Security Oriented e-Infrastructures Supporting Neurological Research and Clinical Trials

Anthony Stell, Richard Sinnott, Oluwafemi Ajayi, Jipu Jiang
National e-Science Centre
University of Glasgow
Glasgow, UK

Abstract - The neurological and wider clinical domains stand to gain greatly from the vision of the Grid in providing seamless yet secure access to distributed, heterogeneous computational resources and data sets. Whilst a wealth of clinical data exists within local, regional and national healthcare boundaries, access to and usage of these data sets demands that fine grained security is supported and subsequently enforced. This paper explores the security challenges of the e-Health domain, focusing in particular on authorization. The context of these explorations is the MRC funded VOTES (Virtual Organisations for Trials and Epidemiological Studies) and the JISC funded GLASS (Glasgow early adoption of Shibboleth project) which are developing Grid infrastructures for clinical trials with case studies in the brain trauma domain.

Index terms – Grid-based authorization, clinical data federation

I. INTRODUCTION

The neurological and wider healthcare domain has an abundance of data spread across many regions. The scientific and medical value of this data has the potential to be significantly enhanced, if it can be accessed and used across these boundaries. It is currently largely the case however that these data sets exist in silos and cannot be linked even within the healthcare domain with primary and secondary care often driven by non-electronic (paper-based) information exchange.

Grid technology "in principle" provides an infrastructure that would allow federation of data to occur. However, Grid technologies and standards will only be adopted by the health domain if they can visibly demonstrate that they address all potential security concerns of patients, clinicians, IT support staff and ethical bodies. It is essential that any security infrastructure makes available the right data sets to the right people for the right purpose.

In the clinical domain, many IT and data storage solutions already exist for individual GP practises, disease registries, clinical repositories and hospitals. Most of these solutions have been developed in isolation and as a result have widely differing data descriptions and associated security policies. As such, any solution developed to meet the problem described above must meet the idiosyncrasies of these heterogeneous sources. There are however compelling reasons for controlled sharing of these data sets. The National Health Service (NHS) is not focused upon keeping data private or inaccessible, but in delivering health care. Sharing of information is an essential part of this.

The VOTES project (Virtual Organisations for Trials and Epidemiological Studies) [1] is exploring this space, specifically with regard to clinical trials and epidemiological studies. The VOTES project is addressing three key areas of clinical trials, which stand to benefit most from the development of an underlying grid infrastructure:

- *Patient Recruitment* – asking specific demographic and/or clinical questions combined with the individual patient histories which can subsequently be used for identifying patients willing to participate (and give consent for their data to be used) in such a trial, or for conducting trial feasibility studies for example.
- *Data Collection* – typically involves following up on the data produced by a clinical trial. Making sure procedures are being followed according to an agreed protocol, and most importantly, recording and analysing the details of any incidents that have occurred as a result of the trial.
- *Study Management* – monitoring the overall running of a trial and making

sure that the interests of the patient are being maintained, and that the trial is being run according to ethical procedures. This may also involve monitoring and identifying ways to improve the efficacy of the conduct of a trial.

By federating the data pertaining to these various areas from separate domains throughout the country and the world, it is believed that greater and more effective statistical and scientific insight into the results of such trials can be obtained and disseminated.

One example of a needed data federation which is explored within this paper is the brain trauma domain. The BrainIT project [2] is a European wide network looking at brain trauma in particular it focuses on understanding brain trauma, and especially how patients are monitored and treated as a whole. This domain has a rich range of heterogeneous data sets which need to be seamlessly linked including MRI images and physiological data sets amongst others.

At present the network send their data through to a centralised data repository at Glasgow Southern General Hospital. However this has numerous disadvantages. We explore these in this paper and present a model whereby fine grained security can be used to allow sites to maintain their own data sets and define and enforce their own security policies on access and usage.

The rest of the paper is structured as follows. Section 2 provides an overview of Grid security in the large and the challenges of the health domain. Section 3 focuses upon the approaches taken within the VOTES project to address these challenges and section 4 focuses on the implementation work with focus in particular on the brain trauma domain.

