary, secondary and SMR data sets is the existence of the Community Health Index (CHI) number – a ten-digit identifier formed in part by the date of birth of the individual and their sex (male/female). This unique identifier has been rolled out across Scotland since 2006 and allows for identification and linkage of patient records across primary care data, secondary care data and longer term records such as SMR to take place.

We note that almost all of these data sets are geo-spatially referenced in some manner either through postcodes of the individual patients; of the GP practices and/or of the hospitals themselves.

2.2 Social Data Landscape

The social sciences are facing numerous challenges in dealing with and management of data. The DAMES project is exploring the issues associated with addressing these

data management challenges through e-Science technologies and processes. In particular DAMES is focusing upon four key themes associated with social science data management. These include data management activities associated with occupational data; with ethnicity data; with educational data and with e-Health data.

Building upon previous work such as the ESRC funded Grid Enabled Occupational Data Environment (GEODE – http://www.geode.stir.ac.uk), DAMES is exploring and cataloguing the wealth of existing data sets associated with the four themes. Many of these themes include documenting the numerous surveys that have been conducted by a multitude of social science researchers and centres. Some of the more prominent resources available to the social scientist in the UK include the UK Data Archive (www.data-archive.ac.uk), Economic and Social Data Service (www.esds.ac.uk), MIMAS (www.mimas.ac.uk), the Office of National Statistics (ONS - www.ons.gov.uk) and the Scottish Census Results On-Line (SCROL - www.scrol.gov.uk/). The SCROL and ONS resources incorporate the major source of statistical information in the UK related to a broad range of areas: from households, to health to occupational data sets amongst numerous others.

Numerous other international resources also exist such as the Council of European Social Science Data Services (CESSDA - www.nsd.uib.no/cessda) which offers a co-ordinated gateway to various national social survey micro-data data archives. Other more specialised social science resources also exist. Examples of these in the occupational research domain (as one of the many examples that might be selected) include: IPUMS (www.ipums.org) which provides access to extensive volumes of census micro-data and specialist information associated with it, such as occupational information; EurOccupations (www.euroccupations.org/main); the GEODE resource itself and the History of Work Information System (http://historyofwork.iisg.nl).

Linking these data sets with other resources through any infrastructure requires an awareness of the data standards in use. This in turn is greatly benefited from harmonisation of the data resources and standards adopted. Organisations such as the ONS and CESSDA are currently undertaking data harmonisation projects in the UK and across Europe respectively. In this regard, numerous other complementary efforts also exist such as the Data Documentation Initiative (DDI) metadata standard (www.icpsr.umich.edu/DDI) which provides a standard metadata framework widely used in the social sciences.

Many of the data sets currently being explored in the DAMES project include those with variables related to health. Specifically, we have focused initially upon Census related variables that deal with health and well being. These data sets themselves are geo-spatially referenced and include output areas, e.g. the local authorities.

2.3 Geospatial Data Landscape

Many data sets are geo-spatially referenced, e.g. including postcodes, named locations and cities or wider output areas such as local authorities. Such geo-spatial information referencing can be used for in a variety of ways. In the case of projects such as VOTES, this might be to better understand the feasibility of undertaking a particular recruitment or data collection activity for a clinical trial. Thus are there particular hot-spots where patients might be recruited from? For projects such as DAMES, geo-spatially referenced data allows amongst many other possibilities, to understand local social patterns and consider these against a national trend. This comparison can be with regard to health, education, social mobility, immigration amongst other social-related possibilities.

With regard to geo-spatial data resources, the EDINA geo-spatial data service (www.edina.ac.uk) provides a key resource used within the UK for a wide range of geo-spatially oriented research. The EDINA resource is aligned with wider international efforts in accessing and using geo-spatially referenced data including the EU INSPIRE initiative (http://www.inspire-geoportal.eu) and the Open Geospatial Consortium (OGC) efforts to standardise the interfaces and protocols to geo-spatial data (www.opengeospatial.org).

