Supporting UK-wide e-Clinical Trials and Studies

Anthony Stell
National e-Science Centre
University of Glasgow
Glasgow, United Kingdom
+44 (0) 141 330 8648
a.stell@nesc.gla.ac.uk

Richard Sinnott
National e-Science Centre
University of Glasgow
Glasgow, United Kingdom
+44 (0) 141 330 8606
r.sinnott@nesc.gla.ac.uk

Oluwafemi Ajayi
National e-Science Centre
University of Glasgow
Glasgow, United Kingdom
+44 (0) 141 330 2958
o.ajayi@nesc.gla.ac.uk

ABSTRACT
As clinical trials and epidemiological studies become increasingly large, covering wider (national) geographical areas and involving ever broader populations, the need to provide an information management infrastructure that can support such endeavours is essential. A wealth of clinical data now exists at varying levels of care (primary care, secondary care, etc.). Simple, secure access to such data would greatly benefit the key processes involved in clinical trials and epidemiological studies: patient recruitment, data collection and study management. The Grid paradigm provides one model for seamless access to such data and support of these processes.

The VOTES project (Virtual Organisations for Trials and Epidemiological Studies) is a collaboration between several UK institutions to implement a generic framework that effectively leverages the available health-care information across the UK to support more efficient gathering and processing of trial information. The structure of the information available in the health-care domain in the UK itself varies broadly in-line with the national boundaries of the constituent states (England, Scotland, Wales and Northern Ireland). Technologies must address these political boundaries and the impact these boundaries have in terms of for example, information governance, policies, and of course large-scale heterogeneous distribution of the data sets themselves.

This paper outlines the methodology in implementing the framework between three specific data sources that serve as useful case studies: Scottish data from the Scottish Care Information (SCI) Store data repository, data on the General Practice Research Database (GPRD) diabetes trial at Imperial College London, and benign prostate hypoplasia (BPH) data from the University of Nottingham. The design, implementation and wider research issues are discussed along with the technological challenges encountered in the project in the application of Grid technologies.

Keywords
Clinical trials, distributed infrastructures and security, data grids

Mardi Gras Conference '08, Jan 31st – Feb 2nd , 2008, Baton Rouge, Louisiana, USA.

1. INTRODUCTION
The political structure of the United Kingdom (UK) provides a unique and highly relevant case study for some of the issues inherent in conducting any kind of population sampling – including clinical trials – across a broad spectrum. The political state of the UK is made up of several semi-autonomous "nations" – England, Scotland, Wales and Northern Ireland – each having a strong sense of national identity built up throughout history[1]. Around this sense of identity, much infrastructure has been built: parliaments with powers devolved from the central administration at Westminster were granted to Scotland and Wales in 1997. More recently, a victory in the Scottish elections for the Scottish National Party could suggest that the population of Scotland are progressively moving towards the idea of a nation fully independent from the rest of the UK.

As the information age progresses, it is not unreasonable to assume that the infrastructure to support various types of data storage and transfer would progress along the same lines and indeed that is the case. Data within one domain or region provides a certain amount of information. But it is common sense to think that the more data can be linked, the more the value of that data can be enhanced. In the case of health records, the UK scenario is particularly relevant – a lot of primary and secondary care information, relating to the medical history of patients will be of use to clinicians from both sides of the Scotland-England border. Due to this political structure, population migration between the two nations is relatively easy. Yet there are two very distinct and different health infrastructures, which the patients have to be registered and processed in.

An example of this diversity is that the Scottish health infrastructure is indexed upon a value known as the Community Health Index (CHI) number. This value has no meaning in the National Health Service (NHS) infrastructure in England, so for a patient requiring treatment in Scotland, who had only previously resided in England, a history would be required but it could not be searched upon this unique index. In an immediate primary care situation, a patient's life could, in extreme situations, hang in the balance based on this fact.

[1]For the purposes of this project England and Scotland have been focused on primarily, largely due to the locations and resources of the collaborating partners.

In secondary care, the issue is more subtle but also more relevant – many trials wish to recruit participants, and a natural part of any campaign that attempts to sample a population is that the wider the net is cast, the more positive returns are statistically likely to be received. A trial that could ask questions, such as medical history, specific conditions or specific treatments for the patient, over a greater subset of the population (or in cases of specific conditions, a more targeted area) would likely be more successful in recruiting eligible participants. Infrastructures that facilitate this process offer a step change in the progression of clinical trial methodologies from largely paper based human resource intensive activities, to more automated *e-Clinical* trials and studies.

