Enabling Secure, Distributed Collaborations for Adrenal Tumor Research

Anthony STELL, Richard SINNOTT, Jipu JIANG
National e-Science Centre, University of Glasgow, UK

Abstract. Many e-Health strategies rely on the secure integration of datasets that have previously resided in isolated locations, but can now in principle be accessed over the Internet. Of paramount importance in the health domain is the need for the security and privacy of data that is transmitted across these networks. One such collaboration, which spans several specialist centres across France, Germany, Italy and the UK, is ENSAT – the European Network for the Study of Adrenal Tumors. The rarity of the tumors under study means the value of accessing, aggregating and comparing data from many centres is great indeed. However this is especially challenging given that ENSAT require clinical and genomic data to be seamlessly linked, but in such a way that the information governance, ethics and privacy concerns of the patients and associated stakeholders involved are visibly satisfied. Key to this is the clear separation of clinical and genomic data sets and support for rigorous patient-identity protecting access control. This is especially challenging when such data sets exist across different organisational boundaries. In this paper we describe a prototype solution offering a security-oriented tailored portal supported by a layered encryption-driven linkage technology (VANGUARD) that offers precisely such patient-privacy protecting capabilities. We describe the architecture, implementation and use to date of this facility to support the ENSAT adrenal cancer research network.

Keywords. Data Privacy, Grid Portals, Public Health Informatics

1 Introduction

As national and international e-Health strategies gain pace, the desire to link data-sets of specific clinical areas over the Internet becomes ever more prevalent. In domains of rare conditions and rare diseases in particular, there is a strong motivation to pool resources and integrate data-sets that have previously remained predominantly isolated.

The European Network for the Study of Adrenal Tumors (ENSAT – www.ensat.org) is one such group. With centres in France, Germany, Italy and the UK, this group have hitherto only previously shared medical information on adrenal tumors by sending CDs to each other through conventional mail. The security problems inherent in these methods are well understood, but only now are the networked alternatives being realised. With the advent of next-generation security-oriented applications running over the Internet, it is now feasible to consider solutions that allow connection of clinicians and specialists resources in real-time, without the overhead in time and cost of postal-based solutions and with the improved security that such a networked solution brings.

It was with this motivation that a prototype online platform for ENSAT was created, which allows the collection and sharing of data on four different and rare types of adrenal tumor: Adrenocortical Carcinoma (ACC); Aldosterone-Producing Adenoma

(APA); Non-Aldosterone-Producing Adrenocortical Adenomas (NAPACA), and Pheochromocytomas and related paragangliomas (Pheo).

The current ENSAT platform at its heart comprises a security-oriented, user-personalised portal that includes a registry application allowing access to a store of clinical adrenal tumour data sets from the ENSAT partners. This includes for example clinical descriptions of the tumours, their treatments and responses to such treatments. The design and functionality of this facility is based greatly upon a previously successful implementation of another pan-European online registry: the EuroDSD project (www.eurodsd.eu). EuroDSD collects, shares and analyses data on the genetic and biochemical information behind disorders of sexual development in children – another very rare condition benefiting from secure, real-time collaboration on a networked system [1].

The highly sensitive nature of the medical information concerned makes it imperative that the clinical data is accessed, used and transmitted between sites in a secure manner. However, looking further ahead the ENSAT group wish to move beyond the clinical data registry to support richer data linkage scenarios including linkage of clinical data sets with genomic data sets, and ultimately to support large scale clinical trials of new drug treatments targeted specifically to adrenal tumours. Through projects such as EuroDSD we have shown how it is quite possible to bring data together and join/link it in meaningful ways to fully-authenticated and authorised end users exploiting federated access control with advanced authorisation solutions. However moving beyond to support inter-organisational, clinical-genomic research demands novel solutions for data access and linkage where the clinical information and the genetic descriptions of particular individual’s cases are clearly separated and the subsequent linkage between the two is not disclosed to unauthorised individuals. We show how the ENSAT research environment is being augmented with a back-end encryption and anonymisation-oriented joining mechanism, VANGUARD – the Virtual ANonymisation Grid for Unified Access to Remote Data [2] to tackle such clinical-genomic data integration challenges.