EDINA provides access to a range of services that allow access to a variety of geo-spatial data sets. These include licensed data from the Ordnance Survey. These data include UK Borders data sets covering for example local authority boundaries and how these have changed over time; and DigiMap resources that provide a variety of ways in which other information can be overlaid onto a given set of maps or geo-spatial co-ordinates. This includes the support for Web Map Service (WMS) which responds to requests by creating map images of spatial data; Web Coverage Services (WCS) which allow access to the raw data which can then be used for further analysis or for portrayal if required, and Web Feature Sets (WFS) which allow to add a range of features over a given map set. More information on EDINA resources is available in (Higgins 2009).

As well as resources such as EDINA, there are a multitude of other geo-spatial resources and software systems that are now available for dealing with and visualising geographical data sets. GoogleMaps represents one of the more widespread systems for visualisation of mapping information. Many other data sets exist in related centres such as the British Atmospheric Data Centre which includes information on weather, pollution and a variety of other related environmental data sets.

3. E-INFRASTRUCTURE & E-SECURITY MODELS AND TOOL SUPPORT

Access to a range of existing data sets cannot be achieved without due consideration of information governance and associated security requirements. Many data providers, especially clinical data providers need to be convinced that all access to and use of their data sets follows strict protocols to address wider public concerns on data privacy and confidentiality. These security models should ideally support *single sign-on* where once authenticated and authorized a user is allowed to access a range of distributed resources without the need for further re-authentication and/or authorization at those resources. This has to be achieved in a manner that addresses the general usability from the end user considerations, i.e. they do not wish to focus on e-Infrastructure security but instead should be provided transparent access to resources that allows them to conduct their own research – ideally with no or at least minimal direct exposure to the underlying security infrastructure.

As noted, e-Infrastructures and e-Security has to address the requirements of the many protagonists in this space including direct data providers, third party data providers, as well as the end user researchers themselves. One of the main requirements of data providers such as the NHS and organizations such as EDINA is local autonomy. Sites wish to define and enforce their own local access control decisions, whilst users do not wish to immerse themselves in credential management and dealing with certificates used for authentication and/or authorization. To address the usability requirements, many projects developing research oriented and inter-disciplinary research oriented e-Infrastructures address end user ability through portal based solutions which in turn can be used to support virtual research environments offering

a one stop shop where data resources, simulation packages and related resources are made available.

Given this model, access to this portal should be made simple and secure for end user researchers and aligned with the way in which they wish to work. Furthermore, the contents of this portal should recognize the privileges and credentials of the end users themselves and be configured accordingly. It is key that data providers are able to define and enforce their own local access control decisions and not simple delegate all security enforcement to the portal itself. Many portal frameworks including GridSphere (www.gridsphere.org) have been developed for Grid-oriented research infrastructures, with recent evolutions of portal frameworks such as Sakai (www.sakaiproject.org) and LifeRay (www.liferay.com) supporting aspects of user–oriented access and configuration. The Open Middleware Infrastructure Institute (OMII-UK – www.omii.ac.uk) SPAM-GP project (www.nesc.ac.uk/hub/projects/omii-sp) has developed a family of JSR-168 (JSR-168, 2003) compliant portlets that allow for secure access and configuration of portals with specific focus upon the exploitation of the Internet2 Shibboleth federated access control technologies. These include the Scoped Attribute Management Portlet (SCAMP) for defining which sites across the UK Access Management Federation to trust (the default is implicitly trusting all sites); the Content Configuration Portal (CCP) for configuring the contents of the portal based upon the privileges of the end users being transferred by Shibboleth from their home Identity Provider (IdP); and the Attribute Certificate Portlet (ACP) which allows for creation of X509 Attribute Certificates (AC) which can be used by remote service providers to define and enforce their own local access (authorization) policies based upon these digitally signed and tamperproof credentials. Currently we have exploited the PERMIS authorization technology for this purpose (Chadwick 2003) however other authorization engines can equally well be applied, e.g. XACML (XACML 2005). A typical architecture for the SPAM-GP portlets and their use in accessing Web Feature Sets from EDINA and Census data sets from MIMAS (www.mimas.ac.uk) is shown in Figure 1.

