Secure, Reliable and Dynamic Access to Distributed Clinical Data

Anthony Stell, Richard Sinnott and Oluwafemi Ajayi

University of Glasgow, National e-Science Centre, Glasgow, G12 8QQ, UK
a.stell@nesc.gla.ac.uk r.sinnott@nesc.gla.ac.uk o.ajayi@nesc.gla.ac.uk

Abstract. An abundance of statistical and scientific data exists in the area of clinical and epidemiological studies. Much of this data is distributed across regional, national and international boundaries with different policies on access and usage, and a multitude of different schemata for the data often complicated by the variety of supporting clinical coding schemes. This prevents the wide scale collation and analysis of such data as is often needed to infer clinical outcomes and to determine the often moderate effect of drugs. Through grid technologies it is possible to overcome the barriers introduced by distribution of heterogeneous data and services. However reliability, dynamicity and fine-grained security are essential in this domain, and are not typically offered by current grids. The MRC funded VOTES project (Virtual Organisations for Trials and Epidemiological Studies) has implemented a prototype infrastructure specifically designed to meet these challenges. This paper describes this on-going implementation effort and the lessons learned in building grid frameworks for and within a clinical environment.

1 Introduction

Evaluation of new drugs or treatments requires infrastructure support to allow procedures to be validated and verifiable results to be achieved within reasonable timescales. Rather than building bespoke infrastructures for each clinical trial or observational study, it is far more efficient to develop a framework that can be reused and applied across numerous studies. A key challenge in supporting this is to efficiently and effectively harness the plethora of clinical data that exists across regional, national and international boundaries for clinical trials and unbiased evaluations of their outcome.

Such a framework places critical demands on the infrastructure. Three of these which are the focus of this paper are: *security* in ensuring that the right data sets are made available to the right people for the right purpose; *reliability*in ensuring that a failure of any parts of the system does not automatically mean that the rest of the system is unavailable; *dynamicity* where a multitude of different clinical trials or observational studies can be rapidly supported where different clinical data sets with different security policies and different users can be brought together for the conduct of a given trial or study. This paper outlines how the Medical Research Council funded Virtual Organisations for Trials and

Epidemiological Studies (VOTES) project has developed an infrastructure that supports such capabilities.

2 Infrastructure Requirements for Clinical Trials

From a data perspective, the inherent challenge is in building an Information Grid [1] matching data across domain boundaries and understanding the relationships between differing classification and indexing systems (if such relationships actually exist). This should support the overall processes involved in establishing and managing a clinical trial including patient recruitment, data collection throughout the trial and overall trial management. Once these data sets are federated they can then be presented under a single global schema description and have distributed, federated queries executed upon them as one ostensibly unified resource. Data is only one aspect of such a framework and a fundamental necessity is to provide a security infrastructure where cross-institutional policies on data access and usage can be agreed upon, and visibly enforced to the satisfaction of all data providers, data owners, infrastructure administrators, ethical and legal advisees and the numerous other actors that play a significant and important role in clinical trials.

Clinical trials and studies are driven by statistics. Understanding the often moderate impacts of drugs often requires significant numbers of people to be involved in a study. This level of scale necessitates an infrastructure capable of handling large volumes of data. As an example, the UK Biobank [2] will be gathering information on 500,000 participants aged between 40 and 69. A grid framework provides one infrastructure capable of dealing with such volumes.

A major requirement of an infrastructure built to support this kind of initiative is that of reliability. A simple definition of this can be expressed as: *Can a user depend on the software and system behaving in a manner that they expect?* [3]? Hence the users of a service provisioned by this technology must have a guarantee that it will function and behave as they expect it to. For any production system, this guarantee, also known as an acceptable quality of service (QoS), can only be provided if the system can be described as:

- Reliable protected against system damage, interruption, resource scarcity or degradation.
- Secure protected against hostile action by a malicious third party.
- Scaleable able to function smoothly and without incident, despite simultaneous access by large numbers of users.
- Dynamic able to function smoothly despite being run across heterogeneous environments, with differing policies governing access and use.

Only if all four of these items are sufficiently addressed, can the end-users be confident that the system will behave in a predictable manner as described by the service providers.

