Digital Health and Informatics Innovations for Sustainable Health Care Systems
J. Mantas et al. (Eds.)

doi:10.3233/SHTI240627

Providing Provenance in eHealth

Silvia LLORENTE, Jaime DELGADO[1], Daniel NARO and Nikolaos FOTOS
Departament d'Arquitectura de Computadors (DAC)
Universitat Politècnica de Catalunya · BarcelonaTECH (UPC), Spain
ORCiD ID: Silvia LLORENTE https://orcid.org/0000-0003-2000-6912
Jaime DELGADO https://orcid.org/0000-0003-1366-663X
Daniel NARO https://orcid.org/0000-0003-3114-1298
Nikolaos FOTOS https://orcid.org/0009-0009-8173-9555

Abstract. eHealth data can be generated by different health professionals. Documenting and storing the provenance of medical data will help in providing patients and medical institutions with trustable and traceable information regarding medical procedures and their results. The use of a modular architecture like HIPAMS allows covering different eHealth information representation formats as well as providing protection mechanisms and access control.

Keywords. Provenance, Metadata, eHealth, Privacy, Security

1. Introduction

Adding provenance information to eHealth data is a way to guarantee that patients' medical data history is maintained and secured. Therefore, there are some issues to be considered, like the provenance information representation format, the type of medical data whose provenance is being described or how to link provenance information with medical data. Consider the case, for example, of a report over an elbow radiography. We need to make sure that this report applies to a specific radiography (from the image generation point of view) and that it corresponds to a specific patient at a given moment (corresponding to a specific health condition). If any of these relationships fails, everything fails, even leading to a wrong physical action over the patient.

In this paper we describe how provenance information can be applied to eHealth data, based on existing standardization initiatives. Then, we show how to protect and manage provenance information using the Health Information And Management System (HIPAMS) [1], a modular architecture which provides several functionalities for health content creation and management. We also present how these modules interact in a specific use case, which is the creation and protection of provenance information. Finally, we present some conclusions and future work.

1.1. Background

From now on, we will use the term provenance, to simplify, when we mean the information needed to describe data provenance. On this regard, there are several

[1] Corresponding Author: Jaime Delgado; E-mail: jaime.delgado@upc.edu.

initiatives to describe provenance information, some of them related to eHealth, but also in other contexts like, to mention a few, the ones coming from different ISO standardization committees. Therefore, we can find, for example, provenance being defined for preservation purposes in Self-contained Information Retention Format (SIRF) Specification (ISO/IEC 23681) [2] or Metamodel for data set registration (ISO/IEC 11179-33) [3]. Moreover, there is work in progress in ISO/IEC JTC 1/SC 27/WG 4 (Security controls and services) on ISO/IEC AWI 5181, Data provenance [4]. This standard provides guidelines, methodology and techniques for securely managing provenance metadata during the whole lifecycle of data use and data manipulation.

Also inside ISO committees, but related to biological samples and genomics, we can find provenance work developed in ISO/TC 276 (Biotechnology) for biological samples [5] which is currently being combined with ISO/IEC 23092 (Genomic Information Representation, MPEG-G) [6] thanks to the joint work done between these two standardization initiatives [7].

It is worth noting that most of the identified initiatives use as a basis the provenance model defined by W3C, W3C-PROV [8], adding specific information for their respective research fields.

Moreover, survey [9] details how provenance is considered in the area of security and privacy, including techniques used to provide and protect provenance.

Finally, data provenance in healthcare is surveyed in papers [10]-[12], where different aspects related to data provenance are considered. Their conclusions show that some common provenance formats, like W3C-PROV, are mainly used. Furthermore, these papers identify which are some of the current trends in provenance protection.

2. Methods

To implement provenance, we need to consider first the generation of the provenance information itself. Second, the association of provenance to the eHealth data it applies to. Therefore, depending on the format of the eHealth data, different strategies apply.

Our approach for providing and managing provenance is to use HIPAMS. As described in [1], it is an architecture that defines different modules to deal with both medical information and its associated metadata, also considering access control and protection aspects. The Provenance Service module is devoted to handle provenance information. HIPAMS modules implement operations to address the information they manage. In the specific case of provenance, the following operations are provided:

- Provenance information creation. This operation creates the provenance information associated to some eHealth data. This metadata may be stored inside the medical information, if allowed by its format, and/or stored in an external data structure, which links to the data to which it refers to. Metadata could even be just a URL pointing to a service, which could be external to HIPAMS, managing provenance. Finally, provenance information has to keep its integrity (digital signatures and/or encryption, among other protection techniques, may apply), in order to guarantee that it is not tampered with.