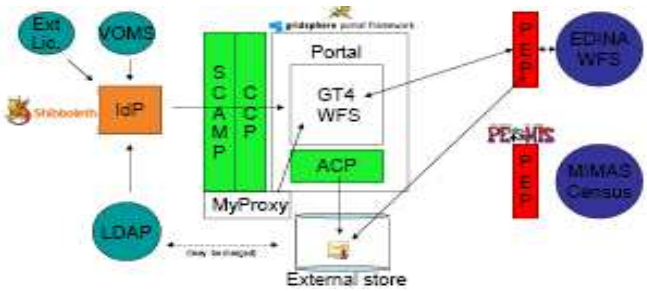

Figure 1: SPAM-GP Architecture

In this model, when a user attempts to access a Shibboleth protected portal, they are redirected to the UK Access Management Federation Where-Are-You-From (WAYF) service and select their home institution. After authenticating at their home institution with their own user names and passwords, a signed Security Assertion Markup Language (SAML) assertion is returned which includes a selection of *eduPerson* attributes (standardized across the UK Federation). In this, virtual organization specific attributes can be embedded. We have adopted the *eduPersonEntitlement* attribute for this person. This SAML assertion is intercepted and checked whether it meets the local access policy as defined by the SCAMP policy which in essence filters the UK Access Management metadata information, i.e. which IdPs are recognized.

The security information (roles/attributes) are then used by the CCP to configure the contents of the portal which involves associating roles with portlets that are accessible. Both the SCAMP and CCP will be used by a virtual organization administrator responsible for deciding who is allowed access to which resource (portlet). Following this, a user is allowed access to a selection of portlets which can be used for a variety of purposes, e.g. to invoke remote data or service providers. However we recognize that many providers will not simply delegate responsibility for access control to a remote portal. To this end, the user can then use the ACP to create an X509 based attribute certificate which is stored in a local attribute store associated with the portal. Creation of X509 proxy credentials is currently undertaken through a further portlet which gives access to a MyProxy server (Novotny 2001), however work is on-going to remove this step and automatically create proxy credentials using information returned from the IdP, namely the encrypted username and password for the MyProxy server. This X509 proxy credential is used when invoking a remote service through a portlet. Upon identifying the distinguished name (DN) of the user, the service (assuming it has finer grained authorization requirements) will automatically pull the necessary attributes from the portal attribute store for that particular user, and use this to make a local access control decision.

We note that richer scenarios are also possible, e.g. where multiple attribute certificates from multiple attribute authorities are required for a single authorisation decision. Such scenarios are currently being explored in the ShinTau (http://sec.cs.kent.ac.uk/shintau/) project where three separate attribute authorities are to be used to make a single authorization decision (Inman 2007).

To understand how the portlets that have been created in the SPAM-GP can be applied to support a variety of inter-disciplinary access control models, we outline a case study in the area of depression, self harm and suicide – one of the focal points for research in the DAMES project.

4. CASE STUDY ON DEPRESSION AND SELF HARM

Depression, self-harm and suicide are recognised as increasingly important areas of public health research and a wealth of public health research topics are currently being explored in this space. Understanding the patterns and classification of depression across populations and society (sex, age groups, occupation, marital status, educational background, accommodation-type, household income, ethnicity, lifestyle, access to green fields/parklands) are all of relevance to a better understanding of depression, suicide and self-harm. Explorations of subjective well-being using social survey data have increasingly matched in geographical and environmental indicators.

Key to supporting such research is the ability to catalogue and link relevant datasets with a view to developing a better understanding of the multiple psychosocial risk factors for depression, suicide and self harm. In realizing this, access to nationwide prescription drug resources such as e-Pharmacy (which is already rolled out across Tayside and planned across Scotland and to be linked through SHIP) will allow for the first time researchers to securely, access automated clinical and pharmacological resources. This will allow a better understanding of the impact of drugs and prescription drugs, dependencies on drugs such as anti-depressants and drug combinations, and their impact and risk on depression/self harm and suicide. Linking such clinical resources with environmental and geospatial resources, e.g. with EDINA for maps/services that allow to capture green fields/parkland; information of proximity to major traffic junctions or airports/flight paths. Currently such research

requires a variety of data resources to be accessed individually, each with their own specific security requirements including licensing arrangements for mapping data through to ethical approval for access to and usage of clinical data.

