# Security-oriented Portals for the Life Sciences

*R.O. Sinnott, T. Doherty, J. Jiang, S. McCafferty, A. Stell, J. Watt
National e-Science Centre
University of Glasgow
Glasgow G12 8QQ
r.sinnott@nesc.gla.ac.uk

**ABSTRACT**
**Motivation:** The life sciences are broad in scope and cover multi- and inter-disciplinary domains as well as the biological domain. These domains can for example involve researchers from the clinical, social, geo-spatial and computer sciences amongst others, e.g. in understanding genetic variations across a population as might be undertaken through a genome-wide association study. Given, this it is essential that portals for these communities are targeted to the individual expertise of the particular domain scientists. Thus tools available to a bioinformatician through a portal might well be meaningless to a social scientist and vice versa. Furthermore certain domains demand that fine-grained access control on data is supported. In this paper we outline how a portfolio of life science related projects at the National e-Science Centre (NeSC) at the University of Glasgow have benefited from security-oriented portals focused upon ease of access, configuration and usage, where data providers are assumed to be autonomous and able to make their own local fine-grained access control decisions. We describe the basic technologies that underlie these solutions and outline specific case studies in their application in the areas of depression, self-harm and suicide, and in the area of paediatric endocrinology focusing in particular on rare diseases associated with sex development.

## 1 INTRODUCTION

The life sciences have much to benefit from the application of e-Science and Grid based technologies: seamless access to large scale high performance computing (HPC) facilities; technologies and processes that help in coping with the explosion of data sets in high throughput post-genomic life sciences, and supporting multi- and inter-disciplinary research communities in tackling major research questions are some of the most obvious ones. Despite this, the impact envisaged in the application of e-Science and Grids in the life sciences domain has not been as major as first hoped for. There are a variety of reasons for this. These include:

- the complexity of the middleware that is used;
- the lack of clear understanding and expression of life science research community requirements;
- a moving set of life science research questions;
- a moving set of scientific data and understanding of that data;
- a perceived lack of security of Grids and issues this has on data access and release.

This list is not complete and there are doubtless other issues that could be brought to bear on life science community take-up. Some of these are described in more detail in [1]. In this paper we argue that many these issues can now be tackled. In particular we focus upon solutions that tackle several major uptake considerations: usability of the e-Infrastructures that are developed; support for inter-disciplinary research and security considerations for all protagonists involved in life science research. We claim that this is now possible through security-oriented portals targeted to the specific needs of life science researchers. This is demonstrated through two case studies in the area of public health and genetics of rare diseases.

The rest of this paper is structured as follows. Section 2 describes portal based technologies and how they have been applied at the National e-Science Centre (NeSC) at the University of Glasgow (www.nesc.gla.ac.uk). Section 3 describes technology support for secure access to and configuration of portals. Section 4 describes the application of these solutions to particular case studies. Finally we draw some conclusions on the work as a whole and outline areas of future development.

## 2 GRID PORTALS

Web portals provide a single point of access where a variety of information is aggregated and personalised to individuals to improve their experience in accessing and using a range of Internet resources. Common features of web portals include support for categorization of web content and advanced search facilities. Grid portals build upon the general web portal model to deliver the benefits of Grid computing to virtual communities of users, providing a single access point to Grid services and resources. Web 2.0 based solu-

*To whom correspondence should be addressed.

tions whether this be wikis, social networking capabilities, lightweight tools, e.g. for visualisation or mash-ups, can also be made available through portals.

The major difference between a Web portal and a Grid portal is that Grid portals provide a single point of access for Grid resources specific to a given domain, rather than more general Internet-based web pages or content. Grid portals provide end users with a customized view of software and hardware resources specific to their particular problem domain. This customization can be based upon the privileges that end users have. This can be used to restrict or authorize access to collections of remote services and data sets. Grid portals should ideally allow researchers to focus on their research problems by making the Grid a transparent extension of their desktop computing environment.

The development of targeted portals for the life sciences in the large, i.e. without focusing on particular specializations, offers a direct way in which a rich variety of applications and resources can be made available in a transparent manner to users who do not wish to become Grid experts. Given the depth and range of life science research currently being pursued, it is highly that a single one-stop shop portal could be established for all researchers.