This paper describes in more detail the clinical context in which the ENSAT platform has been built to address; the informatics challenges associated with supporting such research; the overall architecture of the proposed solution, the implementation details of both the current prototype and the plans for future development and usage for enhanced biobanking scenarios.

2. Background Context

2.1 Adrenal Tumors

The European Network for the Study of Adrenal Tumors was founded in 2002 through the merging of three existing but largely independent adrenal tumor research networks in France, Germany and Italy, with research teams from the UK. The central aim of the ENSAT consortium is to improve the prediction and management of malignant tumors. It is hoped that the study of the genetics and treatment of adrenal tumor patients will reveal new molecular mechanisms of tumor growth and provide insight into the role of peptides and steroids in hypertension in general.

The diversity and often aggressive/fatal nature of adrenal tumors make them an important condition to address. However, their comparative rarity requires many international resources to be drawn upon in order to make significant progress in the field. Given this, a long-term goal of ENSAT is to bring together a critical mass of expertise and resources to achieve significant clinical and biological conclusions and to eventually combat adrenal tumors.

The clinical basis for adrenal research can be summarized as follows: the adrenal consists of two functionally distinct endocrine glands, both parts playing an essential role in adaptation to major stress. In response to severe infection, trauma and shock, the organism responds with activation of the adrenal gland, leading to enhanced cortisol and catecholamine secretion. This response is necessary to maintain homeostasis during stress. In case of adrenal impairment, which can be caused by tumor, the stress response will often be severely diminished, which in turn may lead to irreversible shock. Adrenal tumors can be benign (non-cancerous) or malignant (cancer). Often this separation is difficult to make and long term close follow up is necessary after removal to detect recurrences early in patients who have adrenal cancer.

The ENSAT consortium is especially interested in four major types and causes of adrenal tumors:

- Adreno-Cortical Carcinoma (ACC) – this is a rare malignancy, of which pathogenesis and prognosis is incompletely understood. Patients with ACC will usually be identified with a hormone excess or a local mass effect (for example, with a median tumor size at diagnosis of greater than 10cm). Post-operative disease-free survival of patients diagnosed with ACC is below 50% over a 5-year period.
- Aldosterone Producing Adenoma (APA) – is a type of cancer which results in secondary hypertension accounting for up to 5-10% of all hypertensive patients.
- Non-Aldosterone Adreno-Cortical Adenomas (NAPACA) – Non-aldosterone secreting cortical tumors represent the most common benign adrenal tumor. These may be non-functioning – not associated with any hormonal excess and usually detected in patients undergoing radiological investigations, such as ultrasound scanning. Autopsy studies show that up to 5% of the population may harbor so-called adrenal incidentalomas. However, in some rare cases patients may have a genetic problem that results in autonomous production of cortisol from adenomas within the adrenals.
- Pheochromocytomas and related paragangliomas (Pheo) - Catecholamine-producing tumors may arise in the adrenal medulla (pheochromocytomas) or in extra-adrenal chromaffin cells (paragangliomas). The tumors themselves may be sporadic or part of any of several genetic diseases including familial syndromes. About 10% of these tumors are malignant either at first operation or during follow-up.

2.2 Data Distribution and Security Requirements

Tumor information as described above form the mainstay of the work of the ENSAT consortium. Their details are diverse yet rare, making isolated repositories highly valuable in clinical terms, and creating a correspondingly high motivation for data-sharing. However, a parallel effect of this rarity is to make the identity, and hence security, of the patient involved inherently tied to the same information points. These are the competing technological and informatics tensions that the registry and associated e-Infrastructure – including potential data linkage outside the consortium – must address directly.

There are many initiatives, which encounter similar problems. Some examples are the SAIL databank [9] – a linkage of national clinical data-sets across Wales, the UK Office of National Statistics Virtual Microdata Laboratory [10] and the ESRC Secure Data Service (SDS) hosted by the UK Data Archives [11]. These projects attempt to solve the problem of inferential security by introducing "safe havens" – physical environments which are strictly controlled and require physical presence to access and use data. Whilst maintaining security, this approach does not allow the flexibility required to fully utilize the distributed network technology and essentially allow inter-organisational data linkage. Rather it promotes a data silo model, albeit a secure one.

