A Secure and Distributed Infrastructure for Health Record Access

Victoriano Giralt
Central ICT Services
University of Málaga
Málaga, Spain
e-mail: victoriano@uma.es

***Abstract*—The present paper describes the initial ideas for the author PhD thesis dissertation. The main goal of the research is to use Federated Identity and Access Management techniques in widespread use in academic networks, and more every day on the whole Internet, to the controlled, accountable and open access to health information over the Internet as well as controlling and securing the linkage of such data to a given individual. The challenge is to open the data buried in health records for research without giving out information that will allow to identify the individual persons. All of it keeping the real owners of the data, the individuals, in control of the information release. For this, we propose federated identity use to control access to linkage information about medical acts made publicly available. Using this technique, it would be even possible to provide totally anonymous informed health care.**

Keywords-health record; security; accountability; Federated Identity and Access Management.

I. INTRODUCTION

Health related data has the highest level of privacy protection in most countries data protection laws, but, at the same time, it is in the best interest of the whole medical science and the individuals themselves, that health data can be readily available.

The emergency room scenario has been used many times as an use case for expedited access to the whole heath record of an individual, where consent cannot be requested in the most life threatening situations. [1]

On the other hand, free access to high volumes of anonymous, but traceable (not to a real person only to an anonymous single individual), patient data, could be an invaluable resource for clinical research.

Access to health data should, in most cases, be granted by the individual to whom such data pertains, and should be accountable to those who see those data.

The present paper will propose a system than can be built using already available, and in use, protocols and tools that can both allow free access to anonymous health data and provide controlled and accountable means for de-anonymising the health records and tracing them back to the original person to whom they are related. [4][5][11][10] [12]

The proposed work builds upon the author experience in dealing with personal data in diverse scenarios, with some award winning results. [9] The driving force in the past eight years have been to put persons in the centre of their on-line lives and in control of personal data about them. [14][15] In this case, we propose a change of the status quo. At present, health records are owned by the institutions or practitioners that produce them, instead of the persons that are the subjects of those records. The main reason for our work is to put these persons (all of us) at centre stage and give them control over their own information, regardless of who has produced or created it. The present paper has resulted both from experience and the impression that the time is right for connecting two fields, health record management and electronic identity management, that are experiencing rapid development at this point in time. [13][11][1]

By publishing this work in progress paper, the author tries to gather as much hindsight as possible from others that might be working on ideas that could cross-pollinate and contribute to the final proposed landscape.

We will present the different scenarios of access and creation of health records by means of user stories:

- Individual enrolment
- Creation of health record in clinical practice
- Access to health records in clinical practice
- Access to health records from the emergency room
- Access to health information for research purposes
- Access to personal identity information

Finally, we will describe the technologies that will be used to create a demonstrator.

II. TECHNICAL TERMINOLOGY

The proposed work involves several domains with specialised terminologies that are not commonly understood. The author main field of work, despite his academic background, is electronic identity and privacy and access control, thus making this the main domain for the work.

A. Electronic identity terms

- Identifiable individual: A single physical person than can be identified by a set of personal data that constitutes their identity record.
- Attribute: A property of an identity record consisting of one or more values. All the values of an identity attribute are related by a common purpose or meaning. For example, the collection of telephone numbers belonging to

 ISBN: 978-1-61208-176-2

a person might form an identity attribute on the identity record that represents that individual.
- Principal: a person for whom another entity acts as an agent or representative.
- Pseudonym: an identifier that can single out an individual without revealing the real identity.
- Biometric information: personal information attributes derived from physical or biological characteristics of an individual.

III. General information processing and storage terms

- Hash: the result of using a hash function on an element of a data set. This functions transform larger data sets into smaller ones and produce the same result given the same input.
- Universally Unique Identifier (UUID): a 16 byte (128 bits) string that is guaranteed to be different from all other UUIDs generated before 3603 A.D., if the recommended algorithms are used [2].
- Resolver: an entity that can link pseudonymous identifiers like UUIDs to information about principals with or without identifying them.