Issues arising from the situation described above are exactly what grid technologies attempt to provide solutions to. Whilst maintaining the security and usability of a certain application, the data may be harnessed from many different resources, which may or may not have similar underlying data classifications (dictionaries/ontologies), and may have different access possibilities based upon different security infrastructures realising different information governance policies. By construction of such a security-oriented "data grid" useful links can be made between disparate infrastructures.

2. CURRENT UK HEALTH INFRASTRUCTURES

One of the paradigms of Grid technology is that it must be able to harness and leverage existing data storage/access technologies – presenting them as a single unified resource to the user, but with the additional enhanced value of the data. It is essential therefore that the technological infrastructures already in place in the health services that VOTES is attempting to work with are analysed.

2.1 Scotland

The main health-care information technology systems currently used in Scotland are SCI Store [1] and General Practice Administration System for Scotland (GPASS) [2].

GPASS is an administrative system used by 85% of general practitioners in Scotland, as a facility for managing and uploading patient records. At periodic intervals the patient information is uploaded to a central repository which is hosted by SCI Store. SCI Store can be accessed by a variety of web services which have been specially sanctioned by the NHS. These provide a uniform method of access and data retrieval, however it has been noted that regional variations of SCI Store have appeared in recent years, causing issues in terms of heterogeneous data matching.

Whilst the clinical IT infrastructure in Scotland is relatively well-developed compared to the rest of the UK, it is still a largely paper-based system that is currently used. There are many reasons for this, ranging from unwillingness on the part of healthcare professionals to learn new software processes to the limited success of large-scale healthcare IT implementations. This helps and hinders the VOTES project in equal measure: on the one hand there is clearly an immediate need for a system like this to be developed; on the other, how can technology be leveraged if the building blocks are not securely there in the first place?

2.2 England

In England, there are a number of initiatives to achieve a federated clinical information infrastructure, however most have not gone beyond the development of standard specifications. A high-profile example project has been undertaken by the parliamentary initiative Connecting for Health [3], which attempts to standardise the interfaces used by individual practices, is MIQUEST [4].

MIQUEST provides standard interfaces to be used by individual general practices across the country, so that central facilitators can manually upload and transfer data between nodes, and perform analysis over a largely standard data-set. It is laid down as an industry best-practice to have these databases "MIQUEST-enabled". However, there is one major drawback to this technology, namely that there is a lack of real-time communication between the central repository and the distributed practices. Again, this state of the infrastructure helps and hinders the VOTES project in equal measure, for the same reasons as described previously.

Beyond the technologies used to attempt to link systems however, there have been a number of studies conducted that provide very complete and provisioned data-sets. An example of such is the GPRD (General Practice Research Database) data-set [5], which has been used as the basis for a number of large-scale analyses in England and Wales. The availability of these data sets allow testing to be performed of sampling technologies, producing benchmarks that can be verified against procedures that have been carried out manually previously.

The technological infrastructure underpinning the health system in England appears to be in a marginally less-developed state than that of Scotland. However, both suffer from issues of too many standards, not enough widespread adoption of a single, clear leader, and consequently, a lack of mature, stable platforms upon which to build concrete distributed systems. These standards [6-8] and initiatives are numerous and with their continued development it is hoped that they will achieve just such a solid platform. However without clear control from authoritative agencies, useful, distributed solutions may still be a long way off. This is one of the primary reasons for the research effort in the VOTES project.

3. VOTES INFRASTRUCTURE

Clinical trials are procedures and processes by which new medical drugs, treatments and interventions achieve validation to demonstrably improve quality and length of life for patients. As has been mentioned previously, the central idea behind the VOTES project is to tap the information stored on primary and secondary care patients, and use this to efficiently target patient recruitment for clinical trials and manage the conduct and processes of those trials more generally. Aligned with the vision of data grids, a virtual organisation (in this case known as a Clinical Virtual Organisation, or CVO) is set up that allows various data repositories to be available to the different partners, which would not have been otherwise available. Figure 1 shows the conceptual schematic of a CVO.