II. AUTHORIZATION IN THE CLINICAL DOMAIN

Due to the sensitivity of data in the clinical/health domain, the establishment of security policies and their rigorous enforcement is of paramount importance. However, due to the variety of scenarios and ranges of different clinical trials and studies requirements each with their own particular data or security demands and nuances, there is a need for a flexible solution. In short, we do not wish to build a single static system accessing a closed/fixed set of data, but rather we need to build infrastructures where a variety of data accessible to different individuals for different times for different reasons is supported. In any of these systems however fine-grained access to individual data sets is essential.

As an example, when collating data for statistical purposes, e.g. to check for the feasibility of conducting a given trial in a given area, it may be necessary for users with low privileges to view broad demographical data. However, it may also be necessary for users of higher privilege to view more personal data, and in some cases, data that will positively identify patients. In a given institution or hospital, this might be achievable, however when multiple institutions with numerous stakeholders are involved this becomes a much more fraught process. Extensible systems that support the secure establishment of primary and secondary care systems to more quickly and efficiently provide immediate or enhanced medical care, across a broad range of medical areas and studies are highly desirable.

It is an unavoidable fact however that the specific privileges required in a trial will not be known when the system is first created. Similarly a doctor in one hospital may have privilege to access various systems in that hospital, but these do not transfer directly when this doctor attends a different hospital for example. Therefore a dynamic architecture capable of adding and removing resources *"on the fly"* is necessary, where the corresponding allocation of privileges can be added or removed depending on the needs of different trials or healthcare systems. Grid-based solutions offer one possibility to address this.

A. Grid Security

The basic tenets of Grid-security can be broadly broken down into the "AAA" categories:

Authentication – establishing the identity of the person requesting access to a resource.

Authorization – having established identity, establishing and enforcing what that person is allowed to do on a given resource.

Accountability – being able to establish the activities, and time of activities, of a particular person on that resource (or resources) so that they cannot subsequently deny potential misuse later on (non-repudiation). Though this

paper will not discuss this aspect in depth, it is still an important consideration in the security of any system.

Of course there are other important aspects when considering the wider challenges of building secure systems, but in this paper we restrict ourselves to authentication and authorisation as these are arguably the most important things to get right in the first instance. Put another way if authentication and authorisation are not adequately addressed, then other aspects of security are largely redundant

Methods of authentication with grid technology tend to favour two methods, either username/password combinations or public key infrastructures (PKIs) to set up and use a safer, encrypted communication channels through trusting a third party root of trust Certification Authority. Authentication is focused upon identity management hence a process exists through which a user establishes their identity in the process of obtaining their X.509 digital certificate.

Typically this is through showing some form of physical identification to a local registration authority at their institution. However there are issues with this process, not least of which is the complexity of converting certificates to Grid formats and the lack of local monitoring. Thus a user might be expelled from their institution but still have access to a valid Grid certificate.

One middleware solution that is being widely touted as the answer to this issue is that of Shibboleth [3]. Shibboleth provides a method of securely transferring attributes between institutions subscribing to an over-arching federation. The basic model of Shibboleth is that users attempting to access a remote resource are redirected to their home institution (typically through a Where Are You From service) where they log in locally. A digitally signed SAML assertion showing that the user has authenticated is then delivered to the target resource which may then decide whether access is granted or not. Often further information such as attributes for authorisation need to be returned.

This whole process however is transparent to the end users who only log in to their local system with the normal usernames and passwords. This aspect should not be ignored. Usability of secure systems is one of the primary causes of security incidents. Strong password policies and software that ensures passwords are of sufficient strength are made redundant by users who write them down. (The UK Certificate Authority currently mandates that private key passwords are 16 characters long and made up of upper/low alpha and non-alphanumeric characters. As such it has been documented [18, 19] that users write their passwords down or share them with colleagues.

