Designing Privacy for Scalable Electronic Healthcare Linkage

Anthony Stell, Richard Sinnott, Oluwafemi Ajayi, Jipu Jiang
National e-Science Centre
University of Glasgow
Glasgow, UK
a.stell@nesc.gla.ac.uk

***Abstract*—A unified electronic health record (EHR) has potentially immeasurable benefits to society, and the current healthcare industry drive to create a single EHR reflects this. However, adoption is slow due to two major factors: the disparate nature of data storage facilities of current healthcare systems, and the security ramifications of accessing, using, and potential misuse of that data. To attempt to address these issues this paper presents the VANGUARD (Virtual ANonymisation Grid for Unified Access of Remote Data) system which supports adaptive security-oriented linkage of distributed clinical data-sets to support a variety of virtual EHRs avoiding the need for a single schematic standard and the natural concerns of data owners and other stakeholders on data access and usage. VANGUARD has been designed explicitly with security in mind and supports clear delineation of roles for data linkage and usage.**

Keywords - privacy-preserving computing; system security; privacy engineering

I. INTRODUCTION

The healthcare industry drive to create a single unified electronic health record (EHR) is an inevitable consequence of the progress of the digital age. The benefits that it stands to bring to society are immeasurable: immediate, real-time access to patient histories, conditions, treatments will save many lives in those critical situations where split-second decisions can have a major impact on the medical outcome.

As with other communities that are attempting to create a unified description of application domain data, a number of standards have been posited and agreed upon. Some examples in the health domain are HL7 [1], OpenEHR [2] and SNOMED CT [3] – which are often intended as complementary entities (HL7 is a messaging protocol for exchanging health data, OpenEHR is the aggregation of that health data per-person, and SNOMED CT is a large-scale clinical dictionary). Unfortunately, in practice a lot of these standards have significant overlap and, driven by different open-source bodies, their development is not necessarily commensurate.

There are also a number of commercial vendors attempting to move into the space of unified health-care system support. The largest protagonists in this field include Google Health [4] and Microsoft HealthVault [5] who provide unified health record services to the general public (though currently still at prototype stage). There is also an open source version – Indivo [6] – which attempts to replicate the commercial functions of these vendors for free, but puts the emphasis on patient privacy and ownership of their own records.

The design and publication of standards is one thing, but actually implementing them in systems that will be used by ground-level healthcare staff is quite another. The experience of program development within the healthcare field has shown that there is a major disconnect between the standards promoted and publicized by management, and the systems in use in individual hospitals and practices. Generally, it is believed that this disconnect exists because of a lack of single-tier management structure within the healthcare community. The community itself is broadly structured as a patchwork of different institutions – either a variety of privately run commercial enterprises (such as the US model), or a combination of private companies and national health boards (as in the UK). Such fragmentation makes it difficult to direct technological strategy for healthcare on a national level. Certainly it appears that private companies with greater capital reserves are more able to adopt the standards mentioned previously but, as capital is unevenly distributed in society, this tends to only account for a privileged minority of any given population.

Clearly, for unified data standards to become a reality that can be applied to all patients currently in the healthcare system of any country, there is a strong need to interface with the patchwork of IT systems and data that has grown up as a result of such complex bureaucracy. Even if a single unified standard is adopted, there is sufficient legacy data and systems already in existence providing healthcare, that unifying standards must accommodate this legacy and deal with the transition to new unified standards and associated systems.

Complexity of data and systems is not the only issue with regard to the adoption of new standards however. Another development of the digital age – and one that people are often fearful of because of the unknown factors involved – is that of data security. The issues surrounding not just who can see your data, but what data is important, and how it can be used to compromise individuals, are very much in the public consciousness. A good example of this is the furore surrounding the Connecting for Health initiative [7] – an attempt to unify healthcare systems throughout the UK – which has so far met with more headlines than success. A major source of those headlines is the insufficient effort that has been

978-0-7695-3823-5/09 $26.00 © 2009 IEEE
DOI 10.1109/CSE.2009.323

put into working out the nuances of the security of this national system. Open access databases with coarse granularity, has led to widespread fear that the average patient record – and all the sensitive data therein – would be accessible by even the most junior of hospital staff.