3 Clinical Virtual Organisations

The concept of a Virtual Organisation (VO) is one that is central to Grid computing. A VO is defined as a collection of resources shared in a coordinated and dynamic, yet highly-controlled manner, in pursuit of a common goal [4]. There are several features that are common to most VOs:

- The need for a common security policy, with a well understood degree of trust between participating parties and resources.
- Sharing of data, storage and processing power within the VO for the achievement and common good of the VO according to VO policies.
- Acknowledgement of the transient lifetime of the VO.

Grids involved in the clinical domain are primarily concerned with information and information flows. VOs in this field impose additional requirements uncommon to VOs in other domains due to the strict security requirements which have to be directly and explicitly upheld. As such, the term Clinical Virtual Organisation (CVO) is used here, to distinguish these groupings from the more common definition of a Virtual Organisation.

The most immediate consequences of this information-centric nature are grounded in security. Issues such as the anonymisation of data and the ability to statistically infer identifying data must be considered, and implemented with the utmost regard for data privacy and integrity, given the potentially high sensitivity of the data involved. Additionally, the lack of unification in the distributed format of clinical information is another challenging and distinctive feature of this domain and the associated CVOs.

There has been much work investigating the methods of applying security to VOs, with particularly relevant sources being the DyVOSE [5] and the GOLD [6] projects. There are also a number of initiatives that attempt to set up VOs specifically for the life sciences domain, with characteristics similar to the CVOs described above [7] [8] [9].

3.1 Trial data across domains

There are many challenges involved with implementing the different aspects of CVOs. Two major issues relating to heterogeneous data federation must be overcome before the concept can become a reality:

- The lack of a unified data structure to represent common clinical concepts and classifications.
- The lack of an index to match unique individual records on either side of a domain boundary with each other.

Several solutions to these problems have been proposed. Health-Level 7 (HL7) [10] has been mooted as a possible global schema by which all clinical data refers to a common global standard. This solution is attractive in that the problems associated with heterogeneous environments would be greatly reduced if this standard was widely adopted. However, to date, it has not had widespread/universal

uptake. HL7 is also a highly centralised solution, which diverges from the ideal grid scenario of equally dispensable, contributing nodes providing the overall functionality of a system.

A proposition that provides a more grid-like solution is to have a peer-to-peer matching of local datasets, where brokers governing inter-domain communications only have knowledge of the schemas belonging to the various other brokers that they encounter. In this way, a knowledge base of local schema could be built up that would be comprehensive for the immediate environment, but was not unnecessarily burdened with knowledge of data structures that it does not require.

However it is implemented, a reliable infrastructure must exist on either side of a domain boundary that supports the mechanism to federate the data. That infrastructure must have a certain amount of knowledge of the schema beyond that boundary, and must be relied upon to effectively implement the federation service.

3.2 De-centralisation of components

The first line of defence in protecting a system against non-hostile malfunctions is to replicate critical system components across several machines and implement application code that automatically tests for, and subsequently uses, live components within this network.

As long as it is efficiently implemented, this architecture allows for system damage, network interruptions and power failures, ideally with no noticeable detriment to the overall quality of service provided to the end user. It should be noted here that such concepts are not new and are implemented as best practices in most current production IT systems. The clinical domain imposes certain other requirements however in ensuring the robustness and tolerance of replicated components. Compromises of any given components should not compromise all nodes in the system where those components have been replicated for example. This is especially important for multi-institutional scenarios where code is replicated.

In terms of grid technology, the emphasis is upon how these networked backup systems relate to each other within the CVO. Again, two solutions are immediately apparent:

- Have each node mirror each other in terms of backup architecture. This is a static solution and runs the risk of having many unused resources.
- Have a notification system within the CVO that updates each node with the set-up necessary for effective production use at any given time. This requires more sophisticated application code to be implemented, but ultimately makes more efficient use of the distributed and disparate resources.

3.3 Dynamic resource allocation

Another aspect that grid technology adds to the established concepts of system reliability is the need for a dynamic and flexible solution. Within a CVO, re-

sources may not be available, not just as a result of malfunctions, but also as a result of restrictions implemented by a local resource administrator, possibly limiting the resource availability to the rest of the CVO.