- Provenance information update. This operation modifies existing provenance information, adding or removing data. Then, the updated information has to be kept connected to the eHealth data, as well as protected and stored in HIPAMS. These updates lead to provenance information chains.

- Provenance information verification. This operation checks that some provenance information linked to some eHealth data is correct and has not been unauthorizedly modified. In case of provenance information chains, they are also verified.

- Provenance information deletion. This operation deletes some provenance information from HIPAMS. It should be only done when the medical data is deleted as well. In any other case, for example, when an error in provenance information is detected, provenance should be updated, not deleted, in order to follow the provenance information chain.

These operations provide the CRUD (Create, Read, Update, Delete) operations over provenance information. Nevertheless, especially on the R case, some specific search operations could be added, to find provenance information applied to some medical data.

Finally, there is another important aspect to consider that is the provenance information representation itself. XML or JSON could be used for formalizing provenance metadata, following the format described by Fast Healthcare Interoperability Resources (FHIR) [13].

3. Results

Now that we have the operations, we can we associate provenance information to medical data. As we already have experience in the management of other kinds of digital information, by using modular architectures like GIPAMS [14] or MIPAMS [15], we now apply a similar approach to eHealth. On the other hand, we also know how to apply provenance information to images, as explained in [16].

The main idea is that, thanks to the different modules being part of HIPAMS, we can define provenance information, protect it and connect it to the medical information. The access to this information can be also controlled by means of policies and rules, even associated to the user or the role, thanks to user authentication. Moreover, using modular architectures improves scalability, as several module instances could be started upon request. It also allows interoperability as one can implement different medical formats by means of new modules as they appear.

To validate our approach, we have developed several use scenarios. For example, Figure 1 shows the workflow diagram of medical content creation including provenance information.

In that workflow, to start using the different services, the user needs to be authenticated in front of the Authentication Service to receive a valid token. Later on, this token will be sent when invoking other services through the user application. This use case involves three services, namely, Provenance Service (PrS), Protection Service (PtS) and Policy Service (PS). Content creation and protection steps are omitted for clarity. After that, provenance information is registered in PrS. In this example, provenance is also protected, so the PtS is also called to protect provenance information. Afterwards, two privacy policies are created by the PS, to control subsequent access to protected medical content and provenance information. The step including provenance information into medical information, if required, is not represented in this diagram, but it should be done after provenance information creation and protection.

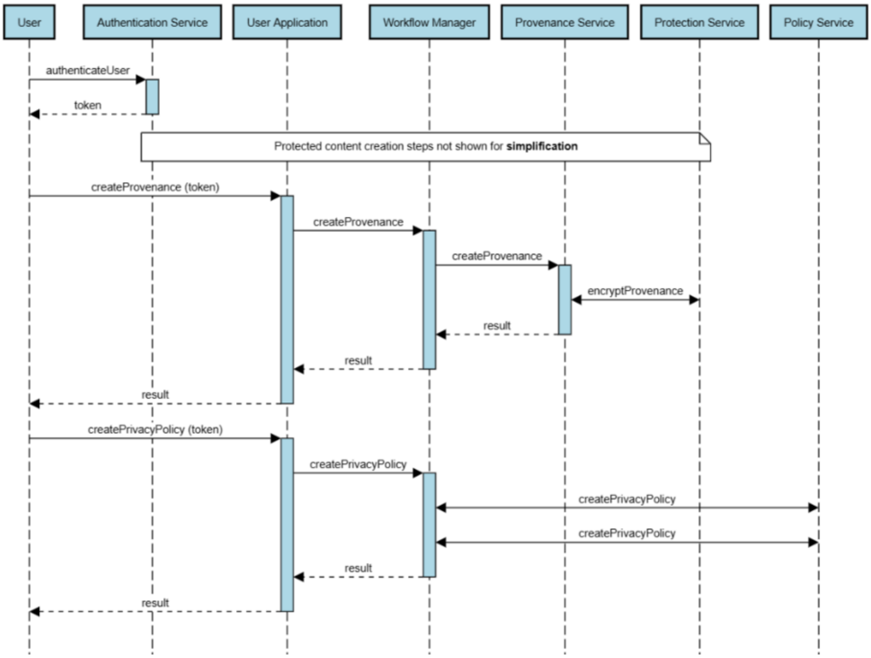

Figure 1. Provenance information creation and protection.

4. Discussion

The use of a modular architecture, like HIPAMS, facilitates adoption of provenance creation and management and its application into different kinds of medical data. Moreover, it eases the support of new medical data representation formats, by implementing new Health Content Services dealing with them.

As a result of the use of HIPAMS, the implementation of common security and protection techniques, like digital signatures, encryption or privacy policies for access control, allows their application into many different medical data representation formats.

