Large Scale Projects in eHealth
B. Blobel et al. (Eds.)
IOS Press, 2012

doi:10.3233/978-1-61499-052-9-23

Nation-Wide Primary Healthcare Research Network: a Privacy Protection Assessment

Etienne DE CLERCQ[a,1], Viviane VAN CASTEREN[b], Nathalie BOSSUYT[b], Sarah MOREELS[b], Geert GODERIS[c], Stefaan BARTHOLOMEEUSEN[c], Pierre BONTE[d], Marc BANGELS[d]

[a] *Research Institute for Health and Society (IRSS), UCL, Belgium*
[b] *Scientific Institute of Public Health (IPH), Unit of Epidemiology, Brussels, Belgium*
[c] *Academisch Centrum voor Huisartsgeneeskunde, KU Leuven, Belgium*
[d] *National Institute for Health and Disability Insurance, Belgium*

Abstract. Efficiency and privacy protection are essential when setting up nation-wide research networks. This paper investigates the extent to which basic services developed to support the provision of care can be re-used, whilst preserving an acceptable privacy protection level, within a large Belgian primary care research network. The generic sustainable confidentiality management model used to assess the privacy protection level of the selected network architecture is described. A short analysis of the current architecture is provided. Our generic model could also be used in other countries.

Keywords: Computerized patient record, primary healthcare, health networks, confidentiality

Introduction

For many years, the secondary use of data from GPs' electronic patient records has been thought to provide many benefits, such as research, education, assessing quality of care, health policy, etc. [1]. However, these activities raise some specific data protection issues, requiring some privacy-enhancing techniques (e.g. anonymization or pseudonymization of the data) [2, 3].

In Belgium during the last decade many permanent or temporary restricted research networks (involving a limited number of GPs) have been set up [4-6]. Now, the ACHIL project (Ambulatory Care Health Information Laboratory), which is funded by the National Institute for Health and Disability Insurance (NIHDI), is setting up a large national Health Research Network that will involve nearly all 10,000 acting Belgian GPs and the 17 different software systems they use [5].

In its first phase (2009 – 2013), this network will focus on assessing the quality of care for approximately 25,000 patients who are currently enrolled in the newly developed ambulatory care pathways (known as 'care trajectories' in Belgium) funded by the NIHDI (cf. www.trajetdesoins.be or www.zorgtrajecten.be). In 2012, only a small amount of structured data will be collected (no free text). Cross-sectional studies

[1] Corresponding Author. Etienne De Clercq, MD, PhD, Université Catholique de Louvain - IRSS, Clos Chapelle-aux-champs 30, bte B1.30.13, 1200 Brussels, Belgium; Email: etienne.declercq@uclouvain.be; Phone: +32-2-764 3262; Fax: +32-2-764 3187; URL: www.uclouvain.be

will be initiated, followed in the near future by longitudinal studies. The research database will not be accessible by the GPs. Benchmarking will be provided for each GP. Such a research network is an important element of the chronic care model that supports these care trajectories [8].

In Belgium, as abroad, semi-anonymous architectures have been developed to meet the specific privacy requirements of this type of research network: researchers only have access to coded data, but re-identification of the patient or of the GP remains possible through a trusted third party (TTP) when appropriate and authorised, for example to send the necessary feedback to the right GP [5, 9, 10].

When setting up a large-scale primary care research network, the efficiency of the privacy protection techniques being used is a major issue [11]. Therefore, to minimize costs (such as implementation, deployment, or help desk-related costs), the NIHDI is particularly keen to use the basic services currently available at a national level as far as is possible. Citizen identification and authentication facilities are provided through the Belgian e-ID card. Services such as professional authentication, data codification, and ciphering are provided through the Belgian eHealth platform[2]. However, these services have primarily been developed to be used in the field of the care and not specifically to support the privacy protection requirements of a research network.

In this research paper, we use a generic model and illustrations from the Belgian experience to investigate the extent to which the basic services developed to support the provision of care can be re-used in a semi-anonymous research network, while preserving the required level of privacy protection.

