Tamás Kifor,
László Z. Varga,
Sergio Álvarez,
Javier Vázquez-Salceda,
Steven Willmott

Privacy Issues of Provenance in Electronic Healthcare Record Systems

Introduction (1)

- advantages of agent based techniques in healthcare information systems: coordination, personalization, dynamic, decentralized, etc.
- but: indivisible healthcare history and therapy of the patient is allocated to independent and autonomous healthcare institutions
- reunification of the different pieces of the therapy of a single patient executed at different places is based on ad-hoc methods and the information provided by the patient

Introduction (2)

- provenance of electronic data in service oriented architectures: enable users to trace how a particular result has been produced by identifying the individual and aggregated services that produced a particular output
- organ transplant application of the Provenance project: we propose the usage of provenance techniques to provide better healthcare services for patients by providing a unified view of the whole health treatment history

Introduction (3)

- As long as the treatment and the data are distributed among the agents of the healthcare information system, privacy protection is focused on the protection of partial information pieces
- with the introduction of provenance into the system we re-integrate the different pieces
- our goals:
    - investigate the privacy aspects of introducing provenance into healthcare information systems
    - propose methods against the new types of risks

Distributed and Heterogeneous EHCR Applications (1)

- fragmented and heterogeneous data resources and services forming islands of information
- the corresponding workflow chunks are distributed among these islands of information
- the treatment of the patient might require viewing these pieces of workflow and data as a whole

Distributed and Heterogeneous EHCR Applications (2)

- ENV 13606 pre-standard developed by CEN/TC251 (European Committee of Normalisation, Technical Committee 251) is vital for the exchange of clinical data
- EHCR architecture defines how to exchange data, but the linking of the workflow pieces which generated the data is not discussed in EHCR standards
- provenance architecture helps to document the way the data was created and link the workflow pieces together

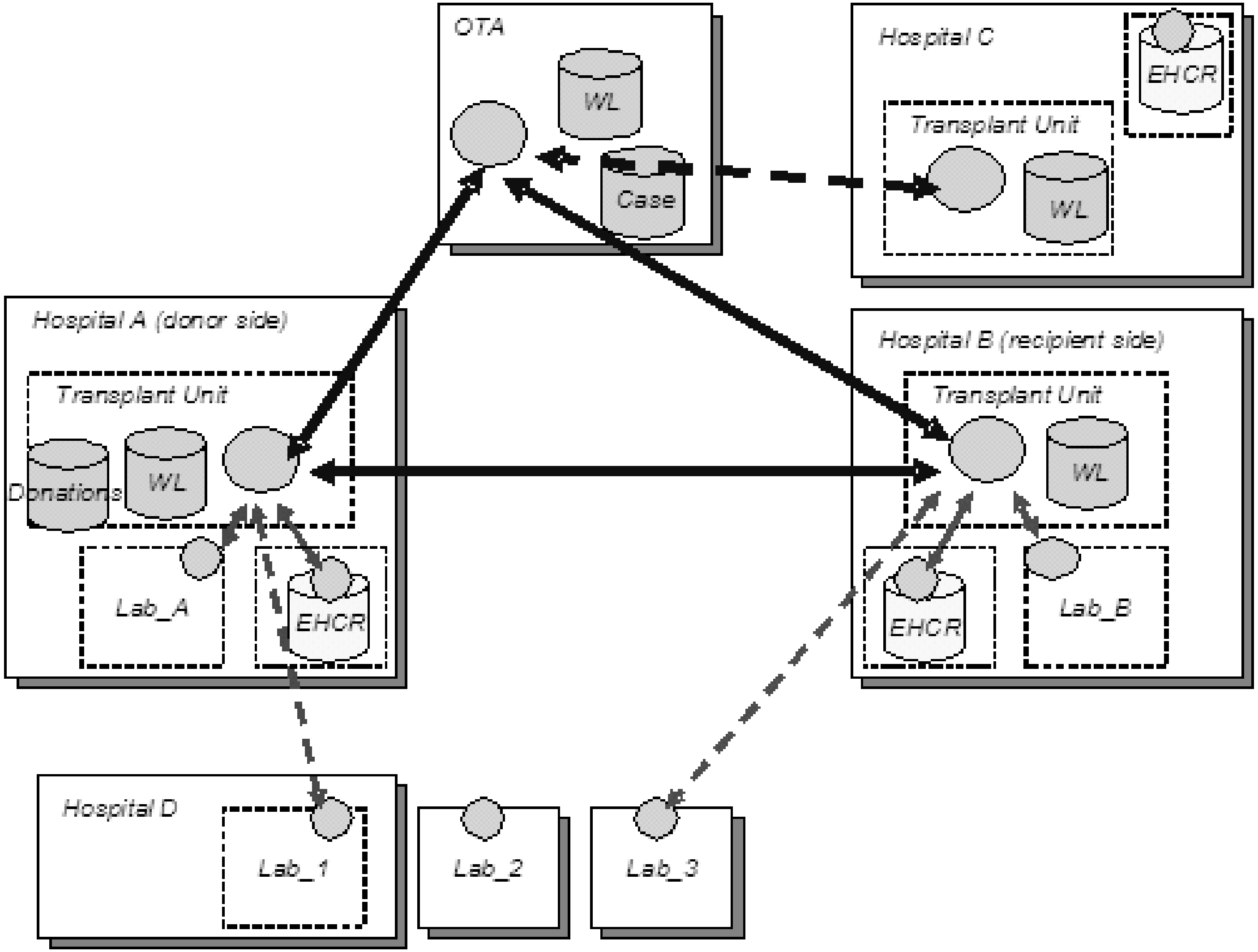
Electronic Healthcare Records and Case Antecedents