2.2.1 Data Structures and Security

In an environment such as ENSAT looking at specific conditions (c.f. nationally collected general practice records), data can be linked deterministically. This is largely due to the streamlined history of the group's foundation and development, where well-defined indexes were decided upon from the start and applied consistently throughout the consortium's lifetime. However, the structure of the data presents issues in terms of multiple form inputs, provenance of these inputs, and how to present them in the final user interface (this structure is described in more detail in section 3.1). It is also a consideration for future implementations of such solutions, that the development of probabilistic matching on secondary data points (such as height, weight, body mass index (BMI), etc.) will allow non-deterministic linkage between data-sets that are related but not matched through any canonical index or identifier. The potential usefulness of this solution would be great indeed, but the commensurate potential for misuse and disclosure of inferred information would also be large.

Furthermore, genetic data is by its very nature identifying. Given this, techniques are needed to provide access to and usage of such data sets without directly making all of the data available, or directly linking it with the clinical data sets themselves. This is especially challenging when multiple organisations are involved and clinical and genetics centres are working together (as is the case with ENSAT).

2.2.2 Data Access and Usage Requirements

The ENSAT consortium is a collaboration of partner groups, but all partners are members of institutions (health boards, hospitals, universities), that are ultimately responsible for the enforcement of good conduct by their employees. With this tenet in mind, secure access to the ENSAT platform is demanded. It is quite possible to establish username/password access to a portal (most portal frameworks provide such capabilities directly), however a better solution is to exploit finer-grained access control in a user-oriented framework. The Internet2 Shibboleth solution provides such a solution. In this model, users authenticate at their home institution and signed assertions (including digitally signed user privileges/roles are used to determine the access privileges to, and inside the ENSAT platform itself) are returned to the service provider (portal) which after verification and validation of these assertions, uses this information to configure the contents of the portal – so called security-oriented personalisation of the research environment. This model has several advantages including the fact that the home institution knows the user best, and will likely act as the first point of revocation of privileges if the user transgresses their policy in any domain related to this consortium.

2.2.3 Legislation Requirements

As the data is situated in several countries, it is subject to separate legislation in terms of transit, identification (anonymised vs. pseudonymised records), and can often have restrictions in terms of where the data itself can physically reside. This is especially so with genomic data sets. Transported using regular query streams into one database, data from France or Germany for example may well be stored physically in a database residing in the UK, and as a result, the ramifications of which legislative body governs the data becomes very important. This can be overcome if the data itself remains at the local provider where federated access to distributed data sets is supported. Many health care providers are especially wary of incoming (ingress) connections through hospital firewalls and the dangers of subsequent data usage. Ideally the data should remain at the local data holder but be accessible subject to ethics and security policy agreements being in place.

3. ENSAT Architecture

The ENSAT platform has been designed with the challenges and contexts discussed above in mind. As a typical n-tier web application, the description of the system can be broken down into the business logic, data and presentation layers.

3.1 ENSAT Database and User Interface

A key part of the registry development is the agreement on a core set of data related to each of the research areas (ACC, APA, NAPACA and Pheo). The ENSAT consortium has developed individual data models for each of these areas previously. Each of the data sets follows a similar broad structure, outlined in table 1.

Identification	*Filled-in once*
Patient History	*Filled-in once*
ENSAT Workup	*Filled-in at diagnosis*
Tumor Form	*Filled-in after operation*
Sample Form	*Filled-in by teams with a Biological Research Centre*
Follow-up Form	*Filled-in at (yearly) follow up*

Table 1: List of tables and usage shared by the four databases

The fields associated with each of these forms define a detailed structure of the information required for collection and sharing. It is noted that most centres have adopted World Health Organisation (WHO) coding systems for many of the disease specific and associated drug related data sets. An example of the ACC Patient History form currently collected is shown in table 2.