That said, Grid-based life science portal solutions should meet the following set of requirements:

- *Usability* – the portal should be developed with both the experienced and inexperienced research communities in mind. This might benefit from use of backend servers to manage user certificates and Grid credentials required across the life science resources.
- *Single sign-on* – in addition to secure access to a Grid-based life science portal itself, seamless access to a range of life science resources without the need for multiple authentications should be supported. Access should, of course, depend on user privileges. The concept of single sign-on is one of the characterizing features of the Grid
- *Interoperability* – it should be possible for research communities to develop their own services using potentially different middleware on their own local resources, but be able to make these available to remote researchers through portal technologies.
- *Support for research* – it is essential that the services and data sets made available through the portals meet the real needs of life science researchers. Their input and feedback should drive the design and development of these portals and their content.
- *Support for collaboration* – Grid-based life science portals should facilitate collaboration between researchers at all levels – within an institution, between institutions, across national and international levels.
- *Portal administration and management* – user communities should be able to establish and ultimately manage their own services and their own user base. The shared resources underlying these communities need to be autonomous, however, and under the control of these communities.
- *Monitoring* – administrators and users should have direct access to monitoring information about various aspects of the Grid for their virtual organization, for their institution, and for those using the shared resources. This might include notifications of new data sets or new tools of interest to life science communities.
- Security – certain life science communities and data providers demand fine-grained security for tailoring access and usage of Grid resources. This might be based on specific roles particular to a virtual organization.

Whilst it is possible to develop hand-crafted portals, recent advances in this area have resulted in Grid portal frameworks which facilitate re-use of code and support various forms of structuring portal pages. Grid portal frameworks provide a set of basic functionalities and infrastructure for developing further portal components as plug-ins. Common components are offered for security (e.g. access management), for personalisation (e.g. user/group profiles), and for different presentation capabilities (e.g. JSP, XSP, XML/XSLT).

Portals themselves provide access to families of portlets or other hosted applications. Portlets are typically Java-based web components managed by a portlet container that processes requests and generates dynamic content. Portals use portlets as pluggable user interface components, providing a presentation or access layer to systems. Portlets support modular and user centric web applications. Portlets are the building blocks of portals and are typically small units of functionality within a portal. Each portlet typically provides an interface to a Grid service offering some well defined functionality. Users and administrators of communities or virtual organisations more generally can build customized environments by adding portlets.

As we shall see in section 3, for advanced scenarios, security techniques can be used to authorize use of particular portlets or the resources available to the services accessible via those portlets.

It is worth noting that to support portlet and portal interoperability, the portal community and wider industry have developed two key standards of relevance to the Grid community: the Java Portlet Specification (JSR-168) and the Web Service for Remote Portlets (WSRP). JSR-168 enables interoperability among portlets and portals. The specification defines the contract between a portlet and portlet container, and a set of portlet APIs that address personalization, presentation, and security. The specification also defines how to package portlets in portal applications. WSRP allows plug-and-play of content sources (portlets) within portals and other aggregating web applications. WSRP standardizes

the consumption of web services in portal front-ends, and the way in which content providers write web services for portals. This allows content producers to maintain control over the code that formats the presentation of their content. By reducing the cost for aggregators to access their content, WSRP improves the integration of content sources into pages for end users.

WSRP and JSR-168 are complementary specifications. JSR-168 defines a standard portlet API for Java-based portals. WSRP allows content to be hosted in the environment most suitable for its execution, while still being easily accessed by content aggregators. Second generation Grid portals can be produced from pluggable (JSR-168 compliant) Grid portlets. Running inside a portlet container, portlets can be added or removed, thus providing administrators with the ability to customize access and usage of Grid services at portal level. A portal built from Grid portlets can provide users with the ability to integrate services provided by different Grid-enabling technologies. This aspect is critical to the success of life sciences since a range of distributed services will likely be developed by different communities and institutions, and subsequently made accessible through common research specific portals (VREs).

The NeSC at Glasgow have developed a large portfolio of JSR-168 compliant portlets that are available in numerous portals supporting various life science research communities. Furthermore, through the Open Middleware Infrastructure Institute (OMII-UK) Security Portlets project (SPAM-GP) at NeSC Glasgow (www.nesc.gla.ac.uk/projects/omii-sp) we have also developed a collection of JSR-168 compliant security-oriented portlets that support simple, user-driven portal security based upon information provided by the Internet2 Shibboleth technology (http://shibboleth.internet2.edu) and the UK Access Management Federation (www.ukfederation.org.uk).