A. Federated Identity and Access Management terms

- Identity Provider (IdP): An entity able to identify individuals and provide attributes pertaining to their identity.
- Relying Party (RP): An entity that trusts the federation and accepts identities asserted by IdPs.
- Federation: Infrastructure supporting the trust links between IdPs and RPs.
- Authorisation Server (AS): it is a trusted entity that takes access decisions based on attributes of the principals involved in a transaction in support of an RP.
- Attribute Authority (AA): is a trusted entity that asserts attributes about principals with or without revealing their identities to other principals involved in a transaction.
- Level of Assurance (LoA): the level of confidence with which the identity of an individual has been vetted in order to be linked to an electronic identity record.

B. Medical terms

- Health Level Seven (HL7): an international standards organisation that works for the interoperability of health clinical and administrative data. And, it is also used to refer to the standards defined by said organisation. [3]
- Act: one of the three main classes defined in the HL7 reference information model (RIM) [8] that represent actions that are executed and must be documented as various parties provide health care. [3]
- Role: second of the main classes defined in the HL7 RIM that establishes the function played by entities as they participate in health care acts. [3]
- Entity: third of the classes that represents the physical things and beings that are of interest to, and take part in, the health care. [3]
- Act Relationship: represents the binding of one act to another. [3]
- Participation: expresses an act's context, such as who performed it, for whom and where. [3]
- Role Link: represents relationships between individual roles. [3]
- Health Record (HR): a collection of health information related to an act or to the general health state of an individual. [3]

IV. Actors

A. Patient

We will use the term patient to refer to a person that is the subject of a medical act, although both in classical and modern medicine, keeping persons in a healthy condition is the main target of medical practice.

B. Practitioner

Practitioner will refer to any health care professional of any kind that interacts with patients in medical acts.

C. Emergency Room Practitioner

We have singled emergency room (ER) practitioners as they will receive special treatment in the system regarding the way they can access health records.

D. Staff member

This term refers to non medical professionals that have a role in medical acts like clinic receptionists or hospital administrative staff that require access to partial content of the HRs or to personal data of the patients.

E. Relative

A person with a family or other kind of social relationship to a patient that might play a role in authorising access to HR or provide personal information about the patient.

F. Researcher

A person that requires anonymous, or, at most, pseudonymous access to HRs for scientific research work.

V. General System Description

The proposed system aims to provide both freely available anonymous HRs published as HL7 [7] XML [16] documents on common web servers and a privacy controlled way of liking such records to the patients that participated in the corresponding medical acts.

There exist both commercial and non-profit repositories for personal health records, but they are centralised and, in many cases, under tight control of entities like insurance companies. We propose a totally open and distributed system based on trust models proven in higher education, research, government and vertical industries. The level of trust can be as high as to use one of such federations for controlling fusion nuclear reactors remotely or submitting experiments to synchrotron facilities.

 ISBN: 978-1-61208-176-2

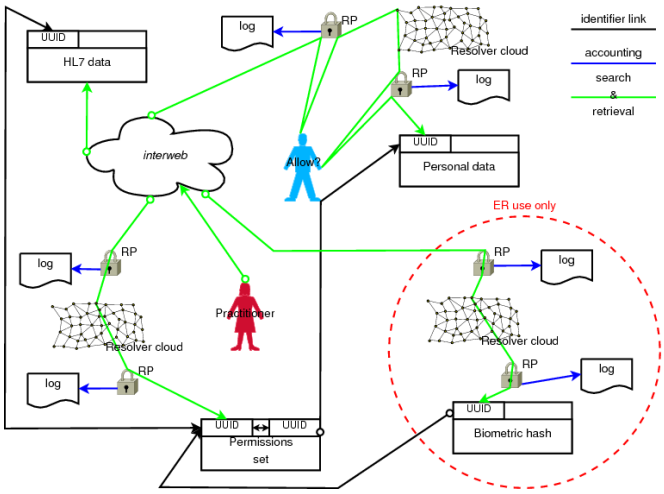

Figure 1. General system description

Identity federations are trust circles that have clearly defined rules for participation that, at the same time, act as codes of conduct for those entities that decide to participate in a certain federation. Even, in this moment in time, federations of federations, also known as inter-federations, are successfully being formed around the globe in the higher education and research sector. [11][10]

A. Pseudonymous identifiers

Any information part in the system, be it a patient personal information, an HR resulting from an act, or a biometric characteristic, will receive an UUID [2] as identification. These identifiers will be used as keys to find the resolvers that will link patients and HR using the UUIDs themselves as double indirection pointers. There will be a resolver finder cloud to locate the proper pointer to resolve a given UUID into the data that it represents.