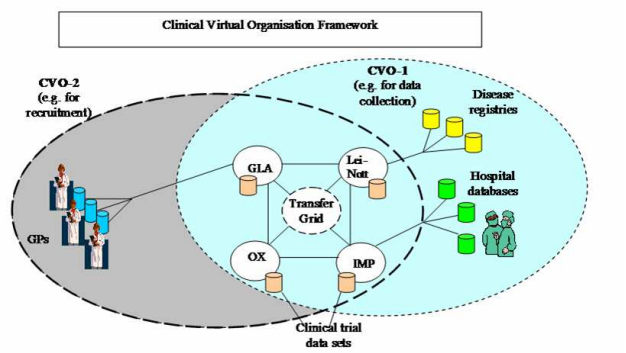

Figure 1: a Clinical Virtual Organisation (CVO) diagram. GP databases are linked to hospital databases and disease registries to allow greater linkage and enhanced data value.

A key point that will be discussed later in the paper is that the partners only have limited trust between each other. Also, in order to realistically generalise the concept, the CVO must also be assumed to be of a transient lifetime. So today's partner may be tomorrow's competitor.

To achieve this technically, the VOTES system has been designed on a modular basis, with each node comprised of architecture as shown in figure 2.

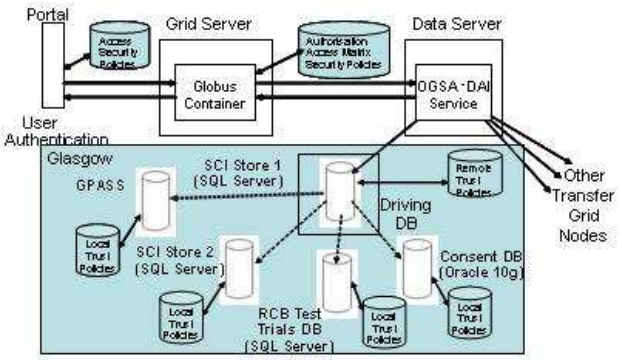

Figure 2: The architecture of a single node on the CVO. Other nodes have analogous structures and inter-node communication occurs between the data server components.

The node is made up of the following components:

- A portal implemented using GridSphere [9], a technology specifically designed to give user-friendly and lightweight access to grid resources.
- A Grid server implemented using version 4.0 of the Globus Toolkit [10]. The methods written here allow linkage between the SQL queries and the data server, but primarily provide an access control point that enforces the CVO-wide security policies.
- The data server is implemented using OGSA-DAI [11]. Until version 3.0 was used, the data server simply served as a conduit for the results of the distributed SQL, executed and joined on the driving database below. However, with the new functionality provided by the latest version of OGSA-DAI, joining of federated queries is now possible at both the data server and the driving database level. This allows a wider range of data resources to be accessed as is discussed later.
- The databases containing the clinical data – the federation of which is the ultimate aim of the node – are grouped together under one guardian database, known as the driving database. This database allows the data from the various sources to be joined together and presented as one resource to the rest of the system.

The main application built on this architecture is a data retrieval portal that allows searches to be run of clinical databases, enhanced through various linkages, yet presented as one resource to the end user as described previously.

Additionally, there are also a variety of supporting features built at the application level, of which a brief description is pertinent:

- Administrative portal – the system has a fully separate portal tab that allows connection information for local and remote node resources to be interrogated and uploaded, as well as trial permissions to be created, with databases and roles added as appropriate.
- Connecting to a consent database – only those patient details will be released if the patient has specifically consented to their viewing/release and usage in particular trials and studies, by means of a flag in this database.
- Meta-data querying – the administrative portal has the ability to query the parameters of the databases and populate the security policy defining the access to and usage of those data sets (see section 4.1).
- The portal uses Google maps [12] to provide geographical information associated with patient records. Currently only individual records are located using the portal, however work is in progress to show how geographical distributions of conditions, treatments, can be shown as well. The implications of this are many – one simple example for instance, could be to identify the prevalence of a certain condition associated with a new treatment under trial, then to focus the patient recruitment campaign in that area, to maximise positive results.

Figure 5 (on the final page) shows a variety of screenshots when a user interacts with the VOTES portal. The results of a distributed query are shown, along with available pictures of MRI brain scans, associated lab data, as well as the geographical location and CHI number of the patient.

The security implications of these features are important. Each feature provides either a means to manipulate the overall use of the system, or, in the case of the maps tool, provide individual identification along with geographical location, the combination of which is a highly sensitive piece of information. The need for rigorous security is therefore paramount and is discussed in detail in the next section.

4. IMPLEMENTATION AND USE

To make a viable solution that can account for the differing infrastructures encountered in a flexible and efficient fashion, the emphasis throughout development has been on modularity of application programming and the "*plug-ability*" of the various

components with a wide variety of resources, whatever their structure.

