Title: *EHCR: An EU Provenance Case Study*

Author: *Tamás Kifor, László Z. Varga, Javier Vazquez-Salceda, Sergio Álvarez, Steven Willmott*

Version: *0.3*

Date: *17th November 2006*

Status: *Public*

Summary

The distributed nature of healthcare institutions sometimes hinders the treatment of patients, because documentation of the healthcare history and therapy of a patient is split into independent healthcare institutions. In order to provide better, user-centered healthcare services, the treatment of a patient requires viewing the processes and data as a whole. In healthcare systems there is a need to provide an integrated view of the execution of treatment processes, to analyze the performance of distributed healthcare services, and to be able to carry out audits of the system to assess that, for a given patient, the proper decisions were made and the proper procedures were followed.

The PROVENANCE project receives research funding from the European Community's Sixth Framework Programme

Members of the PROVENANCE consortium:

- IBM United Kingdom Limited United Kingdom
- University of Southampton United Kingdom
- University of Wales, Cardiff United Kingdom
- Deutsches Zentrum fur Luft- und Raumfahrt e.V. Germany
- Universitat Politecnica de Catalunya Spain
- Magyar Tudományos Akadémia Számítástechnikai és Automatizálási Kutató Intézet Hungary

The PROVENANCE project receives research funding from the European Community's Sixth Framework Programme

1 Introduction

Cooperation among people using electronic information and techniques is an increasingly common practice in every field, including healthcare applications. In the case of distributed medical applications, the data (containing the healthcare history of a single patient), the workflow (of the procedures carried out on that patient) and the logs (recording meaningful events in those procedures) are distributed among several heterogeneous and autonomous information systems. These information systems are under the authority of different healthcare actors such as general practitioners, hospitals, hospital departments, etc. which form disconnected *islands of information*.

The distributed nature of healthcare institutions sometimes hinders the treatment of patients, because documentation of the healthcare history and therapy of a patient is split into independent healthcare institutions. In order to provide better, user-centered healthcare services, the treatment of a patient requires viewing the processes and data as a whole. Although agent-based cooperation techniques and standardized electronic healthcare record exchange techniques support the semantic interoperation between healthcare providers, we still face the problem of the reunification of the different pieces of the therapy of a single patient executed at different places. Currently there are some countries that have no unification method for patient healthcare records; each region in the country or even each institution inside a region may have its own medical record system, sometimes not even fully electronic, and with no automatic health care record exchange mechanisms. Therefore, it is not uncommon for doctors to depend on the patients themselves in order to include data from previous treatments and tests.

Making electronic systems *provenance aware* enables users to trace how a particular result has been arrived at by identifying the individual and aggregated services that produced a particular output. In healthcare systems there is a need to provide an integrated view of the execution of treatment processes, to analyze the performance of distributed healthcare services, and to be able to carry out audits of the system to assess that, for a given patient, the proper decisions were made and the proper procedures were followed. All of these tasks depend on being able to trace back the origins of decisions and processes, the information that was available at each step, and where that information came from. Note that the provenance of a piece of data is primarily about the causal dependencies of execution steps, although time sequences can also be handled. The provenance architecture of focuses on making service-oriented systems provenance aware, but making healthcare systems provenance aware needs additional techniques, because healthcare actors are autonomous and, unlike in service oriented systems, they may participate in the same process without directly interacting with each other.

2 Overview of Application

The EHCR architecture we are implementing provides a way to build electronic health records as well as a unified view of a patient's medical record. The architecture provides the structures to build a part of or the full patient's healthcare record drawn from any number of heterogeneous databases systems in order to exchange it with other healthcare information systems. It uses the ENV13606 pre-standard, which defines the messages, the retrievable objects, the healthcare agents and the distribution rules. Current healthcare systems in many countries work by storing master copies data about individual medical treatments on a patient at the place where the treatments are carried out. Most commonly a single GP oversees a patient's medical history and thus integrates treatments not carried out under his/her own supervision post event. However there is no standard process for forwarding medical details which might form part of the record to a central registry or a master copy a particular patient's record. Information is retrieved from different healthcare providers on the basis of the patients Identity Number (ID). A healthcare provider A may only ask for record information from another provider B

The PROVENANCE project receives research funding from the European Community's Sixth Framework Programme

for a patient X if the patient X is physically being treated at A. Usually there is no central health authority database that could be relied upon to have a complete medical history.