The key benefit of Shibboleth is that end users have simple ways to access resources. Furthermore depending upon local policies and trust relationships across the federation, users are able to access a range of distributed resources thereby supporting *single sign-on*. Once a secure session between a user and a resource is established, the user can access further resources (in the same federation) without the need to further authenticate – based upon browser-based information. Additionally since their authentication is tied to their local institution, they will have their federation privileges revoked, if they are revoked locally (which is the most likely scenario).

Authorization which typically involves access to remote resources (or vice versa) remains a far more intractable issue than authentication. Many technologies have attempted to address the problem of remote privilege allocation and enforcement, but none provide a single package that meets all the needs of most distributed systems.

One authorization solution that is gaining acceptance in the academic community is PERMIS (Privilege and Role Management Infrastructure Standards Validation) [4]. This is middleware package implements a policy engine to act upon hierarchical role based access control policies encoded in XML.

The application is based on the X.812 | ISO 10181-3 Access Control framework [5], depicted in Figure 1, providing a policy decision and enforcement point between the initiator and the target.

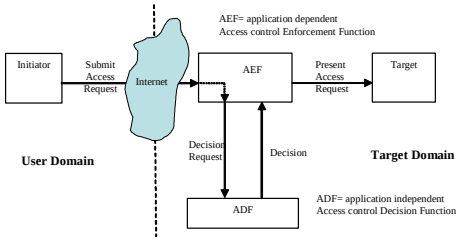

Figure 1: overview of X.812 Access Control Function.

PERMIS has been coded to plug in to the Globus Toolkit Grid middleware [9], protecting grid services by the use of the Security Assertion Markup Language (SAML) protocol [17]. As services protected in this way necessarily use Globus constructs and packages, there is only a limited number of ways that the application can be used.

Another issue is that the application is also designed to be used on a "per-service" basis - i.e. the granularity of authorization is such that only distinct services can be individually secured. In the field of data federation, a finer-grained method of authorization is required and hence the PERMIS solution is not ideal for providing the flexibility needed in this field.

Another example of such authorization applications is that of VOMS (Virtual Organisations Management System) [6]. The VOMS system implements a series of grid-map files – basic authorization files used by the Globus Toolkit that match user DNs (Distinguished Names) to services – to provide greater flexibility in establishing a secure authorization protocol.

However, the system does not implement a decision enforcement engine (c.f. the policy engine in PERMIS) and is also strongly-bound to the middleware applications developed for the particle physics community. In terms of implementation, this makes its integration with software for other scientific areas, with other application interfaces, notoriously difficult.

At the other end of the authorization process from these VO-wide technologies – the authorization methods of the local database resources – it is actually possible to set very fine-grained security policies that allocate privileges of individual data fields to the various roles. However, in the context of distributed grid systems, the problem is that this model does not scale up to the level of large scale dynamic virtual organisations. By the same token, the administrator of the local resource must have the final say as to what roles can and cannot get access to their resource.

There is also the issue of trust between entities at a higher level of abstraction (captured in the VOTES project as the establishment of site-to-site trust). The central issue is the method of establishing a trust relationship between two entities that only indirectly trust each other by means of trusting a common third party. This issue is not central to this paper but is still a consideration and can be referenced at [7].

Any effective security solution to be applied to the field of distributed data federation in the health domain, must address both of these paradigms. On the one hand, the local resource administrator must find a way to translate their local security policies to a common interface that can be understood by remote users as well as local. On the other hand, there must be a way for the remote users to gain fine-grained and secure access to individual data fields and parameters. One dichotomy with this is with regard to how such information is made available, since in general security information is not typically published and this is especially so in the clinical domain.

III. VOTES SECURITY

The VOTES project is an attempt to realise a software solution to the challenges inherent in federating distributed clinical data. The Grid enabling of the BrainIT project being undertaken within the GLASS project is a similar endeavour that deals directly with secure, remote access to distributed medical data specific to the neurological domain. As the two projects have similar goals and challenges the solutions have dove-tailed into the one application, the architectural details of which can be found in section 4. This section focuses on the security theory and outlines the solution that has been incorporated into the design of the VOTES portal.