However, to implement this aspect effectively, an over-riding static context is required within which dynamic resource allocation can then occur. This static context would most likely take the form of a pre-defined legal agreement between the participating nodes of the CVO.

The mechanics of dynamically allocating resources would require the ability to poll and monitor the availability of resources within the CVO. An initial static list of potential resources would then be available to each node, and the use of those resources would be dependent on the immediate results of the polling.

A foreseeable extension to this system would be the ability for the application to automatically assess the load in a given area of the CVO and re-distribute it for optimal overall performance. Such resource broking and information services are common to grids in other less security focused domains, yet largely uncommon in the clinical domain where resources typically are not advertised.

4 VOTES Project Overview

The VOTES project (Virtual Organisations for Trials and Epidemiological Studies) [11] is a 3-year, 2.8 million project funded by the UK Medical Research Council, and is specifically detailed to investigate and address these large-scale clinical challenges using grid technology. The VOTES project itself is a collaboration between the universities of Glasgow, Oxford, Imperial College London, Leicester, Nottingham and Manchester.

The alpha and beta prototype applications that have been developed so far provide a portal gateway to clinical data distributed across several representative test databases to form a data federation system [12], by use of distributed grid and data services. The software used is a combination of GridSphere [13], Globus Toolkit v4 [14] and OGSA-DAI [15]. The clinical databases contain realistic clinical data that accurately represent the clinical datasets and associated software in current use in the Scottish National Health Service (SCI Store [16], GPASS [17], SMR [18]) with on-going negotiations for access to live clinical data.

4.1 VOTES architecture

In terms of critical component services, the architecture of a single node in the CVO is shown in Figure 1. This architecture has been implemented and is available for testing at [19].

A portal connects to a grid server, which in turn connects to a data server, a driving database (a database through which the distributed SQL is submitted to the pool), and finally several auxiliary databases from which the clinical data is retrieved. The modular design allows components to be inserted and removed easily. The interface to allow for future expansion occurs at the data server, where it is envisioned that the node will connect to other nodes in the CVO

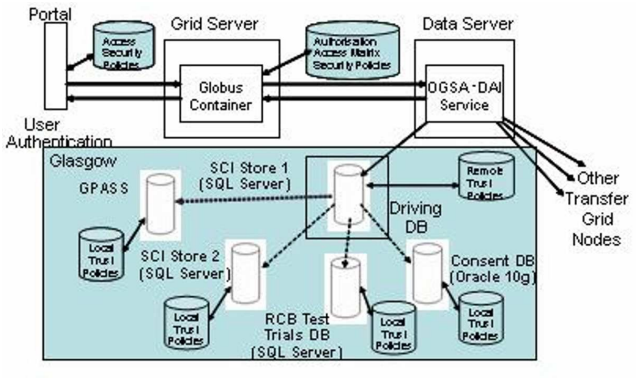

Fig. 1. Glasgow CVO node architecture

(at the other university partner sites and clinical data centres in the NHS for example). Trust policies represented as CVO database rules on legal (authorised) connections are used to ensure that only authorised people can issue queries and importantly that those queries adhere to the agreed contracts between the data providers and data owners to the different roles associated with a clinical trial, e.g. investigators, nurses etc.

To illustrate the reliability factoring of the current set-up, Figure 2 shows the system in terms of real machines at NeSC Glasgow.

With the knowledge of the systems modus operandi from Figure 1, Figure 2 shows the possible paths of operation between the different components and their duplicates on the Glasgow node of the CVO. In order for this architecture to operate effectively it is necessary to implement application code that will, in the first instance, test the live-ness of a preferred component and only attempt subsequent code execution if that test comes back positive, i.e. the component is available.