- In order to pull together the medical history of a patient we have essentially three options:
  - 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).
  - 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.
  - Build a hybrid system which stores fragments of data with providers but records high level events in a central master record.

Provenance in Service Oriented Architectures

- *provenance*: “the provenance of a piece of data is the process that led to the data”
- provenance of a piece of data will be represented in a computer system by some suitable documentation (a set of *p-assertions*)
- provenance lifecycle:
  - actors create p-assertions
  - p-assertions are stored in a provenance store
  - users or applications can query the provenance store
  - the provenance store and its contents can be managed

Organ Transplant Management Application



Privacy Issues

- disclosures are necessary to treat patients, process claims, measure outcomes, and fight disease
- privacy protection should not be focused on nondisclosure, but on controlled and irreversible disclosure
- which mainly means the protection of the identity of the patient

Privacy in Healthcare Record Management

- technical measures that must be taken
  - separation must ensure that no unauthorized person can connect the identity of the patient with his medical or genetic data
  - data must be protected against any kind of unauthorized processing
  - unauthorized inputs, queries, modifications or deletions of the data while they are stored in the computer memory of the information system, as well as while the data are sent through the network from a computer to another, must be avoided
  - no unauthorized access
  - protect the data against accidental destruction and loss
  - access and data input logging

Privacy and Provenance (1)

- we introduce an additional agent type into the system: the provenance agent
- healthcare agents give up part of the control over the data, the autonomy of the healthcare information is shared with the provenance store
- provenance store is then able to link data and the workflow pieces that generated the data
- the provenance system helps the integration of the information islands which raises additional privacy risks

Privacy and Provenance (2)

- for provenance we need as much information as possible about the whole process 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
- we identified two main risks:
  - *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
  - *event trail risk*: the risk to be able to identify a person by connecting the events and actions related to that person

Protecting Privacy in the OTM Application

- event trail risk: information not available in the healthcare information system has to be matched with the information in the healthcare information system => requires more effort and information to exploit
- cross-link risk: can be exploited using information available only in the healthcare information system => we focus on this
- techniques to reduce the cross-link risk:
  - we do not put medical data in the provenance store
  - we anonymise the patient data

Medical Data and the Provenance Store (1)

- Can we put medical data into p-assertions which will be stored within provenance stores?
- Can we put person identifiers into p-assertions?
- Is it enough if we anonymise patients in p-assertions?
- How can we safely anonymise the patient?
- If we do not store medical information in the provenance store, then how can we retrieve the provenance of medical data?

Medical Data and the Provenance Store (2)

- we do not store sensitive medical data in the provenance store, but only references to such data
- public identifiers of patients are not stored in the provenance store, only anonymised identifiers are used
- 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
- retrieval is done by an EHCR system which is completely under the control of EHCR access rules

Anonym Identity in the Provenance Store (1)

- the fact that the patient was treated or the place of treatment, can be sensitive information => at least the patient identity has to be anonymised
- anonymisation process requirements:
  - 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
  - anonymisation procedure should be irreversible
  - as a consequence, no component in the system should store the real identifier of the patient and its anonymised identifier together

Anonym Identity in the Provenance Store (2)

- in the OTM application the EHCRS system applies anonymous case identifiers
- to connect different cases:
  - we store case identifiers in the p-assertions and define relationships between them (the physician says that the current case is related to a previous one), or
  - we use a global anonymous tracer of the patient and connect each case identifier to that anonymous tracer in p-assertions (helps to connect cases which are not explicitly known)

Conclusion

- provenance may increase the quality of medical services by providing a unified view of the medical history of patients
- but raises new privacy issues which we investigated
- we identified two new privacy risks
- we reasoned that the most critical of these risks is the cross-link risk and proposed methods to eliminate this risk
- we are working on the implementation of the OTM application with provenance extension and we are implementing the outlined privacy protection methods
- privacy issues investigated and methods proposed may be relevant beyond the specific implementation discussed here

Acknowledgements

- IST-2002-511085 Provenance project
  - IBM United Kingdom Limited,
  - University of Southampton,
  - University of Wales, Cardiff,
  - Deutsches Zentrum fur Luft- und Raumfahrt s.V,
  - Universitat Politècnica de Catalunya,
  - MTA SZTAKI
- Javier Vázquez-Salceda’s work has been also partially funded by the “Ramón y Cajal” program of the Spanish Ministry of Education and Science