## 3 SECURITY INFRASTRUCTURES

As mentioned many communities are dissuaded from accessing and using Grid resources due to the complexity of the middleware and associated processes. Much of this stems from the demands to acquire and use X.509 based certificates as used to support the public key infrastructure (PKI) allowing single-sign on to distributed Grid resources. The issues with PKIs including their limitations are discussed in [2-4]. To overcome this, the UK and many international communities are moving to federated access control based upon the Internet2 Shibboleth technologies.

When a user attempts to access a Shibboleth protected service or Service Provider (SP) more generally, they are typically redirected to a WAYF server that exists as part of the federation that asks the user to pick their home Identity Provider (IdP) from a list of known and trusted sites. The service provider site already has a pre-established trust relationship with each home site, and trusts the home site to authenticate its users properly.

After the user has picked their home site, their browser is redirected to their site's authentication server, e.g. an LDAP repository, and the user is invited to log in. After successful authentication, the home site redirects the user back to the SP and the message carries a digitally signed Security Assertion Markup Language (SAML) authentication assertion message from the home site, asserting that the user has been successfully authenticated (or not!) by a particular means. The actual authentication mechanism used is specific to the IdP.

If the digital signature on the SAML authentication assertion is verified and the user has successfully authenticated themselves at their home site, then the SP has a trusted message providing it with a temporary pseudonym for the user (the handle), the location of the attribute authority at the IdP site and the service provider URL that the user was previously trying to access. The resource site then returns the handle to the IdP's attribute authority in a SAML attribute query message and is returned a signed SAML attribute assertion message. The Shibboleth trust model is that the target site trusts the IdP to manage each user's attributes correctly, in whatever way it wishes. So the returned SAML attribute assertion message, digitally signed by the origin, provides proof to the target that the authenticated user does have these attributes.

The attributes used in this assertion may then be used to authorise the user to access particular areas of the resource site. Once authenticated through Shibboleth, the notion of single sign-on is supported whereby a user may redirect their browser to other protected Shibboleth resources with no need for re-authentication.

Underlying Shibboleth-based SAML token exchanges are a core set of attributes based upon the eduPerson object class (www.educause.edu/eduperson/) that are pre-agreed across the federation so that an SP can make its own local access control decision. It is essential that interoperability exists between attribute authorities issuing attribute assertions, policy writers defining access policies, and access decision functions that make decisions based on the initiator's attributes and sites target and resource policy.

However given the fact that Grids can be used to establish e-Infrastructures and more security-oriented VOs, the requirement to have VO specific attributes defined and embedded in core eduPerson attributes are highly desirable. The most likely attribute for this purpose is the *eduPersonEntitlement* attribute. The *eduPersonEntitlement* attribute can utilise structured XML data representative of large scale Grid infrastructure users and IdPs. This might include the VO they are involved in, the roles that they might have in that VO etc.

The SPAM-GP exploits Shibboleth-based access to Grid portals and provides tooling that allows usage of Shibboleth information, including VO-specific attributes to be utilized. In particular SPAM-GP developed a family of JSR-168 portlets that support:

- scoped attribute management (SCAMP) which allows restricted and syntactically correct manipulation of the Shibboleth attribute acceptance policy, streamlining the subset of IdPs from whom a portal will accept user attributes across the federation.
- creation and usage of X509 attribute certificates (ACP) to allow distributed service providers to make their own local *authorisation* decisions when users attempt to invoke remote (protected) services;
- content configuration allowing dynamic configuration of portal content based on Shibboleth attributes and knowledge of available services. Once authenticated to a portal via Shibboleth, users are presented with a filtered view of available portlets (and hence access to a restricted set of services). This portlet has been targeted specifically to extend the GridSphere portal framework.

The detailed implementation of these portlets and the infrastructure that is required to support them is described in detail in [5]. Our focus here is to show how these portlets support life science researchers. We demonstrate this in the ESRC funded DAMES project and the EU FW7 EuroDSD projects.

## 4 CASE STUDIES

### 4.1 DAMES

The social sciences as with many other domains are awash with data. The ESRC funded DAMES project is developing systems that will help in tackling the problems facing this community. The DAMES project as a whole has a variety of themes. These include occupational data management; ethic and minority data management; educational data management and the one of concern here: e-Health data management. The project is exploring the data management challenges associated with access to a wide range of data sets cross multiple e-Health related disciplines. These include clinical data sets such as the Scottish Morbidity Records (SMR) covering hospital admissions, mental health and psychosis, cancer and birth and death related data sets; Census related data sets and geospatial data sets.