A. Portal Overview

Usability is at the heart of VOTES efforts and portal based solutions have been prototyped with this in mind. This should be usable for end users, administrators, investigators involved in clinical trials. To support this, a node infrastructure that can be replicated to add or remove various institutions that wish to contribute or query clinical data by becoming a member of the VO. Figure 2 shows an overall schematic representation of such a VO:

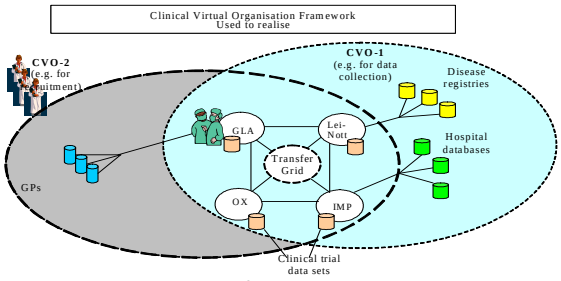

Figure 2: A diagram of the envisaged clinical virtual organisation (CVO) between the VOTES partners.

Each node is populated with a replicated and mirrored portal, grid and data servers. Through these components, queries are sent to a local pool of databases, guarded by a driving database, which joins the data to present a seemingly unified resource back to the end user. Figure 7 (in section 4) shows the architecture of the Glasgow node of this virtual organisation, with these components all marked.

These nodes are to be connected by way of the data servers, providing seamless access to a rich range of clinical data sets. Subject to the establishment of trust between two nodes, the exchange of data can be readily available to the entire VO, once the necessary authorization privilege files are appropriately set.

The basic authorization is based on the roles assigned to the user. A privileged user will be allowed to query a wider set of parameters on the various resources than a less privileged one. Figure 3 shows the different parameter sets available for querying for the "investigator" and "nurse" roles that are attempting to run queries on the same trial (labelled "votes2"). Ultimately these roles represent levels of privileges – which in this case correspond to the data sets that are accessible. Defining the relationships between a role and a view of a federated set of data is essential in this domain and supported within VOTES.

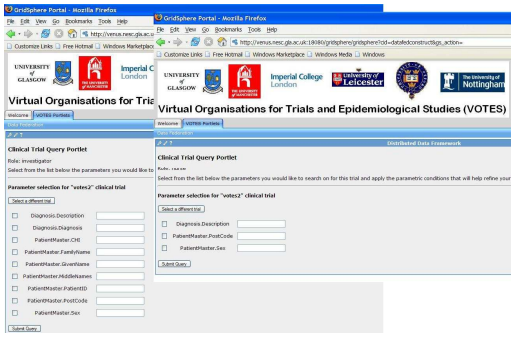

Figure 3: the left screen shows the parameters available for querying to the investigator role; the right shows those available to the nurse. Note how the latter are less in number and, critically, do not contain any identifying data.

B. Authorization and Administration

No Grid authorisation middleware today provides sufficient granularity to adequately express and enforce fine grained authorisation policies across large scale federated, heterogeneous sets of clinical data. To address this, the VOTES project has adopted a security solution based upon an Access Matrix model. Figure 4 shows a conceptual model of the access matrix: essentially, it comprises several tables, one for each database resource (or more precisely the schema for the database resource), which lists fully-qualified parameter names within that database, against the roles of the trial. A "1" digit signifies that that role will have access to that parameter, a "0" digit means that access to that parameter will be denied to that role.

	R_1	R_2	R_3	R_4
U_1		h_1	h_2	h_3	h_n
U_2	U_1	0	0	1	0
U_3	U_2	0	0	0	1
U_4	U_3	1	1	1	1
	U_4	0	1	0	0

Figure 4: The access matrix model. For each data source, the role versus parameter value shows the different privileges.

The central idea is that this access matrix will be available at every node in the VO and will be regularly updated, as the nature of any VO is transient, with resources potentially changing rapidly over a short period of time. Consequently, every user that has access to the VO in some form will go through this matrix model to access any other resource, be it local or remote, within the VO.

As VOs are assumed to be transient and dynamic in nature, a method for dynamically updating this access matrix is mandatory. As such an administrative portal has been incorporated into the overall design of the portal infrastructure.