In terms of data classification, the ideal solution would be an ontology that can account for different resources, without knowing before run-time how the resource is structured. However, the reality is that, in a manner similar to dictionary construction, an ontology can only be built by knowing the underlying details of the infrastructures being connected.

As such, the VOTES infrastructure is programmed with connecting information for the most popular types of database resources in use by health infrastructures across the UK. The initial requirements gathering phase of the project, and subsequent developments, have shown that these are largely Microsoft-based (SQL Server and Access) but other data technologies are also in use, and must be accounted for.

This section describes how the VOTES infrastructure has been used with the data sets described previously (SCI Store, GPRD and BPH study), and the challenges that have been encountered in the process.

4.1 Security

Due to the clinical, and therefore highly sensitive, nature of the data involved in this project, the top priority in every endeavour is that of security: identifying the inherent risks, analysing their importance, and mitigating against them appropriately.

Because of the highly sensitive nature of the data however, the traditional methods of security threat analysis and prioritisation of the mitigating actions against some cost minimisation model for example cannot be used. In this domain, any potential risk of data disclosure has serious consequences for all parties involved and must be avoided at all costs. There is also the issue of trying to combine different policies by partners with differing levels of trust between each other, for differing time periods, and with varying levels of applied rigour. The solutions presented here developed within VOTES go some way to reconciling these issues.

In terms of the dynamic implementation of technological security, a two-tier system has been introduced: a CVO-wide policy that delimits the fields that various roles within the CVO can view; and a local resource policy, which is entirely at the discretion of the data resource owner, and ultimately overrides the CVO policy. The former can be considered an aggregation of the latter over many sources, updated at periodic intervals.

The CVO-wide security policy can be expressed in terms of an Access Matrix model [13]:

	R_1	R_2	R_3	R_4
	h_1	h_2	h_3	h_4
U_1	U_1	0	0	1	0
U_2	U_2	0	0	0	1
U_3	U_3	1	1	1	1
U_4	U_4	0	1	0	0

Figure 3: a conceptual access matrix model. Depending on assigned privileges, roles can access resources or not. In the above diagram, role U_2 can access resource h_4 but not h_3.

This concept is a familiar one in terms of expression and enforcement of computer security policies. However, one of the main benefits of using this approach has been in terms of implementation. The concept can be neatly encapsulated in a simple database, located locally on each node, updated using secure out-of-band communications with the other nodes, and encrypted using a key that only local users have access to. The database is interrogated using simple SQL queries and this ultimately presents a list of privileges available to that user, and nothing else. The "per-parameter" nature of this implementation, allows a far more flexible security policy to be implemented with the minimum of overhead in application programming from the communicating party, i.e. the design is modular and "pluggable".

At a higher level of abstraction, the need for super-users to administer the system has been identified, to underpin the activities of "regular" users, who will largely be data gatherers of some form. The roles of these super-users fall broadly into two categories: a *node administrator* and a *trial administrator*.

The node administrator designs and enforces security policies with regard to the infrastructural aspects of the system. They would ordinarily be trained in administering computer systems and would sanction the addition or removal of the various components of the system. The trial administrator will be a clinical specialist, and will be responsible for using the underlying technical resources to design and enforce security policies for the actual clinical trial recruitment campaigns and data collection processes.

The final level of abstraction when discussing security, applies to production contexts of a system. This is the requirement for an over-arching, static agreement, which legally binds parties to predefined responsibilities, and outlines the recourse of those parties in the event of any breach of security in the system. Though often overlooked in discussions of technological security solutions, this is a mandatory consideration that no technological solution will ever super-cede.

4.2 Connecting Domains

The reality of establishing a CVO involves the following procedures. The first step in establishing a new resource in the VOTES infrastructure is to enable the connection between partners, which essentially delimits the bounds of the CVO. This is inherently static in nature where agreements on the connection and the reason for the connection have been identified already, e.g. through agreement of a protocol outlining the resources to be accessed and shared which has been independently reviewed

by for example Caldicott guardians or Patient Information Advisory Groups.

In the first instance, firewalls between the participating sites (in this example, the University of Glasgow, the University of Nottingham and Imperial College London) must be opened to specific machines across the appropriate ports. Additionally, an account must be created at each site which allows the connection of the remote site to the new resource. It is desirable that this account be as restricted as possible from the remote site, i.e. read-only.