With these concerns in mind, the National e-Science Centre (NeSC – www.nesc.ac.uk) has embarked on several projects involving access to and usage of clinical data in the healthcare domain using a novel system known as VANGUARD (Virtual ANonymisation Grid for Unified Access of Remote Data). VANGUARD attempts to address the issues of remote data linkage and security at a community defined peer-to-peer level. The methods and implementation involved allow relationships between systems to be established that add value to the patient record – which it did not have before the linkage – but also have the security of that relationship designed into the architecture. Critically, this technology does not preclude or position itself as an alternative to healthcare data standards that may be adopted in the future, but attempts to facilitate that adoption. VANGUARD has been designed from the outset to be a scalable technology, allowing many relationships to be built into networks of secure, linked data-sets "greater than the sum of their parts".

The technical basis of VANGUARD is the fine-grained layering of encryption on all data that is passed between parties through mediating agents, which have knowledge of data structures (data models/schema) but not necessarily of the actual data itself. With the requirements of limited visibility designed into the system from the beginning, data can be shared accurately and securely between the participating institutions without actually disclosing identifying information to anyone or any software component. We regard this system as supporting privacy by design.

The rest of this paper is structured as follows. Section 2 describes the healthcare context that VANGUARD has been developed in, along with the predominant data-sets in use in the UK. Section 3 covers the design and implementation of VANGUARD to date in the Virtual Organisations for Trials and Epidemiological Studies (VOTES) project. Finally, section 4 describes the outstanding issues in developing VANGUARD to become a viable health data product, along with the legal and security ramifications of the extended functionality required.

II. HEALTHCARE CONTEXT

The main requirements associated with healthcare data sharing discussed above were arrived at through the work conducted as part of the 3.5 year VOTES project [8].

A. *VOTES*

VOTES was a £2.8m initiative funded by the UK Medical Research Council (MRC), which attempted to bring e-Science and the power of building data grids to the clinical community for a variety of purposes, such as patient recruitment, enhanced data collection, and improved clinical trial management. The project began in October 2005 and recently completed in March 2009. The remit of the project was grand in scale and required the collation and understanding of many of the major healthcare data-sets and systems in use throughout Scotland and England. In Scotland these data sets and systems included:

- The Scottish Morbidity Records (SMR) [9] – which comprise a comprehensive record of inpatients, outpatients, cancer registration, mental health and psychosis, and death records throughout hospitals and practices in Scotland. The records go back several decades and are an authoritative source of a wealth of clinical information across Scotland. These data sets are also augmented and linked with census data available from the General Register Office (GRO) in Edinburgh.
- GPASS [10] - an administrative system used by 85% of general practitioners in Scotland. GPASS is a facility for managing all aspects of primary care patient data. GPASS also supports uploading of patient records to secondary care systems, e.g. to hospitals for further treatments and consultant referrals.
- SCI Store [11] – a central repository used by many hospitals across Scotland that is designed to manage all hospital data from inpatients and outpatients to lab data. SCI Store supports interfaces for periodic uploads of primary care patient data (from applications such as GPASS).

In England the main data-sets and systems encountered were:

- MIQUEST [12] - 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.
- General Practice Research Database (GPRD) [13] – one of the world's largest computerized databases of anonymised patient data from general practice. It comprises demographic, treatment, event, referral and outcome information.
- UK Biobank [14] – a long-term study investigating genetic predisposition and environmental exposure to the development of disease, by collecting data from 500,000 volunteers aged between 40 and 69. During the course of the trial disease events, drug prescriptions and deaths are all recorded.

As the VOTES project developed it became apparent that there is a greater degree of fragmentation and lower data quality in the systems and data used in England than of those in Scotland. This is partly a result of the priorities of the recently devolved parliament in Scotland, and also as a result of the high proportion of poor living conditions in Scottish urban areas (compared to England), which has required a highly proactive public healthcare campaign over the past decade.

The technical result of this higher quality data is a more complete and reliable data-set, referenced through a single index identifier – the Community Health Index (CHI number). A similar identifier exists in English healthcare databases – the NHS number – but this was often found to have patchier coverage, and was not necessarily unique beyond the realms of regional health boards.