Scotland is especially well placed to support e-Health related research. Clinical information has been captured and curated for over 30 years in Scotland and an extensive record of the health of the Scottish population exists. The DAMES project has focused in particular upon data sets associated with mental health and is exploring the issues related to depression, self-harm and suicide. In particular it is attempting to answer questions related to the factors that can impact upon individuals suffering from some form of mental health problem. Thus what is the impact of living alone on suicide rates? What is the impact on suicide rates on access to prescription drugs? What is the impact on suicide rates on access to park land? There are in short a multitude of research questions that could potentially be answered if the appropriate data resources could be integrated and analysed.

To support these kinds of scenarios the various data sets have been acquired and made available through a targeted DAMES portal. Ideally the data sets themselves would remain at the remote data provider sites however in the first instance we have been given direct access to randomized copies of the associated clinical data sets. This was made possible through the previous MRC funded Virtual Organisations for Trials and Epidemiological Studies (VOTES – www.nesc.ac.uk/hub/projects/votes) project. The data sets are identical in structure to the actual SMR data sets, i.e. they have the same schema. However the data itself has been pseudo-anonymised to remove patient identifying information. Nevertheless much of this information can be used directly for building proof of concept systems.

To begin with a family of portlets has been developed that offers user interfaces to targeted data services. For simplicity, we associate a particular portlet with a given role as returned by the Shibboleth IdP. Using the content configuration portlet from SPAM-GP it is possible to associate possession of roles with associated portlets. It is of course feasible to extend this so that possession of a given role can be used to give access to a family of portlets and associated set of group permissions inside of the portal framework.

Originally the DAMES work was based upon the GridSphere technologies (www.gridsphere.org), however more recently we have targeted the LifeRay portal environment (www.liferay.com). This was primarily due to the lack of continued support for the GridSphere framework.

The portlets themselves allow targeted access to a subset of variables available in the particular data sets themselves. Thus in the case of the mental health related scenarios, these include Census variables associated with general health of the population; SMR variables associated with mental health including subsets of SMR01 (hospital admissions data); SMR04 (mental health and psychosis data) and SMR99 (death related variables where the cause of death was indicated as suicide).

To prove the conceptual approach of secure portlet based access to distributed data sets, we developed multiple different portlets for each data set and associated these with particular roles. In particular we identified advanced and basic roles. Thus to see the advanced portlets for SMR04 a user would have to be in possession of *DAMES_SMR04_adv*. The primary difference between the advanced and basic roles was in the variables that could be selected with ad-

vanced roles having a much greater set of variables that could be selected. The DAMES portal showing how it has been configured using the CCP for an individual user with advanced privileges is shown in Figure 1 (left) with the results of running a query shown in Figure 1 (right). This result data set shows the number of individuals who have committed suicide across Scotland who have been at least once to a hospital due to mental health related reasons. This result data set also includes geospatial information. This includes partial postcodes and/or output areas where the individual lived or was treated at that time.

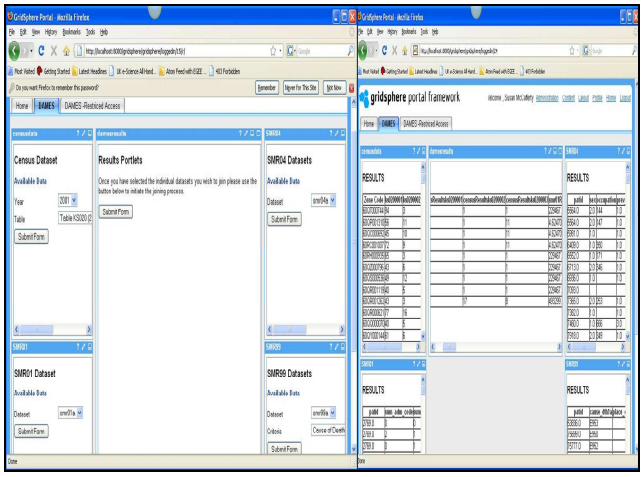


**Figure 1: Census & SMR Data Portlet Query Interfaces (left) and Returned/Joined Results (right)**

Key to this approach is in understanding the actual data sets themselves. That is, knowing what data sets in which data resources are related. At present the clinical data sets have a unique Community Health Index (CHI) number associated with them. This unique variable was rolled out for all individuals in receipt of health care across Scotland mid-2006. However, other data sets, e.g. the Census data, does not including the CHI. Nevertheless there are other data fields that are common, e.g. geospatial output area statistics.