Furthermore, protection techniques can be also applied to provenance information or metadata associated to medical data, obtaining a high degree of protection (theoretically, it could be the same as the medical information itself), which is a valuable asset when considering patients' privacy. In the end, provenance information is itself data and thus can be protected by the same mechanisms as medical data.

5. Conclusions and Future Work

There are several initiatives for providing provenance of digital information, considering different ways of representing provenance information. In this paper, we focus on the provision of provenance information for medical data. To do so, we propose the use of the HIPAMS platform, an extensible modular platform which provides security, authentication and access control functionalities as well as medical content creation and management.

In this way, the implementation of a Provenance Service opens the door for further application of provenance to different kinds of digital information, not only medical. We are also considering this for images, as described in [16], and to genomic information, as presented in [1]. Nevertheless, we will continue exploring other provenance mechanisms, based on security and Artificial Intelligence approaches.

Acknowledgements

The work presented in this paper has been partially supported by the Spanish Government under the project GenClinLab-Sec (PID2020-114394RB-C31) funded by MCIN/AEI/10.13039/501100011033, by the Generalitat de Catalunya (2021 SGR 01252) and by the HORIZON.2.1 - Health Programme of the European Commission, Grant Agreement number: 101095448 - MEDSECURANCE.

References

[1] Delgado J, Llorente S. Provenance and dynamic consents for the management of medical data. Stud Health Technol Inform. 2022 Nov; 299:171-176. Available from: https://ebooks.iospress.nl/doi/10.3233/SHTI220978

[2] ISO/IEC 23681:2019. Information technology - Self-contained Information Retention Format (SIRF) Specification. 2019. Available from: https://www.iso.org/standard/76648.html

[3] ISO/IEC 11179-33:2023. Information technology - Metadata registries (MDR) - Part 33: Metamodel for data set registration. 2023. Available from: https://www.iso.org/standard/81725.html

[4] ISO/IEC AWI 5181. Information technology - Security and privacy - Data provenance. Available from: https://www.iso.org/standard/80971.html

[5] ISO/TC 276 (Biotechnology). ISO/DTS 23494 - Provenance information model for biological material and data. 2022. Available from: https://www.iso.org/standard/80715.html

[6] ISO/IEC. ISO/IEC 23092 - Genomic Information Representation. 2022. Available from: https://www.mpeg.org/standards/MPEG-G/

[7] Wittner R, Gallo M, Frexia F, Leo S, Pireddu L, Mascia C, et al. Linking provenance and its metadata in multi-organizational environments. The University of Manchester. 2024. Available from: https://research.manchester.ac.uk/en/publications/linking-provenance-and-its-metadata-in-multi-organizational-envir

[8] W3C. Provenance Overview. 2013. Available from: https://www.w3.org/TR/prov-overview/

[9] Pan B, Stakhanova N, Ray, S. Data Provenance in Security and Privacy. ACM Comput Surv. 2023 Jul; 55(14-323):1-35. Available from: https://doi.org/10.1145/3593294

[10] Ahmed M, Dar AR, Helfert M, Khan A, Kim J. Data Provenance in Healthcare: Approaches, Challenges, and Future Directions. Sensors. 2023 Jul; 23(14):6495. Available from: https://doi.org/10.3390/s23146495

[11] Sembay MJ, de Macedo DDJ, Pioli Jr L, Braga RMM, Sarasa-Cabezuelo A. Provenance Data Management in Health Information Systems: A Systematic Literature Review. J Pers Med. 2023; 13(6):991. Available from: https://doi.org/10.3390/jpm13060991

[12] Johns M, Meurers T, Wirth FN, Haber AC, Müller A, Halilovic M, et al. Data Provenance in Biomedical Research: Scoping Review. J Med Internet Res. 2023; 25:e42289. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10132013/

[13] FHIR. Provenance Resource. Available from: https://www.hl7.org/fhir/provenance.html

[14] Llorente S, Delgado J. Implementation of Privacy and Security for a Genomic Information System Based on Standards. J Pers Med. 2022; 12(6):915. Available from: https://www.mdpi.com/2075-4426/12/6/915

[15] Llorente S, Rodriguez E, Delgado J, Torres-Padrosa V. Standards-based architectures for content management. IEEE Multimedia. 2013 Oct-Dec; 20(4):62-72. Available from: https://doi.org/10.1109/MMUL.2012.58

[16] Fotos N, Delgado J. Ensuring privacy in provenance information for images. 24th International Conference on Digital Signal Processing (DSP). 2023 Jun. Available from: https://doi.org/10.1109/DSP58604.2023.10167902