In order to connect, the security information associated with the steps above (for instance, a username and password for the account) must be communicated to the participants at the remote site. However, this only takes place once the security at each site has been established to each party's satisfaction. Ideally this step consists of a face-to-face meeting, with inspection of the local security facilities and an interview with the administrator responsible for those facilities.

In terms of heterogeneous resources, this example is useful as the resource presented from Imperial College London is based in a MySQL database, which differs from the assumption of most data sources being Microsoft products.

As such, various modifications were required to the code to allow the presentation of this different data source. As an example, previously the following syntax had been used for most sources:

```
OPENDATASOURCE ("Data Source", "Server name
+ connection information")

[Embedded as representing a table within
the SQL]
```

But this was required to be changed to the more general version, when joining these using MySQL:

```
OPENDATAQUERY ("Linked Server Name", "SQL")
```

With modifications such as these for the most prevalent data sources encountered, the VOTES infrastructure is more flexible and robust in addressing the wide variety of resources in the field.

It should be noted here that the syntax above is a construct of the Transact-SQL language [14], which is a sophisticated aggregation of various "regular" SQL statements, supported only by a limited number of commercial vendors. The benefits of this are that a ready-made tool, for joining data sources in a way that efficiently load-balances, is immediately available. However, partly because of the fact that this is produced by a commercial, competitive entity, the joining of federated queries will not work with every available data source. In order to cover a more comprehensive range of data sources, the OGSA-DAI technology has been enhanced to allow joining between sources such as, say, PostgreSQL databases. These architectural issues are ultimately hidden from the end user but are an important factor in the back-end processing of such a system.

Using these methods, dormant connections between the three participating parties were established, to be used and available when the CVO was required to gather and process data from the different sources.

4.3 Data and Analysis

As stated previously, the data for this case study was drawn from the SCI Store repository in Glasgow, the GPRD study at Imperial and a BPH study in Nottingham.

The data in SCI Store at Glasgow comprised a data-set representation of that used by the live repository and GPASS administration system by GPs and clinicians throughout the various regions of Scotland.

The data provided from the GPRD study was based on a real set of diabetes data, but randomised and "de-linked" in such a manner as to render no identification of real patients possible. So for the main purpose of the data, clinicians with the appropriate privileges would be able to identify the patients as and when necessary, but others would not, despite having access to the statistical information it provided.

The data to be provided from the BPH study at the University of Nottingham is still at an early stage of processing. As such, it is possible for the VOTES project to have input into how the structure will be identified and how the data will be stored electronically. This gives the project a valuable insight into the political reaction to attempts of remote sites to guide the infrastructure implemented, and to see how well the proposed solution will be accepted in new infrastructures in general.

In practical terms, the scenarios implemented involved various combinations of users from remote sites accessing data from the partner sites that they would not have ordinarily had access to without the VOTES infrastructure. A typical example would be to look for occurrences of patients with diabetes in the SCI Store repository then linking this data-set with those in the GPRD. A similar query can be run for benign prostate hypo-plasia, linked to the data-set from Nottingham. In this way, statistical and geographical distributions of these conditions can be gathered in a much more accurate and efficient manner than is currently possible.

In terms of clinical trials, it is often statistical information that is of most benefit to say patient recruitment or follow-up data collection. However, there is also the benefit of being able to link records of any patients that may happen to appear in two or all three of these studies. Currently, the possibility of this occurring is unlikely, but as the number of data-sets increase, so the likelihood of being able to correlate and accurately compile medical histories in this way.

With the wealth of nationwide statistical data that this infrastructure potentially unlocks, comes the ability to run large-scale analyses over that data. As such, several specific queries are now being coded that can be run over the linked data-sets.

Currently, these include:

1) A cross-sectional time trend study on quality of diabetes care in general practice.

2) A cohort study of adverse drug reaction to Rosiglitazone.

These particular queries have been chosen as they are of specific interest to the partners at Imperial College London in the first instance [15], and can be linked to the other two data-sets. As further data-sets become available these kinds of analyses will be greatly expanded and have greater incidence between sets.

5. FURTHER DEVELOPMENTS

This section describes further developments that will affect the direction of the VOTES project in its final year of funding, which are directly relevant to the example presented.

5.1 Web-service protection

As uptake of the VOTES system has progressed, larger clients such the Robertson Centre for Biostatics in Glasgow [16] and the UK Biobank project [17] have expressed an interest in using the infrastructure presented here. Naturally, given the scale and scope of these clients, much more rigorous discussions have taken place regarding the security and viability of the solution presented. For the VOTES project this is the next logical step in terms of acceptance on a production scale.