Building upon this information, DAMES has developed further portlets that support overlaying this health related data across geospatial boundaries. To support this, the project has acquired shapemap files from EDINA (www.edina.ac.uk) which can subsequently be used for rendering and overlaying a variety of data across a given geospatial boundary. These shapefiles cover health authorities in Scotland and Scotland as a whole. The results of overlaying the above data sets across Scotland are shown in Figure 3. We note that access to these shapefiles themselves is restricted as they are under license to EDINA bu the commercial organization Ordnance Survey. Thus access is only for registered UK academics who have agreed to the terms and conditions to the license.

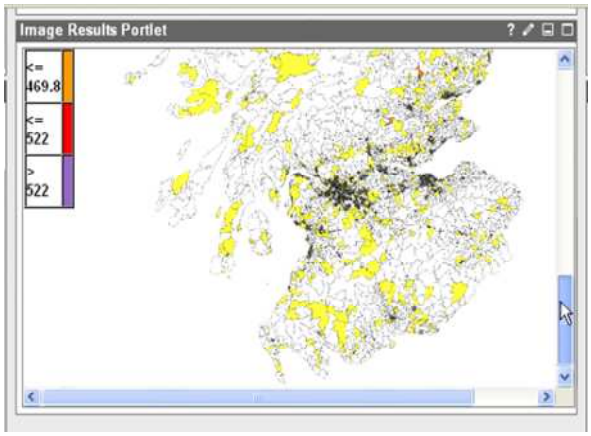


**Figure 2: Geospatial Visualisation of Suicide and Mental Health Data Sets Across Scotland based on representative SMR data**

To demonstrate the functionality of the SPAMP-GP ACP portlet, we specifically defined a data service that had its own local access and usage policy. Thus in reality, it is highly unlikely that any given data provider will simply delegate access to their data sets to a portal hosted at NeSC in Glasgow. To address this, a local policy was defined and enforced on access to the SMR99 data sets. In particular this service would only allow access to the SMR99 data variables for users with associated digitally signed and recognised attribute certificates. Users would thus use the ACP portlet to create an X509 attribute certificate based upon the information that was delivered to the portal through Shibboleth, i.e. the role that they need to use to access a particular remote data set (in this case DAMES_SMR04_adv) would need to be digitally signed and stored in an attribute authority that the remote data provider trusts. In this case, the attribute authority is an LDAP server associated with the DAMES portal port.

To support this process, the back end of the portal hosts a Grid credential repository – a MyProxy server. This is completely transparent to the user however. Thus the username and password to activate the credential in the MyProxy server is sent through as information as part of the Shibboleth SAML assertion. This credential repository is used at the individual invocation of a remote service. Thus the service needs to identify the individual user attempting to access the resource (authentication). Based on this they then need to ensure that the user has the appropriate privileges to run that particular query, i.e. pull the digitally signed attribute certificate from the attribute authority that there local policy recognises.

More information on the DAMES work is described in [6-7]. The project is currently applying for ethical permissions to access and use actual SMR data sets to realise the scenarios outlined previously.

### 4.2 EuroDSD

Disorders of sex development (DSD) are a set of disorders affecting the genito-urinary tract and, in a lot of instances, the endocrine-reproductive system. The physical representations of DSD are manifest in different ways. In the case of newborns this can be through children born with ambiguous genitalia; in the case of teenagers this can be in their lack of entry in to puberty. The condition is extremely rare and can be a fraught and potentially stigmatizing condition for the patients and families involved [8].

At the heart of much DSD research is the androgen receptor which is involved in controlling gene expression for genes associated with DSD and in determining the sex phenotype. In particular much DSD research is focused around androgen receptor co-factors that modulate sex steroid action [9].

Given the rarity of the condition, there is an associated scarcity of data on DSD. The EU funded EuroDSD project (www.eurodsd.eu) has been established to help improve the understanding and provision of therapy and potential treatments or interventions associated with DSD, and increase knowledge of the genetic and biochemical profiles which characterize DSD. Until now, European research and the data sets associated with DSD could best be categorized as being fragmented. There was no possibility to co-ordinate multiple, complementary DSD research areas in order to make a significant step forward in the understanding of DSD. Rather different centres and countries had their own fragmented support networks for patients and their families. The intention of EuroDSD is to bridge the gap between a variety of disciplines including clinical medicine, biochemistry, molecular genetics and molecular biology, through cross-national integration of resources and expertise. Currently, the EuroDSD project involves six European countries including the UK, France, Netherlands, Germany, Italy and Sweden.