If the test is negative then the code must move on to the next assigned resource that provides this component, where it will attempt the same test. Only once all the available resources have been tested and no positives have been returned will the system report to the operator that it has failed and cannot proceed with execution. The code that implements this test in VOTES is as follows for the grid and data servers:

```
String service = http://serviceIP/GridOrDataService;
URL url = new URL(service);
URLConnection dc = url.openConnection();
```

A simple Java URL object is created from the given (standard) URI of the grid or data service. A connection is opened to this service, and a handshake is initiated to establish that the URL is valid and accepting connections. The

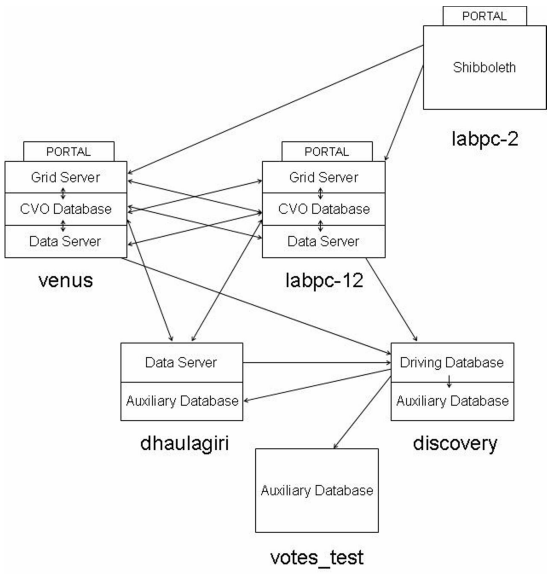

Fig. 2. Actual architecture showing support for reliability at NeSC Glasgow. Note the use of Shibboleth technology to protect the portal access on labpc-2.

code for testing the various databases - CVO, driving and auxiliary - uses (again standard) JDBC connection techniques:

```
String url = jdbc:dbType:dbIP/dbName;
Class.forName(driver);
Connection conn = DriverManager.getConnection(url,username,password);
conn.close();
```

The JDBC URL is constructed, the driver for the particular database type is referenced and a connection is opened with the relevant username and password. (These objects are all included in the standard Java SQL packages.) Once the connection has been successfully opened, it is immediately closed again.

It should be noted that in terms of service provision, the auxiliary databases are termed as such because they are not critical to the overall operation of the entire system. However they do provide the *raison d'etre* of the system the clinical data. If an auxiliary database fails then no change occurs in the operation of the VOTES system, but the data housed on that machine will not be available. Because of this, though they are technically not critical components for the smooth operation of the service, the auxiliary databases must be maintained with equal rigour as the rest of the infrastructure.

It is envisioned that in the pre-defined static agreement, a responsibility for node maintenance would be detailed and within this, a hierarchy between critical

and non-critical components would be established. Also, any changes to the local administration of username/password accounts, and their communication to the rest of the CVO, would have to be outlined in this agreement as well, with the possibility of dynamic tools to automate this procedure being a possible enhancement to the framework tools.

4.2 Implications of architecture

Several implications arise from the application code that has been implemented so far:

Security In order for the connections to succeed the relevant ports must be opened between the portal server and the other machine, which provides a potential security hole. The risk is mitigated somewhat if both server processes are in the same node, by the fact that the communication is between two trusted machines. But if it is a communication between machines that only trust each other to a limited degree, such as between two CVO nodes, the situation is more complicated, and would likely be referred back to the terms of the over-riding static agreement.

Distribution The code above requires specific Java driver libraries to be distributed within the portal server to allow connections to the different types of databases. The main issue here is anticipating the various database types that will be used in the clinical domain. Again, this could be addressed by defining in the initial agreement a set of databases that would be used for implementation within the CVO. A study of the most popular and widely adopted clinical databases would provide a useful idea of what these would be.

Distribution with this knowledge would be less problematic as the pre-requisite drivers would be bundled with the VOTES software package. However, the legal terms of re-distribution of these drivers would also need to be addressed beforehand.

Performance The performance reduction with the added connection code is significant, as establishing one connection with a server or database takes a non-trivial amount of time. This is the area that must be addressed most effectively, as any grid solution *must* be scaleable to very large numbers, with a minimal reduction of performance.

A possible solution here would be to establish only connections with a subset of the resources available, then record in a central list, available to the rest of the CVO, the fact that a particular server or database is recorded as live. However, this would also reduce performance through an increase in network traffic. Additionally the dynamic element would be lost, as connectivity testing necessarily occurs in real-time between primary sources.

As is evident from the discussion of these three areas, the system described still has issues that must be solved in order for it to be widely accepted as a grid

solution. We note that the VOTES project began in October 2005 and is funded for three years - hence it is natural that prototypes explore the problem space to discover issues in rolling out grid infrastructures in this domain.