B. HR publication

Web technologies facilitate the publication of huge amounts of data and also allow for easily indexing, locating and presenting such data. HL7 [7] XML [16] is a text format that provides all the required characteristics for easy web publication of HR, and, at the same time is an accepted interoperable format.

So, the information resulting from a medical act will be published as an HL7 XML document associated to an UUID that the practitioner or the relevant staff can provide to the patient in an electronic format. This UUID will be linked to the patient's UUID in a resolver the patient decides. This HR UUID to resolver relationship is also published through the resolver finder cloud.

The HL7 [7] document should not contain any personal information about the patient and the minimal possible amount of data.

C. Patient identification

The patients themselves will register to an IdP recognised in the global health care federation, using a method that provides an acceptable LoA. The personal data record will receive an UUID that can be published into the resolver cloud.

Patients should also get a hash out of some standardised biometric information. Ideally, this information should be genetic as it is the only type of biometric data that any body part carries. The state of the art does not yet allow for a full genomic characterisation of an individual in a reasonable time and for a reasonable cost, but there is fast progress in that area. Any other biometric information can be standardised, and, for the purpose of the demonstrator, we propose the use of digitised fingerprints, that will be hashed using Automated Fingerprint Identification Systems (AFIS) [17] algorithms.

Using fingerprints could be a handicap for the ER use case that we will present, in case the patient has lost the hands, but the prevalence of such situations is not high enough to render the system useless.

Once the patient has a biometric hash, it is associated to an UUID that will be published in a special resolver finder cloud, that allows for, so to say, backwards searches. This is required mainly for the ER use case.

The patient personal data will also include any relevant information needed for authorisation related contacts, either direct or through a relative.

D. Practitioner, staff and researcher identification

All other principals that participate in health care acts will register to pertinent IdPs in the federation, that could be run by hospitals, physicians or nurses colleges, insurance companies, etc. These IdPs will assert attributes that allow the AS, that control access to the RPs in the resolvers, to take appropriate decisions for granting access to the requested information. Thus, no one will get more information than that required to participate in a given act.

VI. User Stories

For the sake of brevity, we will do a shallow description of the user stories proposed in the introduction.

A. Individual enrolment

I'm a patient and want to publish my HR.

1) I select an IdP or the national health system provides me one.
2) I identify to the IdP using documents to achieve the required LoA and provide contact information for me and my closest relative.
3) I get the UUID that identifies my personal data.
4) My UUID is published by the IdP resolver.
5) My biometric hash is published in the resolver cloud.
6) I get my biometric hash UUID and link it to my UUID.

 ISBN: 978-1-61208-176-2

B. Creation of health record in clinical practice

1) I as a patient go visit a practitioner.
2) All acts are compiled into HR documents.
3) The HR are dated and get UUIDs.
4) The HR UUIDs and my UUID are inserted in my IdP resolver.
5) The HR UUIDs are sent to the resolver finder cloud from the resolver together with the pertinent pointer.

C. Access to health records in clinical practice

1) I as a patient go visit a practitioner.
2) The practitioner requests historic HR information.
3) I provide the practitioner with my UUID.
4) The practitioner identifies to the pertinent IdP and queries the resolver finder cloud and then, the appropriate resolver.
5) The resolver AS sends me a message indicating the practitioner identity, information about the requested data and a request for granting authorisation.
6) I grant the access and set a time limit.
7) The practitioner can access the data.