The central component of the EuroDSD project is the development of a Virtual Research Environment (VRE), designed to facilitate secure, flexible collaboration using state of the art technologies. This technical work is led by the National e-Science Centre (NeSC – www.nesc.ac.uk) in Glasgow.

The EuroDSD VRE currently comprises a portal which hosts a registry which allows for secure upload and searching of clinical case information. This information is based on a common data model agreed through the European Society for Paediatric Endocrinology (ESPE), and in particular the work undertaken through the Scottish Genital Anomaly Network (SGAN - http://www.sgan.nhsscotland.com/). An example of this information is shown in Figure 3.

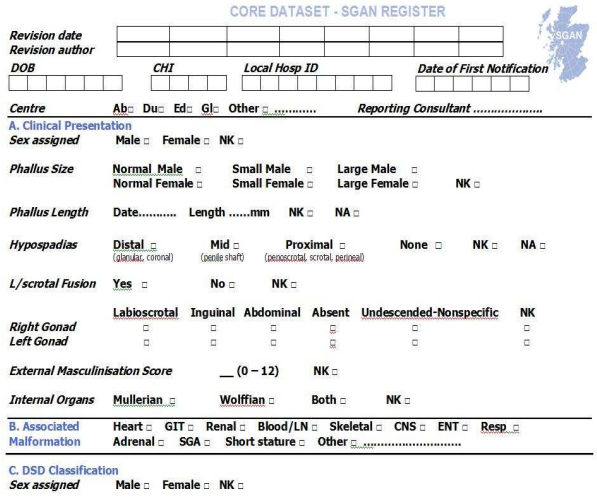


**Figure 3: Core data schema based on the SGAN template**

This information covers patient specific data including the referring clinician; the year of birth of the patient and specific clinical information and associated presentation information. For example, whether the karyotype has been established and if there are any further complications with DSD – often DSD has other physiological presentations, e.g. related to facial anomalies for example. The user interface to upload patient specific information into the registry is shown in Figure 4.

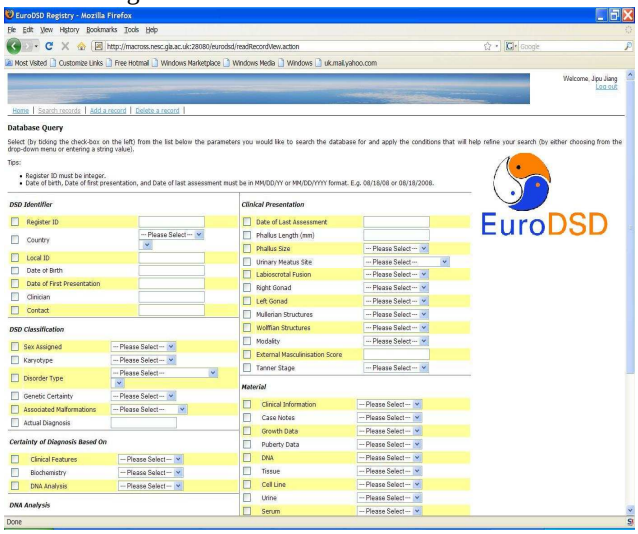


**Figure 4: Registry upload portlet making use of the Struts2 framework**

We note that the system and data models have been developed to be non-patient identifying. This has been a key part in the work as a whole. The only identifying information that exists in the registry is the contact details for the clinician. An optional local patient identifier can also be included, but once again the linkage of this with any given individual cannot be achieved without detailed local information on the patient itself. Indeed numerous other aspects

have been considered with regard to patient security. For example, the data of birth has been refined to year of birth only and exact birth weights have been refined to approximate weights in kilograms. All of these considerations were documented and put before ethics committees so that a decision on data collection could occur. This agreement has now happened and the data is being actively collected. At the time of writing the registry contains patient data on 267 cases from numerous centres across Europe. There are currently 40 registered users of the VRE from multiple sites. It should also be noted that non-EuroDSD partners are also now requesting access to the VRE to add their own clinical data. This includes researchers from countries outwith the European Union.

Each patient that is entered into the system has a unique registry identifier created and clinicians are expected to keep a local record of this in case of future follow up from collaborators, i.e. it is this information that can be used for record linkage.