4.3 Administration

Administration of clinical trials by means of the portal is another major issue that must be addressed when designing a re-usable framework that will effectively utilise grid technology. This is particularly pertinent to the process of patient recruitment where several actors of varying privileges must co-ordinate the details of a trial, which themselves will change over the trials lifetime.

The current VOTES implementation has several trials that provide test data to explore the various issues described above. However the next stage of development is to provide a separate administrative portal that will allow such trial coordination to occur on a CVO-wide basis.

The process we envisage is where a privileged user, likely the trial investigator, needs to create a trial and select from a list of available databases the information to be queried for this particular scenario. The databases will have a live schema querying and description service that will allow the investigator to browse the available information and discover the necessary resources.

From this, the investigator will then be able to assign access rights to specific roles within that trial and the CVO, thus populating data in the authorization access matrix (see next section). Due to the necessarily transient nature of CVOs, a time-limit on these access privileges will also be set at this point. This process would involve the trial investigator completing the following steps:

- Trial creation
- Selection of databases from those available in the CVO (using a descriptive browsing tool)
- View parameters/schemas within these databases and their descriptions
- Select parameters/schema elements relevant to this trial
- Define the roles to be used in this trial
- List the parameter/schema elements that each role can view
- Distribute these roles to known and trusted users

A complication here is the fact that a trial could be created with only a limited set of data shown from other data providers that do not necessarily trust the primary investigator of the trial. This type of privilege allocation would be referred to the overall static agreement and would necessitate a super-user that could monitor and arbitrate data sharing issues throughout the CVO. Another way of considering this is the trial investigator creating a trial will be the most privileged user, and can only issue roles and access to data sets that are within their level of privilege. It might well be the case however that the clinical data sets and schemas will be potentially far more extensive than a given trial investigator is permitted to see and allocate. This administrative portal has yet to be implemented.

5 Security Considerations

The main focus of this paper so far has been on the provision of a reliable service, which behaves in a stable and predictable manner by overcoming the challenges of non-hostile interruptions. However, a consideration that is of equal, if not greater, import is the protection against unpredictability due to disruption of hostile intent. Not only must the system be able to maintain an acceptable QoS but due to the potentially high sensitivity of the data involved, must have a rigorous security policy, which is adhered to in the strictest terms possible.

Grid security is often expressed in terms of AAA:

- *Authentication* the ability for a user to verify their identity to a resource.
- *Authorization* the ability to assign privilege to a user on a particular resource, once their identity has been positively established.
- *Accounting* the ability to unambiguously assign actions to users to allow non-repudiation in the event of a security breach.

In current grid technologies, the predominant method for enforcing authentication is to use the well-established technology of Public Key Infrastructures (PKIs) [20]. Authorization is an area that requires more sophisticated controls and therefore requires more research to give the necessary flexible and secure access control. Applications such as PERMIS [21] and VOMS [22] have been used in different contexts to research this area, making use of underlying paradigms such as role-based access control within a Privilege Management Infrastructure (PMI c.f. PKI) [23], or privilege delegation within a Virtual Organisations. No single authorization technology has yet been established as providing all the solutions for access control on the Grid.

In the current VOTES prototype, authentication is achieved through a username/password combination at the portal to the CVO, and with a second authentication step between the grid/data servers and the local database resources. A successful experimentation in this regard was to add Shibboleth [24] technology to this gateway, to allow a once-only authentication step to a Shibboleth federation [25]. This step also allows users to be mapped to roles within the portal, thereby de-coupling the users identity from the rest of the operation process, and thus making the system more maintainable. An extension to this is likely to be the use of PKI and digital certificates to verify the identity of users. Information on the application of Shibboleth in this domain is given in [26].

Authorization is achieved by enforcing role-based access at the CVO level. A database server is run at each node, known as the CVO Database which maintains a mapping between roles and privileges (the roles having been assigned in the previous step). Based on these roles, the user can see a specific view of the data, depending on their level of privilege. Figures 3 and 4 show screen shots of the different parameters that an *investigator* role can see, and those that a *nurse* can see. Note that the *nurse* view is limited to non-identifying data.