This portal provides a method for privileged super-users to set policies. As this provides a potentially critical security hole in the system, it is crucial that the individual duties of these super-users are clearly delineated, and that the access to the execution of these duties are rigorously secured.

The individual duties for the super-users have been identified as being in two categories:

Node Administrator – this super-user is responsible for publishing what local resources they have. They are responsible for uploading their local resource and connection details to local files accessible to their local portal. They are also responsible for querying the other nodes for their resources, and adding any new remote resources to their local CVO database.

This intermediate validation by the node administrator hampers the dynamic nature of the VO but only to the degree that common sense decrees. There must be some human

intervention to validate the addition of remote resources, though there is a case here that if the other local node administrators publish their local connection details, they can be trusted. The issue here is one of binding legal agreements within the VO and is discussed below.

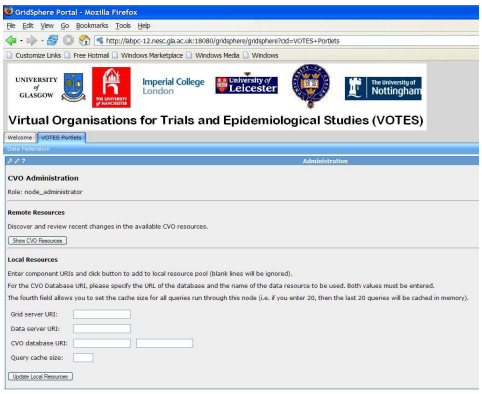

Figure 5: a snapshot of the node administrator home page. The duties are separated into local and remote resource administration.

This role essentially maintains the underlying infrastructure upon which the retrieval of clinical data can be run.

Trial Administrator – this super-user is envisaged to be a clinician more familiar with the running of clinical trials than with the underlying computing infrastructure. They are responsible for the definition and assignment of security policies based on their knowledge of the clinical trial and the data sets been made available by the node administrators.

A typical scenario is that upon login the trial administrator will attempt to create a new trial with a new name, and will be able to search the database resources available to the VO. One challenge we face here is abstracting from the primary database contents and presenting the information in a human-readable form. Once they have obtained the necessary resources. subject to their availability within the VO, they will then be able to create roles for that particular trial and assign access to the various parameters that the databases offer (this will also be subject to the discretion of what the node administrator for that resource has chosen to release for remote access).

C. Static Legal Considerations

Any solution provided must be subject to the approval of an ethical review board and as such, it is not possible to automate the entire procedure of remote federating access to such sensitive data. The example of patient consent for the recruitment of subjects for a clinical trial is one example where human intervention is mandatory before the technology can automate the rest of the process.

In the security scenario described above, there is a need to establish a static legal context between all the nodes within the VO, which establishes a base-level of trust between the parties. From the delineation of super-user duties described above it is clear that a node administrator from one node is potentially granting the power of privilege allocation on their resource to a trial administrator at another node. To do this requires a high level of trust between the acting parties, which in turn will require a legal form of recourse should either party feel that their grant of administrative rights has been abused.

Therefore, a static element to this dynamic system is unavoidable (and indeed necessary). The question then becomes what interactions between the node super-users are delineated in the static agreement and what can be assumed at run-time.

IV. IMPLEMENTATION

The technical implementation of the VOTES portal is based on the architecture shown in Figure 7.

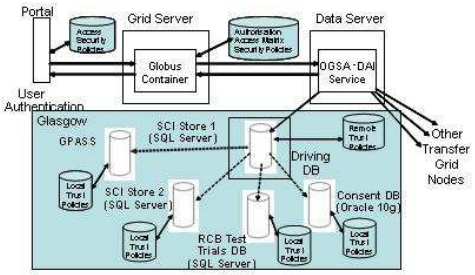

Figure 7: architecture of a single node within the CVO.

The basic operation is as follows:

- The user logs into the portal at a particular node - this can be directly or via Shibboleth (not shown above).
- The portal server checks the local resource files to discover the available grid servers, data servers and driving databases.
- A request containing the user's role and a specific trial is sent to the grid server.
- The grid server consults the local access matrix and returns the parameters for the resources that the user can query for that trial. These are

presented to the user as a list of check-boxes, with the option to specify conditions if desired.