A second consequence of this discrepancy in quality was the fact that the major healthcare data providers had direct involvement with VOTES whilst in England, the engagement and contributions were more difficult to source.

Despite this however, it should be noted that the overall state of healthcare data in Scotland is still far from complete and valid. The major vendor, SCI Store – an allegedly "standard" repository for nationwide secondary care (hospital data) and GPASS data – was reported to have 18 slightly different schema descriptions: one for each Scottish health board. Similarly, CHI numbers – though more reliable than their English counterparts – are still notoriously patchy in coverage (some residents have none, some have two, etc.) It is also the case that many hospitals and practices simply continue to use paper records, in reaction to the relatively short-sighted IT strategies implemented from above.

In terms of technology to integrate with and use healthcare data, the VOTES project developed and tested many systems and interfaces [15]. After trialing of several Grid solutions for data access and management, and integration with security technologies it was agreed that the best way to proceed was to use an in-house authorization system, which could be configured for use by the researchers and clinicians in VOTES.

The results of the technology investigation in VOTES, described in [15], were mixed at best. However, towards the end of the project the consolidation of these efforts and early prototype experiences resulted in the specification and implementation of the VANGUARD system. At the heart of VANGUARD is flexible data linkage (able to bring together many distributed parties) and anonymisation with owner-led security (through dynamically reduced data-sets based on stakeholder privilege). Several proof of concept systems applying VANGUARD have been demonstrated with major new projects also exploiting the *privacy by design* approach.

B. SHIP and Avert-IT

Two projects where the VANGUARD technology will be used with immediate effect are the Scottish Health Informatics Platform for Research (SHIP – www.scot-hip.org.uk) and the Advance Prediction of Adverse Hypotensive Events (Avert-IT) project (www.avert-it.org). The former is a three-year project to establish a research platform upon which a definitive electronic patient record for Scotland can be built. The data-set over which the linkage will occur are those Scottish patient records systems already identified through the VOTES project. Key to this is the use of the CHI number as the referencing index. SHIP has just started in April 2009, with technical papers due for the middle of the year.

The latter, Avert-IT, is a project to develop a "prediction engine" that will allow healthcare professionals to be alerted to imminent hypotensive events in brain trauma patients. Avert-IT is a European collaboration and again, the data-set over which the analysis and development will occur is linked with a single reference index, established from a previous data collection project – Brain-IT (http://www.brainit.org) – and linked at a local level for each participating site (more technical details of this project, which has been running for just over a year, can be found in [16]).

Both of these projects require the secure transmission and linkage of data between two parties through mediating agents, with the assumption that only limited trust exists between data providers, mediators and the end users of that data.

C. Global Datasets

The idea of providing a unified electronic health record is a globally accepted one, and as such, the authors have attempted to investigate the availability of equivalent data-sets in North America, and compare the issues to those in the UK. From collaboration with the caBIG (cancer Biomedical Informatics Grid) project [17], run by the US National Cancer Institute, the situation appears to be that of myriad data-sets all idiosyncratic to the major health insurance providers in the country. In Canada, there appears to be a collection of more canonical data sets (CCRS, DAD, HCD, etc) [18], managed and disseminated by the Canadian Institute for Health Information. It is tempting to assume that this discrepancy between the two countries is a function of the difference in provision of social healthcare – the US notable for being the only developed nation in the world to lack such an infrastructure. This also raises a question of the financial motivation for attempting to unify datasets (which is discussed in section 4 of this paper).

III. VANGUARD - PRIVACY BY DESIGN

In designing the VANGUARD system there were several requirements that the end customer (the UK Biobank team) specified as being paramount. These included the need for firewalls to remain closed to incoming traffic, hence the need for a "pull" model of communication for all component interactions with data providers and their systems. Furthermore from a security standpoint, a major requirement was that all data in transit, at rest, and outside RAM memory would have to use strong encryption to avoid potential eavesdropping by malicious third parties.

From the point of view of application validity, the major requirements were robustness in the event of component failures anywhere in the system; the use of a simple and intuitive user interface (UI), and the avoidance of proprietary platforms and code so that the application would have a long lifetime.

With these considerations in mind the following specification for the VANGUARD application was proposed.