A major result of these discussions was the fact that the clients are unwilling to accept the level of connection required by other CVO partners to their own data sources (described in section 4.2). The alternative presented has been that the individual partners will provide WS front-end implementations of "canned queries" to their own data sources. This method allows these remote sites to provide much greater control over their own data sources, and provision a level of security that satisfies their own remote policies.

With regard to the design of the VOTES infrastructure, schematically the difference can be seen in figure 4 below. Instead of talking to databases (individual or distributed "guardian"), the data server must now also talk to WS interfaces. This has required an extra overhead programmatically, and has required the OGSA-DAI team to provide activities for this specific task (and with version 3.0 it is now possible to join the results of federated queries). Hence the broader appeal of the infrastructure has been greatly enhanced since this Web service approach is likely to be the favoured one when the infrastructure expands to include larger, more competitive, or more sensitive data sources.

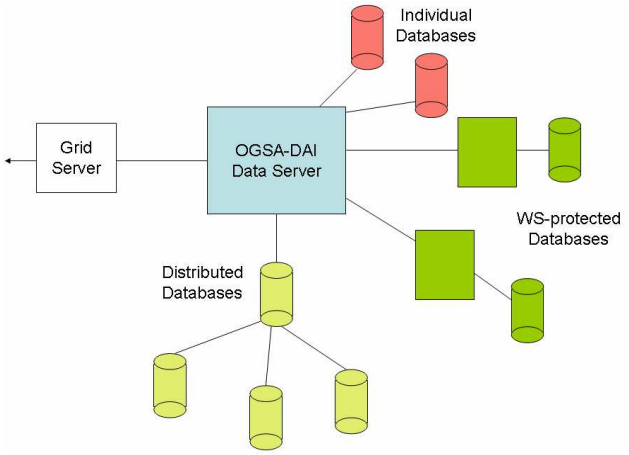

Figure 4: The latest VOTES architecture, now using version 3.0 of the OGSA-DAI data server. The result is that a wider range of data resource types can be queried and joined (including individual databases, distributed databases, and databases protected by Web service front-ends).

5.2 Other Security Technologies

As is the nature of research, various technologies must be experimented with before the best solution available can be identified. This is the mainstay of the work conducted at the National e-Science Centre in Glasgow, and as such, other research projects have direct bearing on the VOTES project and should be described here.

One project investigating the feasibility of security applications in the grid landscape is the VPMan project [18]. The proposal is to look into linking two of the most established authorization technologies available: VOMS (Virtual Organisation Management Software) [19] and PERMIS (PrivilegE and Role Management Infrastructure Standards validation) [20]. Both technologies attempt to allow flexible policies to be developed, which follow the paradigm of virtual organisations within grids – namely to allow a transient and loosely bound collaboration operate with the flexibility required, whilst making no sacrifice in terms of the security demanded by each partner in the VO.

A deliverable of the VPMan project is the application of the solution to a distributed scenario already in operation. In the VPman project we have shown already how VOMS attributes can be used by PERMIS to make an authorisation decision on access to a GT4 service. The service itself was based upon the VOTES project. The results of this experiment and the exploitation of other scenarios, e.g. using VOMS, PERMIS and OMII-UK technologies are described in [21].

A major difference between the VPman architecture and the authorization module currently used by VOTES is the "per-service" method of authorization, i.e. it is fixed stored procedures that are protected (authorised). As a proof-of-concept, the VPMan solution highlights how other technologies can be integrated with VOTES, but in terms of granularity and flexibility, the main VOTES project is likely to continue using the more flexible "per-parameter" method of authorization.

Another technology rapidly gaining acceptance in the academic security community is that of Shibboleth [22]. Shibboleth provides a mechanism by which attributes can be exchanged between parties that provides a flexible and dynamic method of authenticating and authorising users. By modular use of the repositories and transfer mechanisms, a federation is built up which can allow single sign-on (SSO) access to a variety of resources. The different example trials available through the VOTES portal can be accessed through a "shibbolized" version of the portal, housed in Glasgow, but accessible to selected users that are part of the UK Access Management Federation [23] in possession of the appropriate attribute certificates. The scoping of these attributes and their distribution to known and trusted collaborators, along with user oriented attribute release policies is currently being explored within the SPAM-GP project [24].