- The user makes their selection of parameters and submits them. These are constructed into a single query distributed across the various resources.
- The query is sent from the grid server to the data server, where it is wrapped as an OGSA-DAI service request and is passed to the driving database.
- The driving database executes the distributed query over the resources under its guard and joins the various distributed results into one single result.
- This result is sent back to the data server and then to the grid server, where it is transformed into readable HTML and finally presented to the user through the portal again.

The various grid technologies used in this infrastructure are: GridSphere [8] – the portal server is designed to give secure and user-friendly access specifically for grid services; Globus Toolkit v4.0 [9] – the grid server uses the Axis [16] component of this package to implement stateful grid services passing between components the authorization information and the actual data queries; OGSA-DAI [10] – the data server is based on the OGSA-DAI technology and allows to present database resources in the form of a service-oriented architecture using XML.

The databases exist on a test infrastructure at the National e-Science Centre in Glasgow, comprised of variations of SQL Server, MySQL and Oracle. They contain representative, but "dummy", data-sets that use the actual software and schemas that are currently in use by the majority of systems in the NHS in Scotland. These include: GPASS [11] – this is the main software application used by primary healthcare specialists across Scotland; SCI Store [12] – this is a centralised repository of clinical information across Scotland, that is routinely updated by participating practises; Scottish Morbidity Records [13] – this is another centralised data set that records all hospital visits throughout Scotland.

In combination with the BrainIT project the portal has been specifically extended to incorporate brain MRI scans and a variety of physiological data from the Southern General Hospital in Glasgow. Figure 8 shows snapshots of the images, and the extended data repository accessible to authorised users.

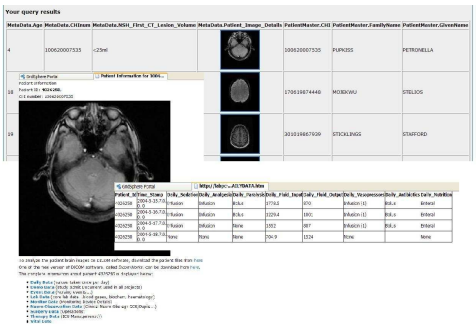

Figure 8: the extended neurological data incorporated into the VOTES portal. The patient data is at the back, the close-up of the image is returned when the original image is clicked upon, and a sample of the in-depth lab data for this patient can be seen on top.

As discussed, security in this infrastructure is implemented on two levels: at the VO level, through the use of the access matrix, and at the local resource level, through the individual database management systems. Through the combination of these paradigms it is possible to marry the necessity of local administrators retaining control of their resources and this control being scaled up to the VO-level.

A further extension of the portal has been to incorporate the Shibboleth model of distributed authentication (see section 2). This allows access to the Glasgow node by a user from a remote site that is subscribed to the same federation as the node. This significantly widens the scope of access to the VOTES resources, whilst maintaining effective security.

Figure 9: the Shibboleth attributes relevant to the user "guest2". Note how the user has different roles for different trials, with varying levels of privilege.

In an attempt to make the portal more user-friendly the implementation has inserted the roles for the different trials into the Shibboleth role database. Through results of the Dynamic Virtual Organisations for e-Science Education (DyVOSE) project [20] we are now able to

push attributes to remote (trusted) authorities for allocation to their own users which can subsequently be used for local authorisation decisions [21, 22]. This overcomes the largely static model of security upon which Shibboleth federations are based. In the UK a core set of eduPerson attributes have been agreed for UK–wide Shibboleth federations. Extending this to support more dynamic VO-specific roles and attributes is essential

V. CONCLUSION

Several security and authorization solutions already exist in the grid community, but none entirely meet the specific requirements of the clinical and neurological domains. As such, other innovative solutions are required to make distributed data federation systems in these domains a viable reality. In this paper we have presented one possible solution from the VOTES project.