Patient History	
Year of diagnosis of ACC	*YYYY*
Previous malignancy	*Yes/no, tick one*
	If yes, specify (free text)
Cushing's syndrome	*Yes/no, tick one*
Virilisation	*Yes/no, tick one*
Menstrual disorders	*Yes/no, tick one*
Feminisation	*Yes/no, tick one*
Hypertension	*Yes/no, tick one*
Diabetes	*Yes/no, tick one*
Hypokalemia	*Yes/no, tick one*
Abdominal pain	*Yes/no, tick one*
Palpable abdominal mass	*Yes/no, tick one*
Venous thrombosis	*Yes/no, tick one*
Incidentaloma	*Yes/no, tick one*
Others	

Table 2: Data fields associated with the patient history table

The only data set that has any identifying data that relates to the patients themselves is the Identification Form. The contents of this form include a local unique identifier (c.f. an ENSAT-wide registry identifier) that is associated with the patient at the clinical centre itself and the contact details for the corresponding clinician. Clinicians are responsible for recording this identifier and its subsequent usage for further information when requested. The generated ID is the only link to the patient's identity and requires a mapping file residing at the patient's hospital to assert that link. This satisfies the anonymisation requirements of nearly all clinical institutions as it is in line with patient-protecting legislation (such as the UK Data Protection Act (1998) [5]).

3.2 ENSAT Distributed Access

The ENSAT prototype system has been protected with a Shibboleth-based single sign-on as part of a federated access management infrastructure. The main purpose behind such a setup is to allow access to multiple distributed (federated) resources (called service providers) with the use of a single set of credentials from a trusted identity provider (IdP) from the federation. Within a ring-fenced domain such as ENSAT - and any data-sets that would be potentially linked to in the future - this appears to be a scaleable solution that provides convenience, but also carries the major security benefit of requiring current privilege assertion by the user's home institution. Shibboleth and similar initiatives are well-documented and more information on the implementation and results can be found at [4].

3.3 ENSAT Layered Encryption

In order to successfully join distributed clinical and genetic data-sets novel solutions are required which offer more than secure data access and encrypted data return to the data requestor. The VANGUARD has been developed to meet the anonymisation-oriented nature demanded by data providers and the often pragmatic concerns of data providers themselves. VANGUARD has been described in more detail in [2,3,4]. In brief, VANGUARD offers *Viewers* (to request and receive data); *Agents* (to mediate information exchange); and *Guardians* (to protect access to data). VANGUARD allows data to be linked, joined and anonymised through these components and targeted use of encryption and hashing keys. It has been developed specifically to meet the demands of clinical organisations such as the NHS. Thus for example, the VANGUARD solution does not mandate that direct queries through hospital firewalls is supported. Rather it is based on a pull-oriented data access and usage model. Similarly, through the VANGUARD approach and use of encryption keys for key components, all data is encrypted and the only entity capable of accessing the actual (decrypted) data is the end user. Further linkage of the linked, joined and anonymised data sets through VANGUARD is not possible. In the VANGUARD architecture, a central mediating agent must have a working, current knowledge of the schemas residing at each data location. It must be able to join these to a degree of accuracy useful to a user hoping to derive value from the data "greater than the sum of its parts". However, the relationship in any joining process, very often does not require that mediating agent to view the data itself, and a major driver in adoption of linkage technologies is the retention of local control by the original data owner. Therefore VANGUARD provides a solution, which enables that joining process without having seen with the actual contents of the data stream, and allowing local data custodians to exercise restrictions over what is released, with full knowledge of what context it will be used in.

In VANGUARD, several overlapping trust agreements drive the interaction: the agent is trusted by all parties to execute policy but not to see data; the guardian, in a specific context, may trust the viewer to see the data; and the guardians do not trust each other with their own local datasets. Joining itself is performed on hashed and encrypted data points hence it is possible to join without knowing the underlying values themselves. To access external non-consortium resources, a mapping can be made of encrypted indexes against the external identification index (e.g. bio-bank barcode), which protects identity whilst allowing traversal of domain boundaries. 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, allowing them to enforce overall policy and accountability.