To support this, proof of concept systems have been established that allow seamless access to a range of clinical resources including health related variables from the Census data; Scottish Morbidity Records related to hospital admissions and inpatient/outpatient information (SMR01); Scottish Morbidity Records related to mental health and psychosis episodes (SMR04), and Scottish Morbidity Records related to death by self harm or alcohol related causes (SMR99). Each of these data sets is geo-spatially referenced. Applying the SPAM-GP solutions results in the interfaces (portlets) to distributed individually protected services as shown on the left of Figure 2. The result of querying and linking the results from these remote data providers is shown on the right of Figure 2. We note that key to the supporting joining of these results is an understanding of common fields across the data sets, e.g. use of the Community Health Index number and/or other fields such as the postcodes or output areas associated with these data sets.

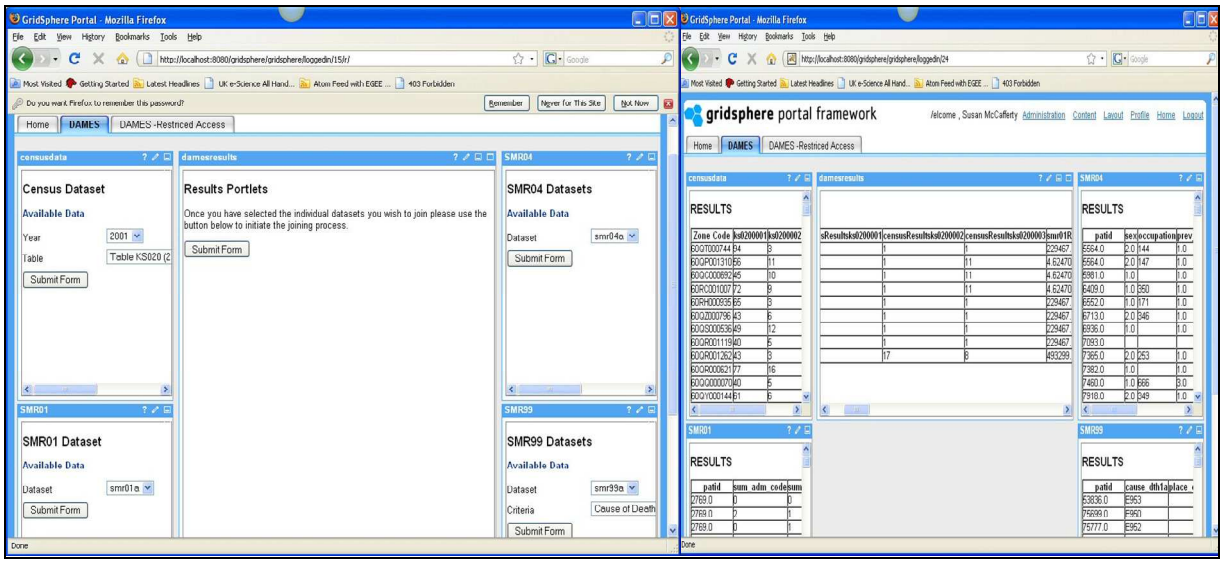

Figure 2: Census & SMR Data Portlet Query Interfaces (left) and Returned/Joined Results (right)

Understanding the geographical distribution of cases related to depression, self-harm and suicide across Scotland and be able to *drill down* into the particular areas with specific focus upon areas of high incidence is especially important for public health researchers. Figure 3 (left) shows the geospatial distribution of depression, self-harm and suicide across Scotland where the darker areas represent lower recorded incident rates. The higher number of depression, self-harm and suicide incidents in the representative SMR data sets that we were given show that there is an especially high level of incidence in the central belt region of Scotland (between Glasgow and Edinburgh). Figure 3 (right) shows the result of drilling down in to the data where the data for the Greater Glasgow region has been selected. This is achieved through use of appropriate shapefiles from EDINA (available subject to licensing agreements) and use of geo-spatial packages. In this instance the overlaying of data has been undertaken using the R-statistical package. We are currently also exploring other geospatial tools for this purpose.