6. CONCLUSIONS

The VOTES project is a pioneering attempt to establish a "proof-of-concept" framework that allows the easy federation of clinical data from around the nation to support a range of trials and studies. The technological solution outlined is, we believe, an extensible and robust architecture that allows the easy addition of new resources and continues to grow and adapt with every data source added. The data-sets that have been federated together have provided much insight into the challenges that exist already, but have also provided the users of the system with a viable and useful tool that can potentially direct the development of clinical trials in the future.

The integration of the VOTES architecture with the various partners and remote sites has brought up not only technological issues, but those of a more political and human nature. Simply put, people are often reticent to provide the type of access required between partners in a loose collaboration where only limited trust models exist. As has been shown, the architecture and approach of the VOTES project has been flexible enough to accommodate such needs, and as such, will likely develop further in this, and other follow-on projects.

As the VOTES project continues (approaching its third and final year), the technological solutions to the problems involved have matured and are now finding structure as the "best" way to approach the federation of clinical data. As such, the final development phase is less likely to concentrate on finding new ways of achieving the goals, but on strengthening the ways that have been found to work so far. In this regard, the aspect of security will be the main focus of the project, which in turn will provide strength in promoting the influence of such an approach to federating clinical data.

Finally, the political structure of the country, and the relation of the technological solution to it, has been emphasised strongly here, because it is one that has similar parallels throughout the developed world. The states in America, the states and territories of Australia or the provinces of Canada, have structures that have many political analogies to that of the UK. With the flexibility, security and robustness of this infrastructure, it is hoped that this approach to federating data can be a possible model for use world-wide.

7. ACKNOWLEDGEMENTS

The authors would like to acknowledge the UK Medical Research Council (MRC), the funders of the VOTES project.

8. REFERENCES

[1] SCI Store – http://www.show.scot.nhs.uk/sci/products/store
[2] GPASS – http://www.gpass.co.uk
[3] Connecting for Health – http://www.connectingforhealth.nhs.uk
[4] MIQUEST - http://www.connectingforhealth.nhs.uk/systemsandservices/data/miquest
[5] GPRD – http://www.gprd.com
[6] ICD coding background - http://www.connectingforhealth.nhs.uk/systemsandservices/data/clinicalcoding/faqs
[7] SNOMED-CT - http://www.connectingforhealth.nhs.uk/systemsandservices/data/snomed
[8] HL7 – http://www.hl7.org
[9] Novotny, Russell, Wehrens - GridSphere: an advanced portal framework, Euromicro conference 2004, 30th Proceedings
[10] Foster, Kesselman – Globus: a metacomputing infrastructure toolkit, International Journal of High Performance Computing Applications, vol. 11, no. 2, 115-128 (1997)
[11] Karasavvas et al. – Introduction to OGSA-DAI services, Lecture Notes in Computer Science, vol 3458/2005
[12] Google maps – http://www.google.co.uk/maps
[13] Access Matrix Model – R. S. Sandhu and P. Samarati, "Access Control: principles and practice" IEEE Communications Magazine, vol. 32, no. 9, pp. 40-48, 1994
[14] Transact-SQL - http://en.wikipedia.org/wiki/Transact-SQL
[15] Personal comm. – Dr Azeem Majeed, Anthea Ng, Imperial College London
[16] RCB – http://www.rcbweb.co.uk
[17] UK Biobank – http://www.ukbiobank.ac.uk
[18] VPMan (also VP-Authz) - http://labserv.nesc.gla.ac.uk/projects/vp-authz/index.html
[19] Alfieri et al., VOMS: an authorization system for virtual organizations, Lecture notes in Computer Science, ISSN 0302-9743
[20] Chadwick D.W., Otenko O., The PERMIS X.509 role based privilege management infrastructure, Future Generation Computer Systems, vol. 19, Issue 2, Feb 2003, pages 277-289
[21] R. Sinnott, D. Chadwick, T. Doherty, D. Martin, B. Nassem, A. Stell, J. Watt, Advanced Security for Virtual Organizations: Exploring the Pro's and Con's of Centralized vs Decentralized Security Models, in preparation for 8th IEEE International Symposium on Cluster Computing and the Grid (CCGrid 2008), May 2008, Lyon, France
[22] Shibboleth - http://shibboleth.internet2.edu/
[23] Access Management Federation - http://www.ukfederation.org.uk/
[24] J. Watt, R.O.Sinnott, J.Jiang, T.Doherty, A.J.Stell, D.Martin, G.Stewart "Federated Authentication & Authorisation for e-Science" Proceedings of APAC'07 Conference, September 2007, Perth, Australia