The portal itself has been implemented using the Apache Struts2 (http://struts.apache.org/2.x/) framework; the Spring Framework (http://www.springsource.org/) and the Hibernate framework (http://www.hibernate.org/). For maximum flexibility, the portal has been based upon a multi-tiered architecture. The upper layer is the user/presentation layer and provides multiple ways to display information, e.g. HTML, JSP, AJAX, etc. Other layers are used to control targeted workflows and to process EuroDSD specific logic. Finally the lower layer is used to give access to the DSD data sets themselves and support for persistence itself.

In this system architecture, Struts2 acts as the controller of the VRE (aligned with the Model View Controller paradigm) and provides presentation of the data. The Spring framework provides a wide range of services for EuroDSD business and workflow logics. Finally Hibernate, at the lower layer, provides data persistence services to the VRE.

This solution significantly improved the overall flexibility, testability, and maintainability of the VRE. It was also compatible with most existing frameworks and standards, including JSR-168 portlets. Based on this n-tier architecture, business logic is properly stripped from the presentation and underlying data layers thus providing better maintainability and enhancements. This model also allows further flexibility for integration with further resources and tools.

To improve system availability, dynamic VRE configuration is support. Thus it is normally the case that changes of configuration will typically require the restart of the Web server containers hosting the services, and hence stopping the services themselves. However, EuroDSD VRE configuration is dynamic and can be loaded into the VRE through scripts and property files.

In terms of portal and VRE security, the Shibboleth system is fully integrated into the VRE and everything inside of the VRE runs through a Shibboleth-based authentication process. One issue we have faced with this, is that users at institutions that are not involved in a Access Management Federation have been offered a virtual home at NeSC-Glasgow. This model works and has up to now been found to scale reasonably well for local administrators. However when the VRE is used by 100-1000's of DSD researchers then this modus operandi may need to be revisited.

Based on consortium security requirements, each user has been assigned with a set of EuroDSD-specific roles, which are subsequently used for fine-grained access control. These include *EuroDSD_investigator*; *EuroDSD_contributor* and *EuroDSD_researcher*. These roles are subsequently used to define and enforce the privileges of the user inside of the portal. A user with the *EuroDSD_contributor* role has the privilege to upload, edit, and withdraw cases in the registry, whilst users assigned the *EuroDSD_researcher* role can only search patient data. EuroDSD_investigators are able to add, edit, delete and search data sets. In addition to these roles, the project also supports the scope of the search. Thus some sites are only able to add data for local centre users; others only able to add data for national data sharing; others for sharing across all of EuroDSD and others for all partners in EuroDSD and potentially with other international collaborators. The scope of data sharing is defined at data entry time by the contributors as part of the consent process, i.e. confirming that they have consent to add the patient data and the level of data sharing that the consent allows.

Every operation in the VRE (search, upload, edit, delete) is secure and implemented according to roles assigned to individuals. HTTPS is used to encrypt communications between user browsers and the VRE, and between the VRE to data sources. End users are completely oblivious of the underlying security used to enforce access and use of clinical data made available in the VRE.

The VRE itself has also been developed with data validation and usability aspects incorporated. Where possible, selection boxes are used instead of text fields. This simplifies data validation and ensures adherence with agreed data formats (aligned with the ESPE core data model) and avoiding data entry confusion. JavaScript is used in the user interface to provide data validation and automatic calculations when possible, e.g. calculating an External Masculinisation Score based on user selections of five clinical fields. Invalid data and actions are shown immediately when moving to next data entry fields or on request submission. AJAX technology is also used to provide coherent flow of information without the need for page refreshing to interrupt user experience.

To support discussion and feedback between the scientists involved in DSD, the VRE hosts a wiki. This wiki is directly linked with the registry itself. Through the wiki, scientists are able to directly comment on specific cases in the registry

and discuss aspects of the cases themselves or the treatment of the individuals for example.

The work on the EuroDSD VRE is still on-going. The current work is focused upon genetic analysis modules. These provide support for understanding the genetic variations of children with DSD. This includes the genes that have been screened for; the mutations/anomalies in genes that have been found; whether further analysis of these genes has been undertaken (including whether results have been or are being published), as well as new and upcoming screening and analysis methods, e.g. based upon mass spectroscopy. The genetics analysis module also typifies the structuring and standardization the EWuroDSD is attempting to achieve. The field is characterized by multiple genes with synonyms. To address this, the genetics analysis module has implemented a core list of widely accepted genes that the contributors can select. Thus textual editing has been avoided to avoid typographical mistakes as well as errors caused by upper/lower case genes representing different genes.