The specific challenges and issues of the clinical domain highlighted in this paper are crucial before further progress in this field can be made. It is clear that a flexible solution is required that will allow access to individual patient data fields, on a "pick and choose" basis, whilst maintaining rigorous levels of access control definition and enforcement.

It is also clear that a secure yet readily accessible administrative access point is required for any flexible solution. The development of an administrative portal in the VOTES project is a step towards this goal but the delineation of duties between the different types of administrator is an issue that still needs more discussion and definition.

The issues tackled in this paper are popular topics of research and other efforts are being made as reflected in the recent call for proposals by the Open Middleware Infrastructure Institute (OMII) for a research project investigating the viability of portals used to define and implement access control policies for generic systems [14].

It is also the case that any of the innovative solutions presented will almost certainly build on the theories of the currently available software, and could perhaps even use modular parts of the packages themselves. For instance, the DyVOSE Delegation Issuing Service (DIS) [15] allows delegation of credentials between two institutions that only have a limited degree of trust between them. Using this technology it is possible to exert access control on remote users through local policies, without having to scale up to an aggregated authorization policy for the entire VO.

The VOTES and BrainIT projects are primarily pioneering, investigative projects. The solutions presented here are only possible scenarios that serve to highlight the challenges and issues as much as providing a production level answer to the questions posed. However, before any national or global clinical infrastructures are created a thorough exploration and demonstration of viable systems are needed. This work is laying the foundation for such future infrastructures with future Scottish-wide efforts looking to adopt these efforts to provide production level facilities for e-Health more generally.

VI. REFERENCES

[1] Virtual Organisations for Trials and Epidemiological Studies (VOTES) – http://www.nesc.ac.uk/hub/pojects/votes
[2] Glasgow Early Adoption of Shibboleth project (GLASS) – http://www.nesc.ac.uk/hub/projects/glass
[3] Shibboleth – http://shibboleth.internet2.edu
[4] PERMIS – http://sec.isi.salford.ac.uk/issrg/permis
[5] ITU-T Rec X.812 (1995) | ISO/IEC 10181-3:1996, Security frameworks for open systems: Access control framework
[6] VOMS Architecture, European Datagrid Authorization Working Group, 5 September 2002
[7] O. Ajayi, R.O. Sinnott, A.J. Stell, Trust Realisation in Collaborative Clinical Trials Systems, to appear in HealthCare Conference HC2007, Harrogate, England, 19-21 March, 2007
[8] GridSphere – http://www.gridsphere.org
[9] Globus – http://www.globus.org
[10] OGSA-DAI – http://www.ogsadai.org.uk
[11] GPASS – http://www.gpass.co.uk
[12] SCI Store - http://www.show.scot.nhs.uk/sci/products/store
[13] Scottish Morbidity Records - http://www.show.scot.nhs.uk/indicators/SMR/Main.htm
[14] OMII-UK Call for proposals in Portlet Development and Integration, November 2006
[15] Supporting Decentralized, Security focused Dynamic Virtual Organizations across the Grid Sinnott,R.O Proceedings of eScience 2006 (http://www.escience-meeting.org/eScience2006/)
[16] Apache Axis – http://ws.apache.org/axis2
[17] Security Assertion Markup Language – http://www.oasis-open.org/committees/security
[18] B. Beckles et al. – A user-friendly approach to computational grid security, e-Science All-Hands Meeting, Nottingham, UK, 2006
[19] R.O. Sinnott - Grid Security Report, http://www.nesc.ac.uk/hub/projects/GridSecurityReport
[20] Dynamic Virtual Organisations in eScience Education (DyVOSE) – http://www.nesc.ac.uk/hub/projects/dyvose
[21] R.O. Sinnott, J. Watt, J. Jiang, O. Ajayi – Shibboleth-based Access to and Usage of Grid Resources, IEEE International Conference on Grid Computing, Barcelona, Spain, 2006
[22] R.O. Sinnott, J. Watt, J. Jiang, A.J. Stell, O.Ajayi – Single Sign-on and Authorization for Dynamic Virtual Organisations, PRO-VE 2006, Helsinki, Finland, 2006