How a clinical information system can support life science research

Jimison Iavindrasana*, Adrien Depeursinge*, Rodolphe Meyer*, Stéphane Spahni*, Antoine Geissbühler*, Henning Müller*a,b

*a Service of Medical Informatics, University and Hospitals of Geneva (HUG), Switzerland
*b Business Information Systems, University of Applied Sciences Sierre, Switzerland
1 http://www.aneurist.org
2 http://www.ogsadai.org.uk

Summary
Management of the cerebral aneurysm can be improved by establishing an IT infrastructure capable of integrating all available knowledge related to the disease. This paper presents the infrastructure implemented at the University Hospital of Geneva to support the @neurIST project. The infrastructure permits sharing of patient data stored in the clinical information system for secondary use and integration of the results derived. The requirements, basic functionalities and security aspect related to patient privacy and clinical information system security are presented. Its main advantages are reusability, scalability and lower maintenance cost while safeguarding patient privacy and clinical information system security.

Introduction
With the wider availability of modern imaging solutions it is possible to diagnose more and more unruptured cerebral aneurysms. Their rupture has fatal and disabling consequences with high societal costs, but the incidence of rupture is low. Recent decades have brought impressive developments in various domains such as determination of the associated risk factors, imaging, rupture prediction based on haemodynamic simulations and treatment of this complex disease. Current knowledge of cerebral aneurysms spans length scales from molecular, through cellular, to tissue, organ and patient representations. However, current management of the disease is based on very limited information such as the size and localisation of the aneurysm. Information technology can contribute valuably to improving research on cerebral aneurysms and their management, by integrating all exploitable information and computational services. The goal of the @neurIST project is to provide a new IT infrastructure to manage, integrate and interrogate the data related to the disease. The @neurIST is an Information Society Technologies (IST) Integrated Project funded as part of the European Commission’s (EC) Sixth Framework Programme.

Five pilot clinical centres are participating in the @neurIST project and are acting as patient recruiters and clinical data providers. Multi-institutional data collection furnishes a statistically significant number of cases in a shorter time period, while making it possible to incorporate a more diverse study population and reduce the bias induced by any individual researcher. The electronic patient record is used as initial data input and the data can be used for more than the treatment of a single patient.

In a multi-institutional prospective research setting, the participating patients’ electronic medical records can be anonymised and exported on a centralised repository as in [1] or stored at each site in an anonymised form as in [2, 3]. In the second case, the data are stored in a repository disconnected from the hospital information system. In the first model the basic research results are also stored on the central repository. In the second the research results are stored where they were produced; a mediation program links all data sources according to the end-user query.

The second architectural solution provides more flexibility with respect to data management, each data producer being responsible for the maintenance of the distributed data source. However, evaluation of the effectiveness of a new therapy or disease management tool takes place inside each clinical centre. To reduce the cost of data management inside the clinical centre and provide researchers with real-time data, the clinical information system (CIS) can be used as a data repository for both clinical management of the patient and for research. In this way each participating patient’s electronic medical record is not duplicated in a specific data repository and all research results related to a patient are directly stored in the clinical information system. Later this involves other issues concerning patient data privacy and confidentiality, as the patient data leaves the control and protection-sphere of medical secrecy and the security of the hospital information system itself.

A first architectural design was proposed in [4] to deal with access to data stored in a CIS for secondary use. In this paper we present the requirements and basic functionalities of first implementation of the architectural solution chosen by the University Hospital of Geneva for
data management in the @neurIST project (section 2 and 3). Patient data security and the hospital information system are also discussed (section 4).

Method
The implementation described in this paper was designed in compliance with legal and technical constraints and based on use case scenarios proposed by the end users.

Architecture requirements
The architecture described in this paper provides two functional cycles: 1) share clinical data produced in the clinical centre with their potential users, i.e. the researchers and/or external computation services; 2) store back research results related to participating patients. Four main requirements guided the design of the architecture: data categorisation, patient identification, data integration and data communication and security.

In the @neurIST project two categories of data are produced: clinical and research data. The CIS is used only to store the first category and research data related to one patient. However, the latter may be voluminous (such as simulation files) and only the metadata serving to reproduce the result may be stored back in the CIS. Moreover, other voluminous research data are produced during the experimental process and are not reproducible without the necessary materials. These results are stored where they were produced but can be linked with the patient data by means of a strong identification mechanism. As a patient has the right to access his electronic health record, the derived data produced by researchers should be identified by other means than the clinical patient identifier. Further, the patient’s identifiable personal information should be removed or reduced to the minimum required before secondary use. This means that all data leaving the clinical context should be depersonalised and that a mechanism to map the two identities should exist (one-to-one relationship). Data integration is defined as the problem of combining data residing at different sources and providing the user with a unified view of these data [5]. In a data integration problem, all data sources are hidden to the end user and viewed as a single data source; the data representation is also unified so that the provenance of the data cannot be identified. As the primary data are collected inside the clinical centre, they are coded with the local terminological system. Normalisation of the data, i.e. their conversion into an agreed and unified format, is done on-the-fly when the data leave the hospital boundary. De-normalisation is also carried out in the same process.