In order to pull together the medical history of a patient we have essentially three options:

1) Build a system mirroring the current one based on fragments of records in different places which can be pulled together to produce a unified view on demand (depending on the permissions of the viewer).
2) Build a system of a more centralised nature with a master record which can be read and written to by authorised healthcare providers (in a controlled fashion) and possible cached at a particular healthcare provider.
3) Build a hybrid system which stores fragments of data with providers but records high level events in a central master record.

In all cases the interchange protocol could be one of the European pre-standards. Our approach follows the third option above.

3 Security Issues

When we extend healthcare systems with provenance in order to provide better services for patients, then we face new privacy issues in addition to those already handled in the healthcare information system.

Organizational and technical measures help to protect the privacy of the patients in usual healthcare information systems. In these systems the patient and the medical data are stored in EHCR management systems and transmitted between these systems. The separation of data and the different kinds of access control techniques protect the identity of the patient. The anonymity of medical data allows controlled and irreversible disclosure for different purposes mentioned earlier. In these systems the data is completely under the control of the actors comprising the distributed system and data sharing is controlled by the agents.

When we want to increase trust in data and to increase the quality of medical services in distributed medical applications by introducing provenance concepts, we introduce new privacy risks as well. We introduce the provenance store into the systems. In order to be able to trace how a particular result has been arrived at by identifying the individual and aggregated services that produced a particular output, the actors of the system must entrust information to the provenance store. This way healthcare actors give up part of the control over the data and the autonomy of the healthcare information is shared with the provenance store which is then able to link data and the workflow pieces that generated the data.

One of the problems of healthcare information systems is that there are information islands. While the healthcare data exchange standards help the information exchange between these islands, the provenance system helps the integration of the islands. This raises additional privacy risks.

The purpose of using a provenance system in the healthcare applications is to be able to trace back the processes that happened whenever an audit is needed to verify, e.g., the chain of decisions made for each case, or the compliance of a treatment with respect to the related regulations. However there may be a conflict between provenance and privacy. While for provenance we need as much information as possible about the whole process (*who* did *what* and *when*) to be able to trace back all that has happened, for privacy we need to restrict as much as possible the information available in order to avoid identification of patients and practitioners by unauthorised users.

The use of distributed provenance stores to register all relevant information in a distributed medical information system poses two main risks:

1) *cross-link risk*: the risk that unauthorised users are able to link some piece of medical data with an identifiable person by cross-linking information from different sources.
2) *event trail risk*: the risk to be able to identify a person by connecting the events and actions related to that person (e.g., the hospitals he has visited in different countries).

The PROVENANCE project receives research funding from the European Community's Sixth Framework Programme

4 Distribution Issues

We introduced two techniques to reduce the cross-link risk: a) we do not put medical data in the provenance store that can be easily used to identify the patient, and b) we anonymise the patient data. When mapping the provenance architecture to the OTM application, we decided not to store sensitive medical data in the provenance store, but only references to such data. This way the provenance store contains only the linkage and the skeleton of the provenance of the medical data, and the healthcare data can be laid on the skeleton by retrieving it from the healthcare information system when needed. The retrieval is done by an EHCR system which is completely under the control of EHCR access rules. With this approach we keep the same privacy degree of medical data as in the original system. Moreover we also minimalise the amount of transferred data.