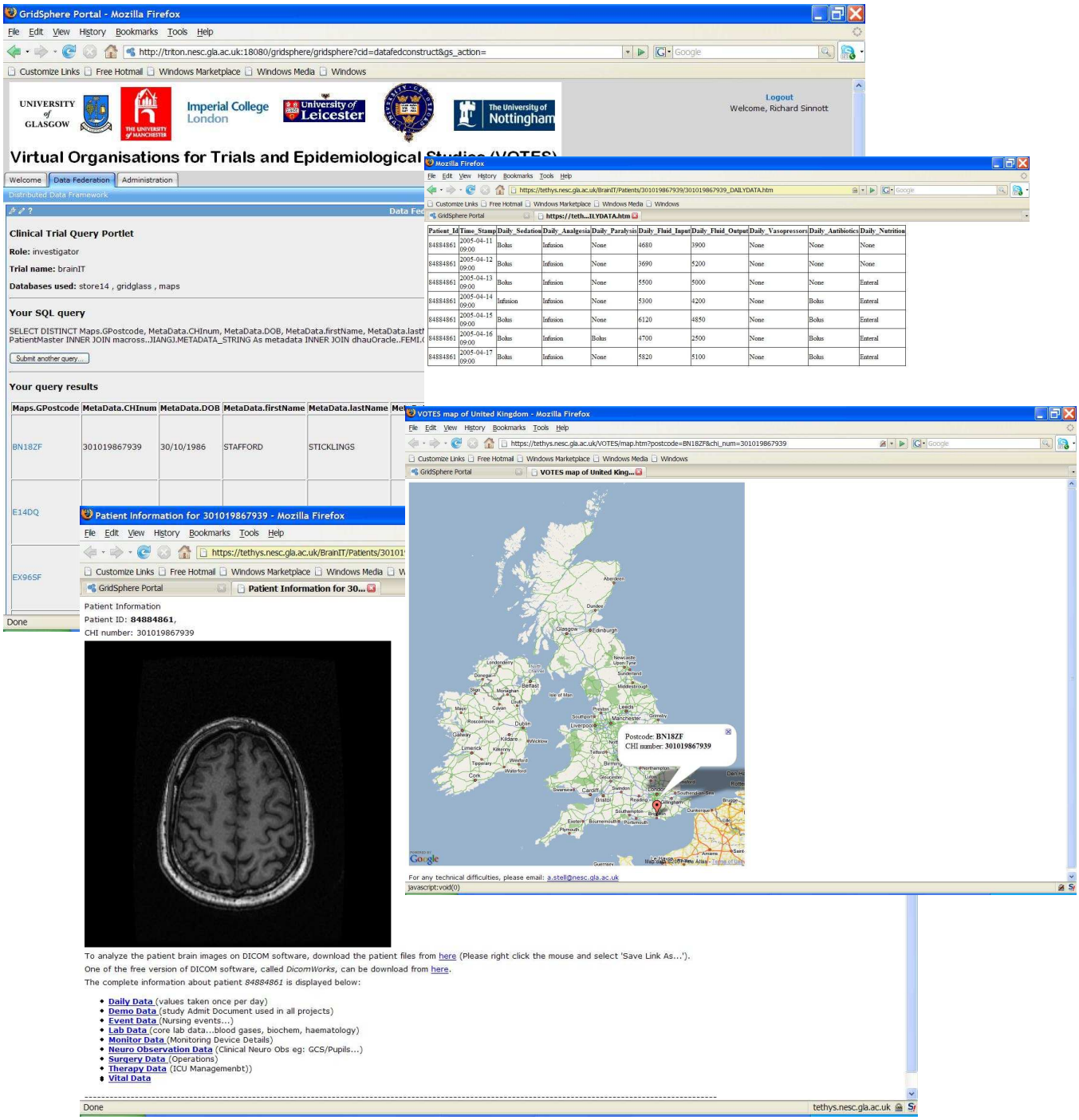

Figure 5: The interaction of a privileged user with the VOTES portal can bring back a variety of clinical information from distributed sources. Shown are patient information lists (on the back left picture) an image of their brain MRI scan, some associated lab data and their location within the UK. This kind of information drawn from many sources has the potential to be greatly beneficial to the conduct and processing of clinical trials.