Within the @neurIST project, GRID and Service-Oriented Architecture (SOA) were chosen as the underlying technologies of the whole IT infrastructure [6]. The elements of the architecture are composed of a set of web services compatible with GRID technology. A detailed description of the security aspect is provided in section 4.

Basic functionalities of the implementation
The architecture implemented to support the @neurIST project has three layers (see fig. 1): the public data service located in the hospital’s demilitarized zone (DMZ), the private data service located in the hospital’s intranet and the CIS located in the hospital’s intranet as the data source. The mechanism for querying data stored in the CIS is presented in the following paragraphs.

The public data service is the entry point to the data stored in the CIS. It is located in the hospital’s DMZ and thus accessible from open networks such as the Internet. It authenticates the user or the application querying the database and all transactions are monitored and logged. A hospital information system is a closed one and the public data service cannot communicate directly with the CIS. It instead queues all authenticated and authorised incoming queries in a repository. This component is based on the GRID middleware OGSA-DAI [3].

The private data service is responsible for recovering all queued incoming queries located in the DMZ. Received queries are transformed to reflect the internal data structure and representation (mediation and de-normalisation) and sent to the CIS. A novel feature of the current architecture is the private data service’s ability to mediate the reception of data generated outside the hospital for incorporation into the CIS – a process which mandates the ability to re-identify the patient. Thus a query may be either a request for data or a notification that data is available for integration. The private data service has two important resources for these steps: the translation rules for the normalisation and de-normalisation service and the ID database which also contains the patient’s consent preferences. This
latter policy is required to control access to the patient's data: only the @neurIST data related to patients registered in the ID database can be queried. The patient data returned by the CIS (query results) are depersonalised and pseudonymised on-the-fly by the private data service. The query results are also filtered by this component. When the results are normalised – ie transformed into an agreed representation – they are written down in a result repository located in the DMZ and can be retrieved by the client. A further access policy consideration is that derived data are not viewable or accessible by unauthorised users (which may, for unverified results, include the clinicians and/or patients) until reasons and methods for release are favourably reviewed by the internal project ethics committee. These accesses are subject to both the clinical centre's policy and the patient's decision on whether they agree or not to the return of health research relevant results.

Security and privacy
The CIS maintains and manages personal medical records in a digital format, containing the first instance information relating to its subjects’ current and historical health, medical conditions and medical tests. Various measures are taken to protect not only the privacy and confidentiality of patients participating in the project but also for the security of the whole hospital infrastructure.

Pseudonymisation
Motivated by the need to store back derived data into the CIS (ie to re-identify the patient), the pseudonymisation procedure was preferred instead of anonymisation of patient data for secondary use. The pseudonymisation procedure is split into two steps, as shown in figure 2. In the first step, all information permitting direct identification of the patient in structured, and unstructured data such as images are removed or reduced to the minimum required to carry out the research. This step is called depersonalisation and data minimisation. In the second step a pseudonym is generated from the patient ID and attached to the depersonalised data, and is reversible to re-identify the patient. The re-identification of the patient from the pseudonym is necessary for query management but is also needed by the ethics committee to recontact the patient when an important finding is obtained from research. It is important to notice that the pseudonym generated is specific for a participant recruited and followed up at a single clinical centre.

Access control
Even if the data accessible through the infrastructure does not permit direct identification of the patient, it should be protected against unauthorised access. Generally, clinical centres have their own security, access right management and privacy protection policy according to the role of the user [7]. However, these security models are designed for local use and should be combined

Figure 1. Access to @neurIST data stored in the clinical information system.

Figure 2. Pseudonymisation system of outgoing data.
with distributed security models to authenticate and authorise external users. Within a security domain all the security is concentrated and placed under the responsibility of this domain. Between different security domains the chosen approach is to designate, in each domain, a security entity (Security Token Service), which will be in charge of issuing and verifying short-term security tokens with the entities of the other security domains.

Accessing an operational CIS may also affect its operational status if the query itself raises a security issue. Incoming queries are analysed to prevent such incidents and only valid queries are executed on the CIS. Access control is also strengthened by control of outgoing results and can be considered complementary to privacy protection. Such filtering technologies can assist in preventing unintentional disclosure of personal identifying information due to problems such as unidentified flaws in the depersonalisation component. A policy for statistical queries may, for example, reject result sets which contain less than four entries to be exposed to the requesting user.

Logging and monitoring

Another important component in the architecture is the logging and monitoring of external requests to the public data service and onwards to the CIS. Even if the CIS has a logging and monitoring mechanism, it is imperative that all queries to retrieve or store data are logged and monitored, in order to identify and intervene in potential misuse of the system. This process is managed at the public data service level.

Discussion and conclusion

In this paper we proposed the first implementation of an infrastructure for better management of the cerebral aneurysm. The infrastructure serves to share clinical data acquired inside the hospital and integrate derived data produced by researchers. The main advantage of the implementation is its reusability, scalability and lower maintenance cost. Most of the implementation is generic and can be adapted for other disease management. Patient security and privacy and CIS security have been discussed. Implementation of an infrastructure of this kind requires open access to the CIS from the technical and political point of view. The next step is integration of the infrastructure into the whole @neurIST IT infrastructure.

Acknowledgements

This work was generated in the framework of the @neurIST Integrated Project, which is co-financed by the European Commission through the contract no. IST-027703.

References