This differentiation between roles is achieved using an Access Matrix model in the CVO database applied to roles within the CVO for a given trial. The

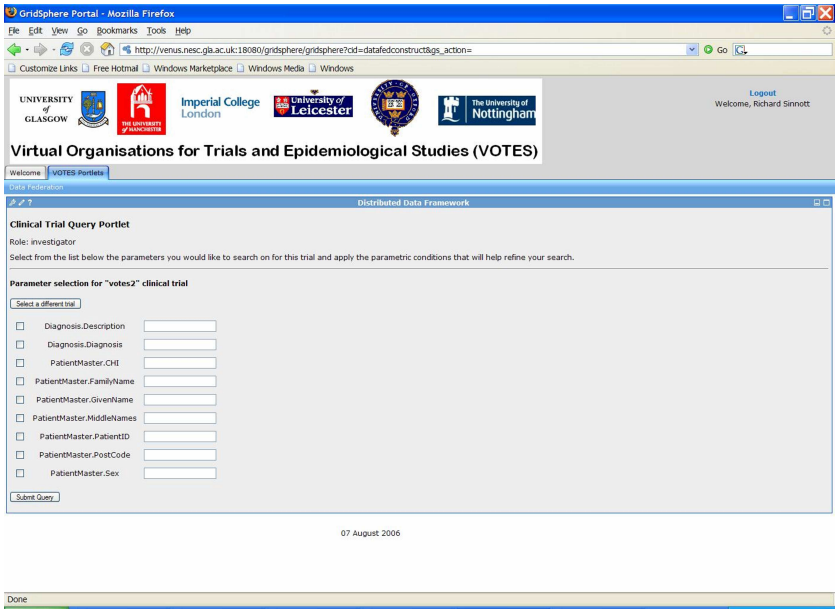

Fig. 3. The parameters shown are those that can be queried by the investigator role. Critically they allow identification of patients where those available to the nurse do not.

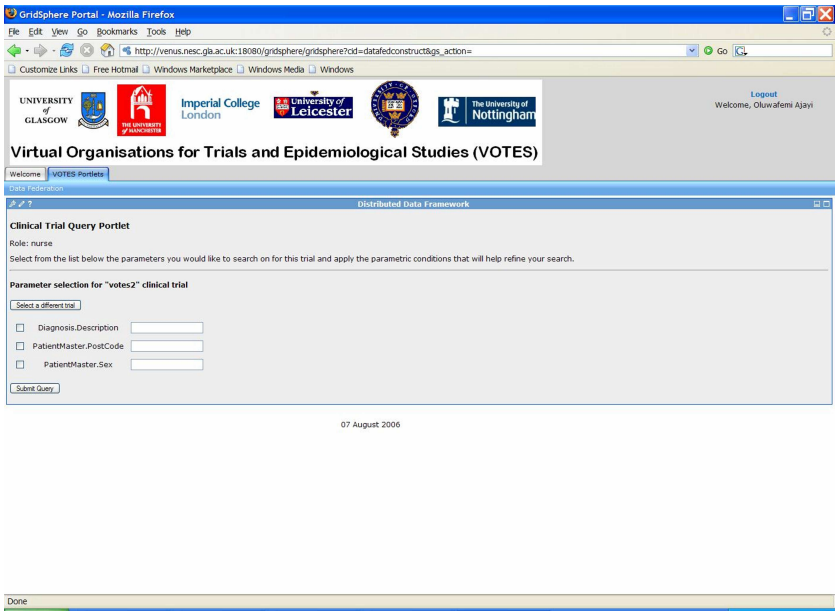

Fig. 4. The parameters shown are those that can be queried by the nurse role.

model is best represented using mathematical matrices of parameter versus role-name with a bitwise representation of access privilege (i.e. if a 1 is present in the value of an element in the matrix where a specific parameter meets a role-name, then that role has access to that parameter).

However the practical implementation of this matrix is most efficiently achieved using a database schema that holds information about trials, participating databases and their parameters. Figure 5 shows the Entity-Relationship diagram for this CVO database roles are defined in the context of specific trials, parameters belong to specific databases and the access privilege information is obtained by a view that matches the roles to the parameters through the primary tables. Previous work researching the application of access controls to data federation systems can be found at [27] and [28].