We emphasise that this geospatial distributed shown is based upon the representative data sets that were provided by the Information Statistics Division of the NHS in Scotland. This is a pseudonymised collection of actual NHS data comprising over 4 million records. This pseudonymisation includes the removal/scrambling of many identifying data elements including the names, addresses and dates of birth of individual patients, but it does include many other fields that have not been tampered with, e.g. output areas, and as such is not simply representative but accurate.

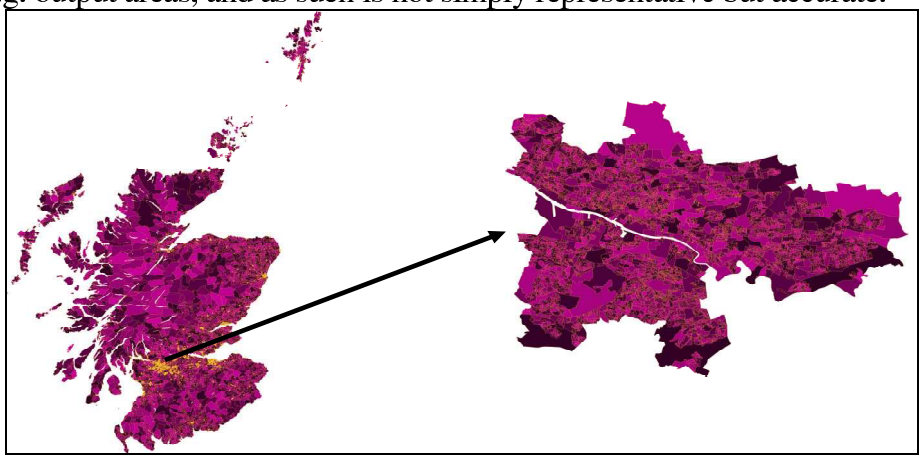

Figure 3: Mental Health / Depression Related Data sets overlaid across Scotland (left) and showing data for Greater Glasgow (right)

5. CONCLUSIONS AND FUTURE WORK

This work has shown that the Grid and e-Science vision of seamless, secure access to inter-disciplinary research data and resources can be supported. Through fine grained expression of access and usage policies, local policy enforcement can be realized and sites remain autonomous. This work is widely applicable and indeed the security models put forward here are being applied in a wide variety of projects at NeSC in Glasgow. We recognize however that many open issues remain to be solved. One of the major obstacles that remain to be addressed is trust. This is not just trust in the underlying technologies that they can and do support the required fine grained security, but also trust in the people that are applying these technologies and the general process of supporting security. Information governance, data protection, confidentiality, privacy and ethical considerations are essential considerations to the successful adoption of any technological solutions. Establishing such trust and demonstrating that the security and e-Infrastructure solutions made possible are not weakening or compromising existing security models and the whole process of security is crucial. As a concrete example of this, the existing SMR data sets comprise many hundreds of fields in many tables. In developing secure solutions that allows linkage of these data sets with other data sets care needs to be taken to avoid potential data disclosure where anonymised or pseudo-anonymised data when linked with other data can lead to identification of individuals. The VOTES Vanguard system (Sinnott 2009, Stell 2009) has been developed specifically to tackle this problem and address another major issue in interfacing with providers such as the NHS, namely their wariness of allowing incoming connections through their firewalls. Initial experiences of applying Vanguard have shown that secure linkage and anonymisation is possible where no entity in the system is in possession of information that could be used for further data linkage and potential statistical disclosure. The work on Vanguard will be

further explored and evaluated in the context of the Scottish Health Informatics Platform for Research (SHIP – www.scot-hip.ac.uk) project which began in April 2009.

5.1Acknowledgements

The DAMES project was funded by a grant for the Economic and Social Science Research Council (ESRC); the VOTES project was funded by a grant from the Medical Research Council (MRC); the SeeGEO project was funded by a grant from the Joint Information Systems Committee (JISC). We gratefully acknowledge their support.

6. REFERENCES

Sinnott, R.O., Watt, J., Ajayi, O., Jiang, J., Shibboleth-based Access to and Usage of Grid Resources, *IEEE International Conference on Grid Computing, Barcelona, Spain*, 2006.