One might think that if we do not store medical information about patients in the provenance store, then there is no need to anonymise the patients and we can use real patient identifiers in the provenance store, because no medical information can be inferred on the patient. However this is not the case. Even the fact that the patient was treated, can be sensitive information, because it may increase the event-trail risk mentioned in the previous section, and also because the reference to the place where the medical data of the treatment was carried out may refer to the type of treatment the patient received. The type of institution can reveal the type of medical intervention for example because if the institution is specialised on heart diseases, then the reference to this institution reveals that the patient was treated with heart problems. Therefore at least the patient identity has to be anonymised.

The anonymisation process has to satisfy two requirements as described in the following. In order to be able to create a complete and interlinked process documentation of the treatment of the patient, if two sets of p-assertions are related to the same patient, then there should be a way to link anonymised patient identifiers referring to the same patient in the different sets of p-assertions. In order to protect the privacy of the patient, the anonymisation procedure should be irreversible in the sense that nobody should be able to tell the real identity of the patient by knowing the anonymised identifier in the provenance store. As a consequence of this, no component in the system should store the real identifier of the patient and its anonymised identifier together.

In the OTM application, the EHCR subsystem applies case identifiers (identifiers created at run-time) as tracers to make connections between sets of p-assertions. The case identifier is anonymous, because it does not contain the identity of the patient. Using this identifier in the provenance store is similar to the anonymisation method of EHCR systems and yields similar privacy degree within the provenance documentation of one case.

When we want to connect different cases, we have two choices: a) we store two or more case identifiers in any of the p-assertions and define relationships between the case identifiers, or b) we use a global anonymous tracer of the patient and connect each case identifier to that anonymous tracer in p-assertions. The first solution corresponds to the current practice where the doctor knows (for example from the patient) that the current case is related to a previous one. In this case the doctor connects the cases with a relationship p-assertion in an ad-hoc way, therefore it is not recommended. The second solution helps to connect those cases as well, that are not explicitly known, therefore we followed this approach.

In the second solution we send the case identifier and the public identifier of the patient to an authorisation agent who is responsible for all authorisations in our systems. The authorisation agent generates from the public identifier of the patient (such as national insurance number) an anonymous tracer, called Global Medical Patient Identifier (GMPID[1]) and makes a p-assertion in the provenance store connecting the case identifier with the GMPID. This way all case identifiers, which are about the

[1] We do not detail here how the GMPID is generated. We assume that the algorithm satisfies the general anonymisation rules of EHCR systems. The patient identifier and the GMPID are never stored together, and they are present at the same time temporarily only in the authorisation agent when the GMPID is generated.

The PROVENANCE project receives research funding from the European Community's Sixth Framework Programme

same patient, will be connected to the GMPID of the patient in the provenance store. When we query the p-assertions that relate to the same patient, we can use any of the case identifiers, because case identifiers related to the same patient are linked together. This way the authorisation agent connects the different identity domains together.

When we want to retrieve the medical history of the patient, then we ask from the provenance store where and when the patient was treated, and then we have to query all the other medical information directly from the hospitals where the patient was treated because they are not stored in the provenance store.

5 Conclusion

In addition to the ability to return the whole process documentation, the method described above has two other advantages as well. One is that the actors can improve the quality of the process documentation and the other is that they can improve the quality of their own activities.

The quality of the process documentation can be improved if the actors discover some relationship from the real processes (e.g. the current illness of the patient is a consequence of a problem in the previous treatment not discovered before) then they can improve the process documentation by augmenting the already existing naming service based links with the direct causal relationships present in the process. This is now possible, because the p-assertions relevant to a single object can be located and linked to following the links.

The actors can improve the quality of their own activities using the linked process documentation, because when an actor executes its treatment process, it can already retrieve the p-assertions of other actors and might be able to discover a relationship from the provenance information. For example if the physician knows the details of the previous treatment of the patient, then he/she might use that information in the current treatment.

The PROVENANCE project receives research funding from the European Community's Sixth Framework Programme