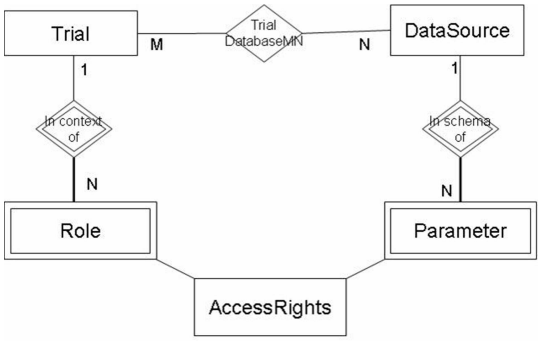

Fig. 5. ER diagram of CVO database

Accounting is an aspect that is usually implemented later in grid application development, as it is a secondary method of protection attempting to assert accountability after an incident has occurred. A rudimentary accounting mechanism has been implemented in the VOTES portal, maintaining the number and nature of the queries executed through the portal. A necessary extension to this will be the assignation of identity to the queries executed, using the digital certificates from the initial authentication step.

In addition to the common concerns of grid security, as has been mentioned when defining the CVO, are issues that relate directly to the process of information sharing in the clinical domain:

- *Anonymisation* the ability to identify records as unique entities but whilst still protecting the individual identity of the patient that the record pertains to.

– *Statistical Inference* the ability to prevent identification of individuals by combining two or more sets of non-identifying criteria to a query (e.g. a patient suffering from an unusual condition in a particular postcode).

A possible solution to these problems is the use of encrypting technology, such as PKI, and using an anonymisation service within the CVO, with a trust hierarchy that is analogous to the use of certificate authorities in PKIs. Non-disclosure of any result sets when only a certain number of records are returned is a common mechanism used to prevent potentially identifying data. More in-depth discussion of these issues can be found in [29] and [30].

Finally, the issue of patient consent is one that has direct implications on the security and reliability of such a system. In any initiative that involves such sensitive data, consent must be obtained from the patients about whom it concerns. This necessitates a point in the process that cannot under any circumstances be automated: where a patient must register with the appropriate authority that they give unambiguous consent to use a specific piece of their clinical data records.

In terms of security and reliability, the interface that the patient uses must therefore be user-friendly, securely implemented and protected against non-hostile interruption. To encourage uptake and confidence in the system, participants must feel that the infrastructure, into which these highly personal details are being input, is secure and reliable. Public understanding of the purpose, benefits, and potential drawbacks and dangers of clinical trials along with the non-technical information on how their data will be used is essential and is currently being explored in detail in several large scale projects across Scotland [31].

6 Conclusion

A reliable system that behaves in a predictable and secure manner is absolutely mandatory for any production enterprise to gain widespread acceptance this is especially so in the clinical domain. There are many solutions already in current IT practice that attempt to mitigate the risks posed by service interruptions of hostile or non-hostile intent, however with the move towards large scale international trials, open architecture based approaches are becoming ever more necessary. The additional challenges posed in using grid infrastructures are the need for flexibility and scalability, whilst still maintaining these rigorous levels of reliability and security.

While technology is fundamental to achieving this vision, arguably the greatest hurdle to be overcome is the human factor. With the implementation of grid applications across heterogeneous domains of differing structures and policies, the need for greater communication between parties and the absolute delineation of duties and responsibilities in a legal context is mandatory for progress to be achieved. Ultimately large scale infrastructures in the clinical domain depend upon trust: trust of people, trust of software and trust of practices.

The VOTES project has engineered a secure and robust grid infrastructure prototype. However, this is still very much a work in progress with new clinical requirements and new data sets arising, combined with changes in the grid technology and standards landscape. In time, the development of this infrastructure will hopefully meet the challenges outlined and be adopted on a global scale and help drive wider grid efforts in this domain.

6.1 Acknowledgements

The VOTES project is funded by a grant from the Medical Research Council in the United Kingdom. The authors would also like to thank the collaborators on the project and NHS Scotland for help in understanding the clinical data sets used in VOTES and the associated software.