D. Access to health records from the emergency room

1) An unconscious and unidentified patient arrives in a life threatening condition.
2) The standard biometric parameters are determined and hashed appropriately.
3) A practitioner in the ER identifies to an IdP connected to an AA that asserts the attributes that verify the ER job.
4) The asserted attributes allow access to the special resolvers for biometric hashes, and to the UUID resolvers without requesting authorisation from the patient or relatives.
5) The resolvers return all HR UUIDs related to the UUID associated to the biometric hash.
6) The ER practitioner can retrieve the whole history of HRs related to the patient, without knowing the identity of the individual.

E. Access to health information for research purposes

1) I am a researcher working on a certain disease.
2) I search the web an collect all pertinent HRs.
3) I need to know about historic HR data about the same individuals that form the population under study.
4) I identify to my IdP that has an AA that asserts attributes to prove my researcher condition.
5) I query the resolvers for other HR UUIDs that belong to the same individuals as the HR UUIDs in the collection under study.
6) Depending on user preferences, data sensitivity or other parameters, patients get a request for granting access to the HR.

F. Access to personal identity information

1) I am a hospital staff member.
2) I need to know a patient identity for billing purposes.
3) I identify to the hospital IdP and the hospital AA asserts attributes to prove my administration staff status.
4) I query the resolver finder cloud to find the resolver for the patient UUID.
5) I query the patient resolver.
6) I get back the data needed to bill the patient.
7) The patient is notified of the personal data request.

VII. Technologies for Implementing the System

There are several options for some of the technologies needed to implement the proposed system. Producing a demonstrator is one of the main aims of the work described in the present paper, so it has been necessary to select a given technology for the different parts of the system. The selection has been mostly based on the author's experience or common practice in the fields in which he is working.

A. Security Assertion Markup Language (SAML)

SAML version 2 [4] is a proven method for expressing trust via electronic means and asserting information about principals that is in widespread use in present identity federations. It allow for inter-domain authentication, authorisation and accounting of access to resources. Such information is carried using XML [16] documents.

SAML2 will be used for the IdPs, AA and some RPs in the system.

B. Open Authorisation (OAuth)

Also in version 2, OAuth is a protocol that allows third party access to data with express authorisation of the owner of that data. [5]

OAuth will be used for the AS and some RPs in the system.

C. Distributed Hash Tables (DHT)

A distributed hash table (DHT) is a class of a decentralised distributed system that provides a look-up service similar to a hash table; (key, value) pairs are stored in a DHT, and any participating node can efficiently retrieve the value associated with a given key. Responsibility for maintaining the mapping from keys to values is distributed among the nodes, in such a way that a change in the set of participants causes a minimal amount of disruption. This allows a DHT to scale to extremely large numbers of nodes and to handle continual node arrivals, departures, and failures [6].

Due to the distributed, decentralised and resilient nature of DHTs, the system will use this technology to implement the resolver finder cloud. The keys will be UUIDs and the values will be URLs pointing to the resolver that can resolve a given UUID. In the special case of biometric hashes, the keys will be the later and the values will be UUIDs to feed into the normal resolver finder cloud.

 ISBN: 978-1-61208-176-2

VIII. A DEEPER VIEW OF USER STORY C

We will do a more detailed description of user story C, *Access to health records in clinical practice*, once we know the ttechnologieswe will be using for the ddemonstratorimplementation.

The actors and elements involved are:

- Patient: The subject in the clinical act.
- Practitioner: The health care professional performing the clinical act.
- IdP: The Identity Provider where the Practitioner authenticates.
- Resolver: The element that resolves the Patient identifier and locates pointers to HR.
- RP: The element that grants access to the Resolver.
- AS: The element inside the RP that permits the retrieval of pointers.
- HR: Relevant health information about the Patient.

Patient and Practitioner are both physical persons and their electronic representations, and computer applications acting in their name as proxies. The elements are computer applications and electronic representations of information.