A. Design

The main components and their functions for VANGUARD application are viewers, agents, guardians and bankers:

- **Viewers** are used to access potentially remote data sets (typically this is associated with a specific clinical research study that has been approved by an independent ethics body).
- **Guardians** protect the data resources being provided to the virtual organization.
- **Agents** mediate the exchange between the guardian and the viewer.

- **Bankers** maintain a record of all transactions that have taken place and limit resource data exchanges based on accountability information. (Though the banker will be an important component, its specification has yet to be outlined and as such, will not be expanded upon in this paper.)

B. Agreements and Threat Models

The overall agreement is assumed to be between the viewer and the guardian, so the agent is an untrusted entity that merely facilitates the joining of data requests to results. The agent can see the data headers and is aware of the guardian's schema but cannot see the data values within. Joining is performed on encrypted data points hence it is possible to join without knowing the underlying values themselves. Through the use of a single overall guardian key, the integrity of the data returned can be secured for the viewer. This also protects the user from establishing the audit trail of which data has come from which guardian. Only the agent has this knowledge, which it can then use to facilitate the banking component.

The major threats that this system attempts to protect against are as follows:

- Guardian impersonation (man in the middle)
- Unauthorised users (authorizing privilege levels appropriately)
- Decoding encrypted data payloads (use of asymmetric encryption on the query-specific hash can protect against this). This would potentially protect against users eavesdropping on agent-guardian interactions and decoding the method to identify what the sub-query is.
- Integrity of payload between guardian and viewer (use of an overall guardian key)
- Anonymity of payload between guardian and viewer (combinations of separate component keys)

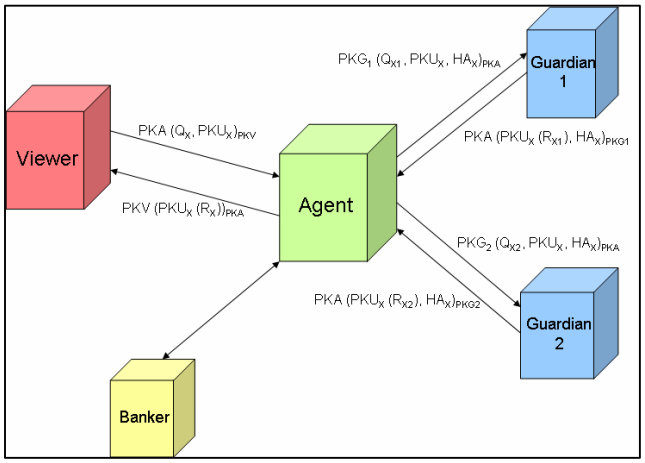

Figure 1: architecture diagram highlighting the messages passing between the VANGUARD components.

Figure 1 shows a diagram of the process. The convention of messages passed is encapsulating brackets for encryption, and trailing subscripts for signatures. The "x" represents a single user request/process. The subscript numbers represent where the query and results divide into sections.

The various process stages are as follows:

- The user selects a study through a viewer;
- The user is presented with data fields that can be queried in that study which have a visibility of either open, hashed or closed;
- The user makes a selection through the viewer and the request is sent along with the user's public key to the agent. The request is signed with the viewer's private key and encrypted with the agent's public key:
  - $PKA\ (Q_X, PKU_X)_{PKV}$
- The agent decrypts the request.
- The agent creates a federated query execution plan based upon the request received and divides the request into components to be forwarded to the relevant guardians (Q_{X1}, Q_{X2}, etc.)
- The agent also generates a unique query-specific hash to be attached to the guardian requests (HA_X).
- The agent puts all this information into the relevant guardian storage mechanism, signed by the agent's private key and encrypted with the guardian's public key:
  - $PKG_1\ (Q_{X1}, PKU_X, HA_X)_{PKA}$
- The guardians periodically check for requests (queries) they should respond to;
- The guardian pulls requests that they should act upon;
- The guardian decrypts and makes a locally defined authorization decision on the query, and when satisfied that it meets all local policy criteria on data access and usage, executes the query.
- The guardian returns the encrypted results (using the user's public key PKU_X) to the agent, signing it with its own private key (for the agent), the overall guardian signature key (for the viewer), and encrypting it with the agent's public key.
  - $PKA\ (PKU_X\ (R_{X1}), HA_X)_{PKG1}$
- The agent decrypts the individual package and joins with other components of HA_X that have been returned from the other guardians. We note that the agent is only able to see and join the hashed data set since the remaining data is encrypted with the user's public key. Once joining has taken place (the assumed process is a non-duplicated inner join, though this would be subject

to the parties' agreement), the information on how the join was made is removed thereby removing the possibility for further direct joining.