Further modules are also being explored. These include modules specific to the surgical treatments of patients with DSD. The classification of surgical treatments and the outcomes of these treatments are subject to international variation between partners.

For centres with a catalogue of DSD cases, the effort to manually enter these cases is off-putting. To address this we are supporting systems that allow bulk upload of patient data. The common technology that is used by most sites is Excel. Services have been implemented that allow Excel spreadsheets developed according to the ESPE core data model to be automatically incorporated into the registry.

Finally we note that some centres have their own local databases and resources. We are in the process of make these available through the EuroDSD VRE. These include patient steroid and metabolomic databases for example.

## 5 CONCLUSIONS AND FUTURE WORK

The work described in this paper has outlined how security-oriented portals can be supported that are aligned with the requirements of a wide range of life science researchers and communities. From experience of numerous projects in NeSC we are acutely aware that usability of e-Infrastructure has to be factored in from the outset. In this regard, Shibboleth-based access, security-oriented portal configuration and support for federated access control at remote, autonomous data providers have been found to offer a model that is closely aligned with many researcher user access requirements. However we also recognize that these models are often non-trivial for data providers such as the NHS. They are for example, extremely reluctant to open up their firewalls, irrespective of whether or not a service exists at their side that has associated authorisation policies. Given this, we are also exploring alternative data access and usage models. This includes the VANGUARD system [10] that has been demonstrated as a proof of concept system which will be refined and hardened as part of the Wellcome Trust funded Scottish Health Informatics Platform for Research (SHIP) project.

We also note that the work described here is also being applied in numerous other research domains. This includes the EPSRC funded nanoCMOS project (www.nanocmos.ac.uk) and the recently funded ENROLLER project (www.enroller.org.uk) and the NeISS project (www.neiss.org.uk).

## ACKNOWLEDGEMENTS

This work has been funded from a variety of sources including the UK Engineering and Physical Sciences Research Council (EPSRC), the Economic and Social Sciences Research Council (ESRC), the European Union, the Joint Information Systems Committee (JISC), the Medical Research Council (MRC) and the Wellcome Trust. We gratefully acknowledge their support.

## REFERENCES

[1] R. O. Sinnott, M. M. Bayer, J. Koetsier, A. J. Stell, Grid Infrastructures for Secure Access to and Use of Bioinformatics Data: Experiences from the BRIDGES Project, 1st International Conference on Availability, Reliability and Security, (ARES'06), Vienna, Austria, April, 2006.

[2] R.O. Sinnott, O. Ajayi, J. Jiang, A. J. Stell, J. Watt, User-oriented Security Supporting Inter-disciplinary Life Science Research across the Grid, New Generation Computing, Special Edition on Life Science Grids, editors A. Konagaya, P. Arzberger, T. W. Tan, R. Sinnott, D. Angulo, pp 339-354, Vol. 25 No. 4, 2007.

[3] R.O. Sinnott, Grid Security, National Centre for e-Social Science book, Grid Computing: Technology, Service and Application, CRC Press, May 2008.

[4] R.O. Sinnott, O. Ajayi, A.J. Stell, J. Watt, J. Jiang, User Oriented Access to Secure Biomedical Resources through the Grid, Life Science Grid Conference, Yokohama, Japan, October 2006.

[5] J. Watt, R.O. Sinnott, J. Jiang, T. Doherty, C. Higgins, M. Koutroumpas, 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.

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

[7] R.O. Sinnott, T. Doherty, C. Higgins, P. Lambert, S. McCafferty, A. Stell, K. J. Turner, J.P. Watt, Supporting Security-oriented, Inter-disciplinary Research: Crossing the Social, Clinical and Geospatial Domains, in Proceedings of International Conference on e-Social Science, Cologne, Germany, June 2009.

[8] O. Hiort, P.M. Holterhus, U. Thyen, *The basis of gender assignment in disorders of somatosexual differentiation*. Horm Res 64 (Suppl.2), 2005, 18-22.

[9] J.H. Bebermeier, J. Brooks S. DePrimo, R. Werner, U. Deppe, J. Demeter, O. Hiort, P.M. Holterhus. Cellline and tissue specific signatures of androgen receptor coregulator transcription. J Mol Med, 2006, 84: 919-31.

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