Patient goes visit PPractitionerfor some clinical Act. Let's assume that is related to ccholesterolblood levels. It is the first time Patient and Practitioner meet. So, Practitioner needs some historic data about blood samples, mostly cholesterol levels and some related values. Thus, Patient provides Practitioner with a UUID that can be linked to published HRs, through the use of resolvers. The process flow proceeds as depicted in figure 2, with the following steps indicated as circled numbers:

1) Patient provides Practitioner with UUID
2) Practitioner goes to the resolver cloud
3) RP on resolver cloud requests Practitioner identity
4) Practitioner identifies to the pertinent IdP and returns to RP
5) AS in resolver cloud RP finds Patient authorisation method and requests access permissions for Practitioner.
6) The resolver AS sends Patient a message indicating Practitioner identity, information about the requested data and a request for granting authorisation.
7) Patient grants access and sets a time limit.
8) Practitioner retrieves a set of UUIDs from the resolvers that belong to previous Patient HRs with relevant information.
9) Practitioner retrieves the needed HRs.

RPs log all resolution requests and authorisation responses with pertinent identity information about the requestor and granter, in order to create audit trails.

In our ddemonstrator SAML2 [4] protocol will be used to carry identity and authentication information, while OAuth2 [5] will carry the authorisation requests and responses.

It is ppossibleto increase the security of the previous flow rrequiringPatient to also authenticate against an IdP for replying to the authorisation request.

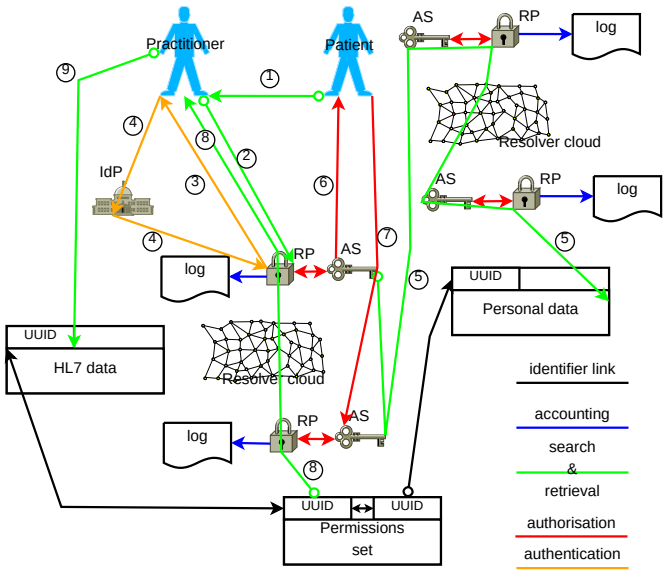

Figure 2. Access to health records in clinical practice

IX. CONCLUSIONS

In case the ideas presented in this paper are deemed worth the effort and such effort produces the expected results, the system will have two main advantages:

- a paradigm shift moving ownership of the data from the hands of those that produce such data into the hands of those to whom the data belongs to,
- and open data availability for many purposes.

ACKNOWLEDGEMENTS

The author wishes to thank all the fruitful conversations he has had with wise people in the Identity Federation space, with special mention of some ones that have seeded the ideas resulting in the work described in the present paper, including, but not restricted to, and in no particular order, Roland Hedberg, Ken Klingenstein, Andrew Cormack, J.A. Accino, Licia Florio, Klaas Wierenga, Milan Sova, RL "Bob" Morgan, Lorenzo Gil, Matthew Gardiner, Dave Birch, David Chadwick, and, last, but not least, for his very special support, Diego Lopez.

REFERENCES

[1] P. Groen, P. Mahootian and D. Goldstein, *Medical informatics: emerging technologies and 'open' health IT solutions for the 21st century*, January 2011, in press.

[2] ITU, *Universally unique identifiers*, URL: http://www.itu.int/ITU-T/asn1/uuid.html retrieved: November 21st, 2011.

[3] R. Gajanayake, R. Iannella and T. Sahama, *Sharing with care: An information accountability perspective*, Internet Computing, IEEE , vol.15, no.4, pp.31-38, July-Aug. 2011 doi: 10.1109/MIC.2011.51 URL: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5749997&isnumber=5934 844 retrieved: November 21st, 2011.