- The agent returns the linked and anonymised results to the viewer, signed with the agent's private key and encrypted using the viewer's public key.
    - PKV $(PKU_X (R_X))_{PKA}$
- The viewer decrypts the results data, while the user is subsequently able to decrypt the full, linked and anonymised result set using their own private key, and checks it's integrity using the overall guardian public key.

The ontology used for each participating site is maintained by the agent, updated by schema request updates from the guardians at regular intervals. The standardized representations of these parameters are presented to the user, with a fine-grained authorization layer between the agent and viewer to evaluate how the user sees the possible data points. The possible states are:

- **Open** – the data is available to this user.
- **Hashed** – the data is available in anonymised form to this user, and can therefore be used for statistical counting and aggregated information.
- **Closed** – the data is not available to this user but can still be used by the agent for joining (the relevant data points are stripped away before the result set is returned to the viewer).

Through this combination of layered encryption, signature and restricted authorization of standardized ontology mappings, the VANGUARD system effectively allows a unified federation of distributed data, whilst maintaining the strict agreement between the requesting user and the resource guardian.

C. Implementation

Version 1.0 of VANGUARD was developed to make use of the interface already used by the VOTES project. As such, the application is currently accessed using a GridSphere portal through a web browser. The resources behind the infrastructure are JDBC-enabled databases from several of the VOTES centres, and the encryption is performed using digitally-signed X.509 certificates.

The component communication is implemented using Axis2 web services, which effectively performs the Agent role. The Viewer and Guardian modules are implemented at the application level beneath the user-facing portal interface. The calls between the Viewer/Guardian and the Agent are as follows:

- Viewer to Agent:
    - *submitQuery (pku, query)*
    - *downloadResults (userID, queryID)*
- Guardian to Agent:
    - *downloadRequests ()*
    - *uploadResult (queryUserID, jobID, queryData)*

The ontology mapping is currently accessed through the application layer of all participating components (Viewer, Guardian and Agent). The mapping is constructed using a relational PostgreSQL database at each site, updated using SQL scripts whenever a new source is added or a site's policy is updated. However, investigations are underway to establish whether using XML and its ability to process semi-structured data-sets would provide more advantage in storing this mapping information.

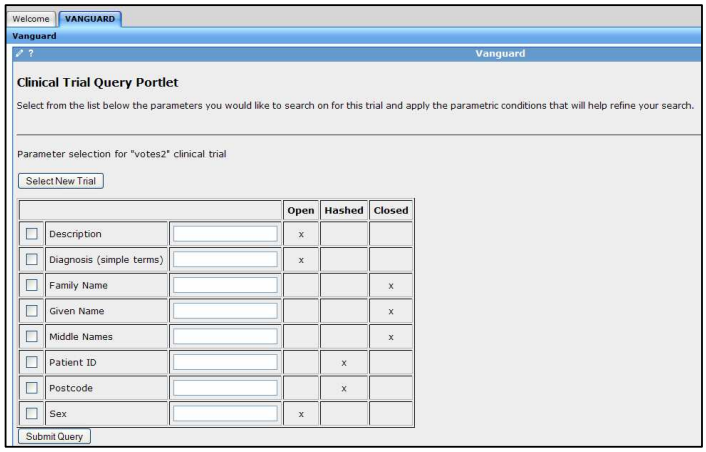

Figure 2: Example shot of viewer input screen, including parameter label, conditions, and availability (open, hashed, closed).

Identity and use of public key infrastructures is managed using Shibboleth [19], a single sign-on (SSO) technology that provides access to distributed resources throughout a pre-existing federation, but authenticated at local institutions. This hides a lot of complexity involved in PKI management from the end users and allows an established authority to maintain the identity assertion component of the system.