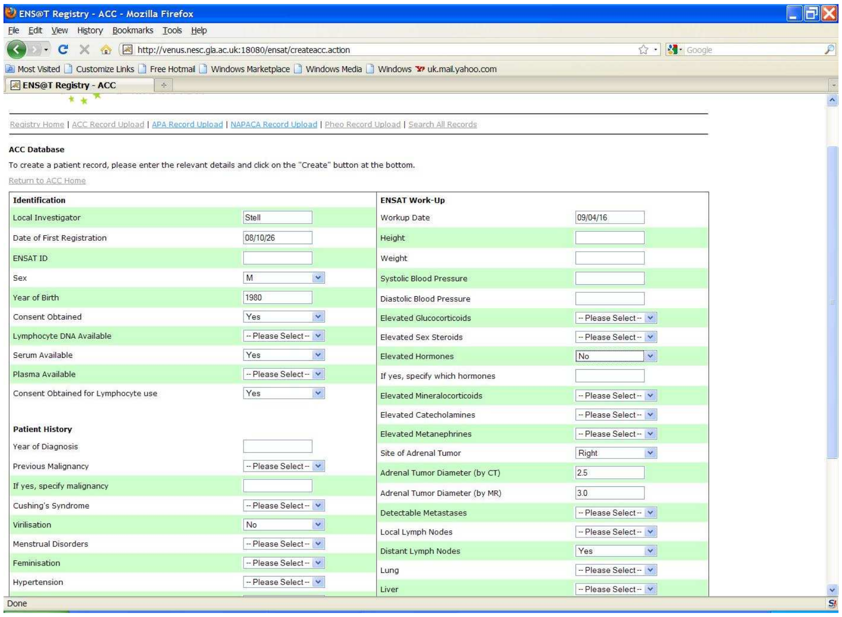
3.4 Case Study Example

In the case of ENSAT, the following example serves to illustrate how VANGUARD can be used as a tool in a wider Virtual Research Environment (VRE), to augment the basic function of the registry with bio-banking capabilities. A typical scenario of the exploitation of VANGUARD in the context of ENSAT is as follows:

- A clinician wants to discover patients with specific ACC biomarkers. This might be as part of a particular clinical trial. They log in to the ENSAT portal and based on their roles/privileges sees the associated Viewer (see Figure 1) that allows selection of a set of ACC specific parameters used to search the ENSAT registry. In addition to the registry specific parameters, the Viewer also supports selection of a family of particular biomarkers that may be available from the Bioinformatics Research Centres (BRC) associated with the ENSAT consortium. It is important to note that the BRC and the ENSAT partners themselves are typically independent organizations.

- The clinician selects which data-sets they want to interrogate and the ACC markers they are interested in. Assuming that the clinician has the necessary privileges in ENSAT (as securely presented to the portal through the targeted Shibboleth Identity Provider hosted at the National e-Science Centre in Glasgow), the result of the interactions that take place between the Agent and various Guardians involved in VANGUARD, allows the clinician to see those patients that match both the ACC registry clinical information and have the associated genetic information (biomarkers). The linkage between the patient information as present in the ENSAT registry and the BRC biomarkers themselves are removed before delivery to the Viewer (by the Agent) before subsequent decryption and review.

A key element of this scenario and the primary justification for VANGUARD is that the genetic markers for specific patients themselves are never disclosed or linked directly with clinical information. Different data Guardians (for the biomarker data resources) demand strict delineation of data governance issues. Rather the fact that a given ACC genetic marker exists and is associated with a particular de-identified clinical record is the only information that is ever released. However, VANGUARD assumes a common identifier to be able to link and anonymise data. Currently this identifier is generated through the registry when adding a clinical case. This unique identifier is not directly accessible through the Viewer but can be selected as a joining field. This identifier is also added to the BRC genetic data management system as part of the barcode through which the samples themselves are indexed.