Watt, J. Sinnott, R.O., Jiang, J., Stewart, G., Stell, A., Martin, D., Doherty, T., Federated Authentication and Authorisation for e-Science, *Proceedings of APAC 2007 conference, Perth, Australia*, 2007.

Sinnott, R.O., Chadwick, D.W., Doherty, T., Martin, D., Stell, A., Stewart, G., Su, L., Watt, J., 2008, Advanced Security for Virtual Organizations: Exploring the Pros and Cons of Centralized vs Decentralized Security Models, *8th IEEE International Symposium on Cluster Computing and the Grid (CCGrid 2008), May 2008, Lyon, France*.

Stell, A.J., Sinnott, R.O. Ajayi, O., Security Oriented e-Infrastructures Supporting Neurological Research and Clinical Trials, *2nd International Conference on Availability, Reliability and Security*, (ARES'07), Vienna, Austria, April, 2007.

Ajayi, O., Sinnott, R.O., Stell, A.J., Trust Realisation in Collaborative Clinical Trials Systems, *HealthCare conference*, Harrogate, UK, March 2007.

Sinnott, R.O., Ajayi, O., Stell, A.J., Supporting Grid Based Clinical Trials in Scotland, *Health Informatics Journal Special Issue on Integrated Health Records*, Vol. 14 (2), June 2008.

Higgins, C., Sinnott, R.O., Koutroumpas, M., Watt, J., Hume, A.C., Turner, A.G.D., Spatial Data e-Infrastructure, *Proceedings of International Conference on e-Social Science*, Cologne, Germany, June 2009.

Watt, J., Sinnott, R.O., Jiang, J., Doherty, T., Higgins, C., Koutroumpas, M., Tool Support for Security-oriented Virtual Research Collaborations, *IEEE International Workshop on Security in e-Science and e-Research* (ISSR-09), Chengdu, China, August 2009.

Novotny, J., Tuecke, S., Welch, V., An Online Credential Repository for the Grid: MyProxy, Proc. of the Tenth International Symposium on High Performance Distributed Computing, IEEE Computer Society Press, 2001.

JSR 168: Portlet Specification, http://jcp.org/en/jsr/detail?id=168

Cantor S., et al., Shibboleth Architecture: Protocols and Profiles, Internet2-MACE (DocumentID: internet2-mace-shibboleth-archprotocols-200509), http://shibboleth.internet2.edu/docs/internet2-mace-shibboleth-arch-protocols-latest.pdf

Inman, G., Chadwick D., Klingenstein, N., Authorisation using Attributes from Multiple Authorities a Study of Requirements, *HCSIT Summit – ePortfolio International Conference*, Oct. 2007, Maastricht, Netherlands.

Chadwick, D.W., Otenko, O., Ball, E., Role-Based Access Control with X.509 Attribute Certificates, *IEEE Internet Computing*, Mar-Apr 2003, pp. 62-69

The Open Geospatial Consortium (OGC) Web Feature Service (WFS) standard, http://www.opengeospatial.org/standards/wfs

Sinnott, R.O. Ajayi, O., Stell, A.J., Data Privacy by Design: Digital Infrastructures for Clinical Collaborations, *International Conference on Security and Privacy*, Orlando, USA, July 2009.

Watt, J., Sinnott, R.O., Jiang, J., Doherty, T., Higgins, C., Koutroumpas, M., Tool Support for Security-oriented Virtual Research Collaborations, *IEEE International Workshop on Security in e-Science and e-Research (ISSR-09)*, Chengdu, China, August 2009.

Stell, A.J., Sinnott, R.O., Ajayi, O., Jiang, J., Designing Privacy for a Scalable Electronic Healthcare Linkage System, *submitted to IEEE International Conference on Information Privacy, Security, Risk and Trust (PASSAT 2009)*, Vancouver, Canada, August 2009.

SAML, Security Assertion Markup Language (SAML), www.oasis-open.org/committees/security/, 2005

XACML, OASIS eXtensible Access Control Markup Language (XACML), www.oasis-open.org/committees/tc_home.php?wg_abbrev=xacml, 2005