IV. Extending Usability

The development of VANGUARD so far has focused on the security requirements inherent in unifying healthcare data, and attempting to address those requirements at the application design stage. However, in order for the application to be widely adopted, and to give it a functionality that truly addresses the needs of the community – whilst maintaining security – there are a number of additional issues, which currently still need to be addressed, and are not yet implemented.

A. Beyond the CHI

In the first instance, VANGUARD works across single data domains because the presence of a single unified index is assumed. In Scotland this is the CHI number; in England the NHS number. For a country such as the UK – made up of

several nation states (England, Scotland, Wales and Northern Ireland), each with differing levels of self-governance – it is possible to maintain a record of how these indexes relate to each other (though only through simple allocation matching, rather than any complex analytical relationship).

In this case, where movement between various health domains is politically easy, it is desirable to maintain this index relationship from a primary care standpoint – allowing effective treatment by knowing the patient's medical history, no matter where in the country that history has been built up.

On a global scale, the issue of maintaining such inter-domain relationships would appear to be less important. Most health data joining is not primarily to track individual patients, but to aggregate overall statistical health trends. Typically this information is reported per-country, and with migration being generally hard to achieve, knowing individual patient histories becomes an issue internal to the country involved.

It should also be noted though, that most countries have a political structure which maintains a relationship between regional and federal governance. In a country such as the United States, the issue of matching domain index relationships becomes incredibly complex, as these are based on the patient's presence within a particular healthcare insurance provider, and such data is not necessarily forthcoming. The issue of only having access to private health-care brings up the financial value of such data. Closely aligned to concerns on privacy of data linkage, there is potentially a strong business driver against such linkage and aggregation of health data. Whether this is to the benefit of the wider population depends on the corporate responsibility of the healthcare insurers involved.

B. Secondary Matching

A further development in the integration of distributed healthcare records would be to remove the dependence on a single primary index discussed above. The first step in doing this is to match patients on criteria other than their identifying index. An example would be to identify a condition, within a specific postcode – in many cases this would return a sufficiently small number of records to allow identification. Combined with a probabilistic calculator of the match confidence and it would be possible to use a few key indicators to identify individuals.

The security and privacy ramifications of this however, are manifold. Such cases are examples of statistical inference, which is a notoriously difficult process to protect against, if used by malefactors. This type of matching would also be of great interest to all governments looking to more efficiently track their citizens as part of the post-9/11 drive to prevent terrorism. The short jump from such technology falling into the hands of governments intent on implementing repressive regimes is not inconceivable.

Therefore, the identification of what data should be available is a mandatory procedure, and has given rise to groups such as the open source Indivo project, which places the emphasis on patient ownership of data and has a philosophy of releasing data only if absolutely necessary to the provision of care.

C. Ontology

Key to the technical development of VANGUARD is the existence of a mapping schema that describes the individual datasets and matches them to the peer datasets. This already exists in the current implementation, but in a static format held in databases by both the Viewer/Guardian and the Agent.

The required development here is a separate module that can automatically inspect a new dataset, translate that into an XML document for use in VANGUARD (along with the privacy level of each parameter), upload that document to the application and have it published to the rest of the components a short time afterwards.

Central to the success of VANGUARD is that no single standard is adopted, but these ontologies simply exist on a peer-to-peer basis. However, the ability to maintain a structure such as this requires a separate module in itself, which will need to manage the inherent performance issues when the structure reaches a large scale.

D. Performance/ Scalability

Maintenance of the ontology structure is one area affected by performance issues, but the strong security provided by VANGUARD is of potentially greater importance in this regard. Multiple encryption calls have a cost in performance, which will only increase as the number of peer-to-peer calls increase. A possible technical solution to this would be the ability to outsource the encryption action to a component of the system residing on (or with access to) a more powerful hardware component. However, it does not solve the resulting bandwidth required by inflated communication calls, and it also adds an additional layer of complexity (which is against the tenets of secure application design).

Though this is an issue that should be solved before adoption, there is a strong argument to suggest that in the immediate term it is not a major problem, given the asynchronous batch processing nature of the VANGUARD system. As such, the drive to increase performance will not impact directly on the user experience of VANGUARD.