ENS@T Registry - ACC - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://venus.nesc.gla.ac.uk:18080/ensat/createacc.action
Google
Most Visited Customize Links Free Hotmail Windows Marketplace Windows Media Windows uk.mail.yahoo.com
ENS@T Registry - ACC
Registry Home | ACC Record Upload | APA Record Upload | NAPACA Record Upload | Pheo Record Upload | Search All Records
Return to ACC Home
Done

ACC Database

To create a patient record, please enter the relevant details and click on the "Create" button at the bottom.

Identification	
Local Investigator	Stell
Date of First Registration	08/10/26
ENSAT ID	
Sex	M
Year of Birth	1980
Consent Obtained	Yes
Lymphocyte DNA Available	-- Please Select --
Serum Available	Yes
Plasma Available	-- Please Select --
Consent Obtained for Lymphocyte use	Yes
Patient History	
Year of Diagnosis	
Previous Malignancy	-- Please Select --
If yes, specify malignancy	
Cushing's Syndrome	-- Please Select --
Virilisation	No
Menstrual Disorders	-- Please Select --
Feminisation	-- Please Select --
Hypertension	-- Please Select --

ENSAT Work-Up	
Workup Date	09/04/16
Height	
Weight	
Systolic Blood Pressure	
Diastolic Blood Pressure	
Elevated Glucocorticoids	-- Please Select --
Elevated Sex Steroids	-- Please Select --
Elevated Hormones	No
If yes, specify which hormones	
Elevated Mineralocorticoids	-- Please Select --
Elevated Catecholamines	-- Please Select --
Elevated Metanephrines	-- Please Select --
Site of Adrenal Tumor	Right
Adrenal Tumor Diameter (by CT)	2.5
Adrenal Tumor Diameter (by MR)	3.0
Detectable Metastases	-- Please Select --
Local Lymph Nodes	-- Please Select --
Distant Lymph Nodes	Yes
Lung	-- Please Select --
Liver	-- Please Select --

Figure 1: Screen-shot of upload screen for the ENSAT registry prototype. The biomarkers and associated BRC resources are available through this form.

The result of running a VANGUARD-enabled query crossing the clinical registry and the BRC genetic data sets will be a set of matched patients with specific biomarkers and associated matching clinical information. The researcher will typically then formally apply through ethics for access to further information and/or access to and use of the biosamples themselves. This aspect is done outwith the ENSAT registry and VANGUARD. However we note that further information on the transfer of biomaterials is made available through the ENSAT research environment including what samples were sent where and the amount of samples that are available. One way that this can be considered is that the ENSAT environment offers a secure data access and bio-matchmaking research environment.

4. ENSAT Implementation

The implementation of this work has consisted of three phases. The first phase was to set up a database and UI to ascertain the needs of the clinicians and how the registry should look and feel. This was successfully completed. The second phase was to develop the VANGUARD linkage and joining mechanism. At present we have established this on an isolated test-bed, using representative dummy genomic data. The third and final phase of this work is to integrate these two components and ultimately move to a full production service.

The database and user interfaces are constructed using MySQL and the Struts2 framework. This provides the benefits of the Model-View-Controller paradigm (MVC), captured by combining Hibernate, Spring and Struts (version 1) to allow more powerful manipulation of data access objects, whilst minimising the amount of coding work required to achieve such goals. This initial prototype of the ENSAT registry was created within this framework and can be found at http://ymir.nesc.gla.ac.uk:28080/ensat.

The VANGUARD component of the system has been implemented on a test-bed platform using the Java Glassfish web service framework. This plugs in directly to the Netbeans IDE and is secured using the WSIT encryption scheme (Web Service Interoperability Technology) [7].

The next and final phase of the work will be final testing and subsequent release of the ENSAT platform as a full production service. In practical terms, the Viewer front-end will exist in the ENSAT portal directly, whilst the client applications – to be installed on the Guardian machines – will be distributed to the various centres consuming the service. The tokens required to encrypt and sign the interactions will be compiled into the Guardian client applications before distribution. This method of client-server interaction is preferred because of one common requirement of distributed clinical access: hospitals generally will not allow incoming connections (or even sanction in specific circumstances) to their firewall.