References

1. Oracle, "Grid Computing with Oracle," Technical White Paper, 2005.
2. "UK Biobank Project." URL. http://www.biobank.ac.uk.
3. Garfinkel and Spafford, "Practical Unix Security," O'Reilly Second Edition.
4. Foster, Kesselman, and Tuecke, "The Anatomy of the Grid: Enabling Scaleable Virtual Organisations," in *IEEE Internet Computing*.
5. "Dynamic Virtual Organisations for e-Science Education." URL. http://www.nesc.ac.uk/hub/projects/dyvose.
6. "GOLD." URL. http://www.goldproject.ac.uk/.
7. "cancer Biomedical Informatics Grid." URL. https://cabig.nci.nih.gov/.
8. "PharmaGrid." URL. http://www.pharmagrid.com.
9. "PRobabilistIc Symbolic Model checker." URL. http://www.cs.bham.ac.uk/~dxp/prism.
10. "Health-Level 7." URL. http://www.hl7.org.
11. "Virtual Organisations for Trials and Epidemiological Studies." URL. http://www.nesc.ac.uk/hub/projects/votes.
12. A. P. Sheth and J. A. Larson, "Federated Database Systems for Managing Distributed, Heterogeneous, and Autonomous Databases," in *ACM Comput. Surv*, June 1990.
13. "GridSphere." URL. http://www.gridsphere.org.
14. "The Globus Toolkit version 4.0." URL. http://www.globus.org.
15. "Open Grid Services Architecture - Database Access and Integration." URL. http://www.ogsadai.org.uk.
16. "Scottish Care Information Store." URL. http://www.show.scot.nhs.uk/sci/products/store/SCIStore_Product_Description.htm.
17. "General Practitioners Administration System for Scotland." URL. http://www.show.scot.nhs.uk/gpass.
18. "Scottish Morbidity Records." URL. http://www.show.scot.nhs.uk/indicators/SMR/Main.htm.
19. "Virtual Organisations for Trials and Epidemiological Studies." URL. http://labpc-12.nesc.gla.ac.uk:18080/gridsphere.
20. "Public Key Infrastructures." URL. http://www.entrust.com/pki.htm.

21. "PrivilEge and Role Management Infrastructure Standards Validation." URL. http://sec.isi.salford.ac.uk/permis.
22. "Virtual Organization Membership Service." URL. http://hep-project-grid-scg.web.cern.ch/hep-project-grid-scg/voms.html.
23. D. Chadwick, A. Otenko, and E. Ball, "Role-based Access Control with X.509 Attribute Certificates," in *IEEE Internet Computing*, pp. 62–69, March-April 2003.
24. "Internet2 Shibboleth Architecture and Protocols." URL. http://shibboleth.internet2.edu.
25. R. Sinnott, O. Ajayi, J. Jiang, A. J. Stell, and J. Watt, "User Oriented Access to Secure Biomedical Resources through the Grid," in *Life Sciences Grid*, (Yokohama, Japan), October 2006.
26. R. Sinnott, J. Watt, O. Ajayi, and J. Jiang, "Shibboleth-based Access to and Usage of Grid Resources," in *IEEE International Conference on Grid Computing*, (Barcelona, Spain), September 2006.
27. S. D. Vimercati and P. Samarati, "An Authorization Model for Federated Systems," in *ESORICS 96: Proceedings of the 4th European Symposium on Research in Computer Security*, 1996 address = London, UK, owner = astell, timestamp = 2006.09.24.
28. K. Taylor and J. Murty, "Implementing Role Based Access Control for Federated Information Systems on the Web," in *CRPITS 03: Proceedings of the Australasian information security workshop conference on ACSW frontiers 2003*, (Darlinghurst, Australia), 2003.
29. R. Sinnott, A. Stell, and O. Ajayi, "Development of Grid Frameworks for Clinical Trials and Epidemiological Studies," in *HealthGrid 2006*, (valencia, Spain), May 2006.
30. A. Stell, R. Sinnott, and O. Ajayi, "Secure Federated Data Retrieval in Clinical Trials," in *IASTED Telemedicine 2006 conference*, (Banff, Canada), July 2006.
31. "Genetics and Healthcare Initiative: Generation Scotland - Scottish Family Health Study." URL. http://www.innogen.ac.uk/Research/The-Scottish-Family-Health-Study.