[4] P. Madsen et al., *SAML V2.0 Executive Overview. OASIS Committee Draft*, April 2005. Document ID sstc-saml-tech-overview-2.0-cd-01-2col URL: http://www.oasis-open.org/committees/download.php/13525/sstc-saml-exec-overview- 2.0-cd-01-2col.pdf retrieved: November 21st, 2011.

 ISBN: 978-1-61208-176-2

[5] E. Hammer-Lahav, D. Recordon and D. Hardt, *The OAuth 2.0 Authorization Protocol*, http://tools.ietf.org/html/draft-ietf-oauth-v2-21 retrieved: November 21st, 2011.

[6] Wikipedia, *Distributed hash table*, http://en.wikipedia.org/wiki/Distributed_hash_table retrieved: November 21st, 2011.

[7] *Health Level Seven Standard Version 2.7 - An Application Protocol for Electronic Data Exchange in Healthcare Environments*, ANSI/HL7 V2.7-2011, National Institute of Standards and Technology, U.S. Government Printing Office: Washington, DC, 2011.

[8] *HL7 Version 3 Standard: Reference Information Model, Release 2*, ANSI/HL7 V3 RIM, R2-2010, National Institute of Standards and Technology, U.S. Government Printing Office: Washington, DC, 2010.

[9] V. Giralt, et al., *Example of Privacy Management in a Public Sector Organizational Electronic Directory* in Cunningham P., Cunningham M. (Eds.) Expanding the Knowledge Economy: Issues, Applications, Case Studies, IOS Press, Amsterdam, pp. 1386-1393, 2007

[10] V. Giralt, et al., *Federated Identity Infrastructure for the Andalusian Universities. Deployment of a Multi-technology Federation* in Cunningham P., Cunningham M. (Eds.) Collaboration and the Knowledge Economy: Issues, Applications, Case Studies, IOS Press, Amsterdam, pp. 1139-1144, 2008

[11] D. Simonsen, *Revealing the Identity of Federations*, the 16th European University Information Systems Organisation (EUNIS) congress, EUNIS 2010, University Information Systems: Selected Problems, University of Warsaw, Poland, pp. 11-20.

[12] M. Ramos, et al., *Design and Implementation Details of the Public Andalusian Universities Identity Federation CONFIA*, the 16th European University Information Systems Organisation (EUNIS) congress, EUNIS 2010, University Information Systems: Selected Problems, University of Warsaw, Poland, pp. 217-224

[13] *Microsoft® "Geneva" Server and Sun OpenSSO: Enabling Unprecedented Collaboration Across Heterogeneous IT Environments*, Microsoft® and Sun Microsystems White Paper, 2009, URL: http://download.microsoft.com/download/C/F/D/CFD1D9C8-EBA4-4780-B34B-DBEB5A4792B F/Geneva%20and%20Sun%20OpenSSO.pdf retrieved: November 21st, 2011.

[14] J.A. Accino, et al., *dUMA: comprehensive personal information management*, the 17th European University Information Systems Organisation (EUNIS) congress, EUNIS 2011, Maintaining a Sustainable Future for IT in Higher Education, Trinity College Dublin, Ireland

[15] J.A. Accino, M. Cebrian, and V. Giralt, *Identity Based Clusters of Applications for Collaboration and eLearning*, the 15th European University Information Systems Organisation (EUNIS) congress, EUNIS 2009, IT: Key of the European Space for Knowledge, University of Santiago de Compostela, Spain

[16] T. Bray, J. Paoli, C.M. Sperberg-McQueen, E. Maler and F. Yergeau eds., *Extensible Markup Language (XML) 1.0 (Fifth Edition)*. W3C Recommendation 26 November 2008, URL: http://www.w3.org/TR/2008/REC-xml-20081126/ retrieved: November 21st, 2011.

[17] K.R. Moses, P Higgins, M. McCabe, S. Probhakar, S. Swann, *Fingerprint Sourcebook-Chapter 6: Automated Fingerprint Identification System (AFIS)*, National Institute of Justice/NCJRS 225326, 2010, URL: http://www.ncjrs.gov/pdffiles1/nij/225326.pdf retrieved: November 21st, 2011.

 ISBN: 978-1-61208-176-2