5. Conclusions

The ENSAT online platform is a practical clinical tool that allows the sharing of data on rare conditions throughout Europe, between centres that are greatly distributed in terms of geography, but securely connected using the Internet backbone. We have detailed a solution, currently in the prototype phase but set to enter production in the coming weeks, which addresses the needs of these clinicians, but also goes one step further in assuring the security of data-sets that are distributed and require joining, without assuming the full trust of any intermediate joining agents or individuals.

Initial feedback from use of the prototype system has been positive. Focus has been largely on the presentation layer with much feedback on the data validation requirements and expectations including automated conversion between units used.

A key component of the platform itself, not documented here, are the tools that are used to confirm the encryption performance and capabilities, e.g. tcpdump, switching on http.output in Glassfish logfiles, etc. All of these have been used to confirm the data linkage and anonymisation capabilities of VANGUARD. Further security testing will include attempting to subvert the encryption through various eavesdropping techniques and attempting to extract meaningful information by various side-channels. Furthermore, demonstration of the secure, anonymisation-driven data linkage to the ENSAT protagonists and their associated ethics and legal bodies will be a key aspect of the work as a whole when it moves to full production mode. This aspect is currently on-going in the UK in the NeSC Scottish Health Informatics Platform (www.scot-ship.ac.uk) with the NHS where information governance is very much at the forefront of data provider demands.

Performance of data transmission will be a key part of the production system, i.e. to ascertain whether it is necessary – given factors such as volume of data, or regularity of use – to streamline the encryption used, perhaps by use of homo-morphic encryption schemes (which allow addition or multiplication of data points whilst encrypted, and still give accurate results once unencrypted). With the extension of the joining ability to include probabilistic matching techniques, whilst still under encryption, the integration and roll-out of this system will provide massive potential for many future applications in the e-Health domain.

Further work that will be undertaken as part of ENSAT is supporting seamless access to wider, non-ENSAT related research resources. This includes microarray data resources such as ArrayExpress [12] and GeneCards [13] and allowing seamless upload to such resources when deemed appropriate by the researchers.

References

[1] Jiang, Sinnott, Stell, Watt, Ahmed – Towards a Virtual Research Environment for Paediatric Endocrinology across Europe, Proceedings of CCGrid 2009, Shanghai, China

[2] Stell, Sinnott, Ajayi, Jiang – Designing Privacy for Scalable Electronic Healthcare Linkage, Privacy and Security (PASSAT) '09 Conference, Vancouver, Canada

[3] Sinnott, Ajayi, Stell – Data Privacy by Design: Digital Infrastructures for Clinical Collaborations, International Conference on Security and Privacy, Orlando, USA 2009

[4] Sinnott, Stell, Ajayi, Young – Towards a Virtual Anonymisation Grid for Unified Access to Remote Clinical Data, Proceedings of the 6th International HealthGrid conference, Chicago, USA, 2008

[5] UK Data Protection Act 1998, Office of Public Sector Information - http://www.opsi.gov.uk/acts/acts1998/ukpga_19980029_en_1

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

[7] Web Services Interoperability Technology (WSIT) - https://wsit.dev.java.net/

[8] World Health Organisation – International Classification of Diseases (ICD) - http://www.who.int/classifications/icd/en/

[9] The SAIL Databank: building a national architecture for e-health research and evaluation – BMC Health Services Research 2009 9:157, http://www.biomedcentral.com/1472-6963/9/157

[10] Secure access to confidential microdata: four years of the Virtual Microdata Laboratory – Economic and Labour Market Review, Vol 2, No 5, May 2008

[11] UK Data Archives, Secure Data Service (SDS), http://securedata.ukda.ac.uk/

[12] EMBL-EBI, ArrayExpress, http://www.ebi.ac.uk/microarray-as/ae/

[13] Lancet D, Safran M, Olender T, Dalah I, Iny-Stein T, Inger A, Harel A and Stelzer G. GeneCards tools for combinatorial annotation and dissemination of human genome information GIACS Conference on Data in Complex Systems April, 2008