E. User Interface

This is not a new problem, but key to software adoption is the experience of the user – is the interface and operation intuitive and easy to follow? Is there a minimum of new processes to learn? Does it follow the businesses processes that the user is already used to? All these concerns are solved by carrying out good requirement capture and adhering to these throughout the development cycle.

A key additional question in the health domain is often whether to stay with legacy interfaces or not. The driver tends to be that the learning process will be so much quicker if the program operates in the same way as the tools that were previously used. However, the competing concern is that the design of legacy applications is often so poor that much compromise in terms of security and efficiency has to be made.

Also important in the UI regard is the management of complex PKI credentials. Shibboleth and related federated technologies are relatively established methods of addressing this, though their widespread adoption is yet to be total.

F. Legal and Social Ramifications

A finding of the VOTES project was that irrespective of the technical capabilities of the system implemented, an over-riding agreement between the parties involved that is signed and understood *a priori* is essential. This is particularly important when dealing with the potential ramifications of security breaches or data protection violations. Within these agreements it must be made explicitly clear who has what privileges; who is able to see what data sets; where potential conflicts of interest exist and how to make efforts to prevent system abuse.

Also for consideration is the potentially great impact that a technology such as VANGUARD can have socially. Such a system would be of interest to government agencies and malefactors alike, particularly if the ability to perform secondary matching with high accuracy were achieved. Such considerations are part of the "privacy erosion" issue, which legislation such as Title II of the Health Insurance Portability and Accountability Act (HIPAA), attempts to address.

G. Extended Implementation

The expression and implementation of these additional issues will be investigated as part of the SHIP project throughout the coming year. In the first instance, secondary matching will be implemented using data points other than the central index. The provision of more sophisticated ontology development will follow, along with investigations into the performance impact on simulated production systems.

CONCLUSION

In this paper we have presented a schematic for the VANGUARD system, in an attempt to address the issues of security in linking healthcare data sets, from ground-level design. The design and architecture has followed a set of requirements provided by major customers in the sector, and has been implemented in a test environment. The additional considerations for turning VANGUARD into a viable production system available for widespread adoption have been discussed, along with the legal and social implications and impact on patient privacy. Future work in the context of the SHIP project will develop these ideas and provide further insight into the security issues surrounding global electronic health records.

REFERENCES

[1] HL7 – http://www.hl7.org

[2] The OpenEHR Foundation – IOS Press, 0926-9630, Regional Health Economies and ICT Services: The PICNIC Experience, 2005, p153-173

[3] SNOMED Clinical Terms: overview of the development process and project status – Stearns, Price, Spackman, Wang, American Medical Informatics Association, Proc. AMIA symposium 2001; 662-666

[4] Google Health – http://www.google.com/health

[5] Microsoft Healthvault – http://www.healthvault.com

[6] Indivo: a personally controlled health record for health information exchange and communication – Mandl, Simons, Crawford, Abbett, BMC - Medical Informatics and Decision Making 2007, 7:25

[7] Will Connecting for Health deliver its promises? – Michael Cross, British Medical Journal, 2006; 332:599-601

[8] Development of Grid Frameworks for Clinical Trials and Epidemiological Studies – Sinnott, Stell, Ajayi, Proceedings of HealthGrid conference 2006, Valencia, Spain

[9] Quality of Scottish Morbidity Record (SMR) data – Harley, Jones, PubMed 8936810

[10] GPASS - http://www.gpass.scot.nhs.uk/

[11] SCI Store - http://www.sci.scot.nhs.uk/products/store/store_main.htm

[12] MIQUEST - http://www.connectingforhealth.nhs.uk/systemsandservices/data/miquest

[13] GPRD – http://www.gprd.com

[14] UK Biobank – http://www.ukbiobank.ac.uk

[15] Technical Challenges in Leveraging Distributed Clinical Data – Stell, Sinnott, Ajayi, Proceedings of IASTED Telehealth conference 2008, Baltimore, USA

[16] Federating Distributed Clinical Data for the Prediction of Adverse Hypotensive Events – Stell et al., Philosophical Transactions 'A' of the Royal Society of London, 2009

[17] CaBIG - https://cabig.nci.nih.gov/

[18] Canadian Institute for Health Information – http://www.cihi.ca

[19] Shibboleth – http://shibboleth.internet2.edu