1. Methods

Drawing on previous work [3], and on over ten years of experience held by our research teams [5, 6], we defined a few basic privacy protection requirements and attempted to design the best (i.e. with the highest level of privacy protection) generic sustainable data flow and confidentiality management model. To ensure the sustainability of this model in the Belgian context, we identified the few additional technical services required to implement the model on a national scale. In 2010, this model was discussed and refined by various NIHDI working groups. Several potential national network architectures that would re-use some existing national services were considered, and their privacy protection levels were assessed using the generic model. The final network architecture selection was made by the NIHDI based on the acceptable privacy protection level of the selected architecture (i.e. a level that could be accepted by the Belgian Privacy Commission) and on many additional constraints such as time constraints, expected costs (according to the expected budgets available), and acceptability for professional associations and health insurance companies. The identification of an acceptable level of privacy protection is itself related to additional contextual constraints that may develop.

[2] cf. www.ehealth.fgov.be

2. Results

2.1. The Best Generic Sustainable Model

The key basic privacy requirement is that nobody should have access to any unnecessary information. If a "confidentiality hole" (i.e. access to information not required to perform a legitimate function) is unavoidable, then it should be reduced as much as possible and suppressed as soon as possible.

Patient and GP identification data is not required by the final research team. Therefore only coded (clinical) data needs to be available at the research institution, and they should be able to identify neither the patient nor the GP. So that data relating to a particular patient or GP may be linked, pseudonyms should be used. According to the Belgian Privacy law, these pseudonyms must be managed by a TTP. However, this TTP does not require access to clinical data (i.e. any data but the identification data) in order to function properly. Clinical data should therefore be encrypted from end to end.

Based on these requirements, a two-layered model was set up (cf. Figure 1). The lower level is dedicated to the message routing functions: identifying the receiver, de-identifying the sender, and managing multipart messages. At this level, all the patient data (i.e. medical and identification data) is encrypted. Only the information required for the message routing functions is accessible.

Patient data is only accessible at the upper level (the research and care level) by GPs and by the research team. GPs have access to all the identification data (IDs), while the research team only has access to the GP and patient codes (pseudonyms). Only the TTPs (in the upper layer) have access to the patient or GP IDs and codes simultaneously. A simplified view of the encrypted data flows is provided in Figure 2.

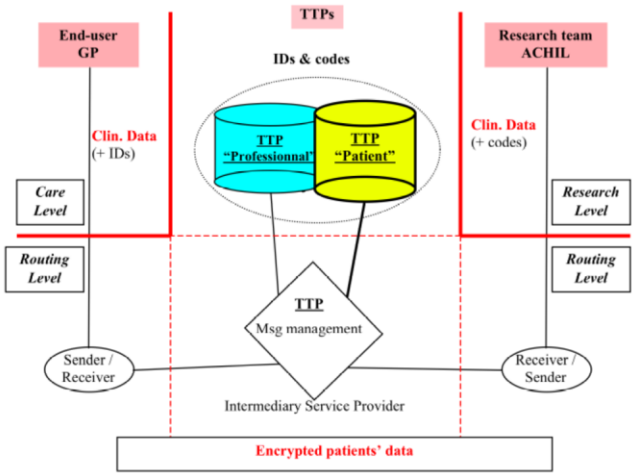

Figure 1. Generic confidentiality management model

In this "best" model, there are still some confidentiality holes. To avoid the uncontrolled coupling of research databases, research project-specific pseudonyms are being used. This means that when one requests the specific pseudonym of patient X for

research network Y, the TTP can deduce that patient X is suffering from the disease related to research network Y (in this case, to a specific care trajectory). Given the message-related architecture of the model, the TTP can tell whether a GP has not sent any message and therefore whether the GP has not fulfilled his/her contractual obligations. If several TTP functions are ensured by the same agent, further flaws could become apparent, such as evidence of a therapeutic link between a patient and a GP. To reduce these confidentiality holes, no sensitive information is stored at the TTP level.

The proposed model is sustainable in the Belgian context. We already have national patient IDs and codification services. We have an end-to-end encryption system with a strong sender identification mechanism (an essential requirement in the care field) and with certificates linking receiver IDs and public keys (asymmetric encryption service). To implement our model on a national scale we therefore need an encryption system that allows the de-identification of the sender and that has certificates linking public keys with pseudonyms (for instance when sending feedback to a GP, the research team knows only the pseudonym of that GP).

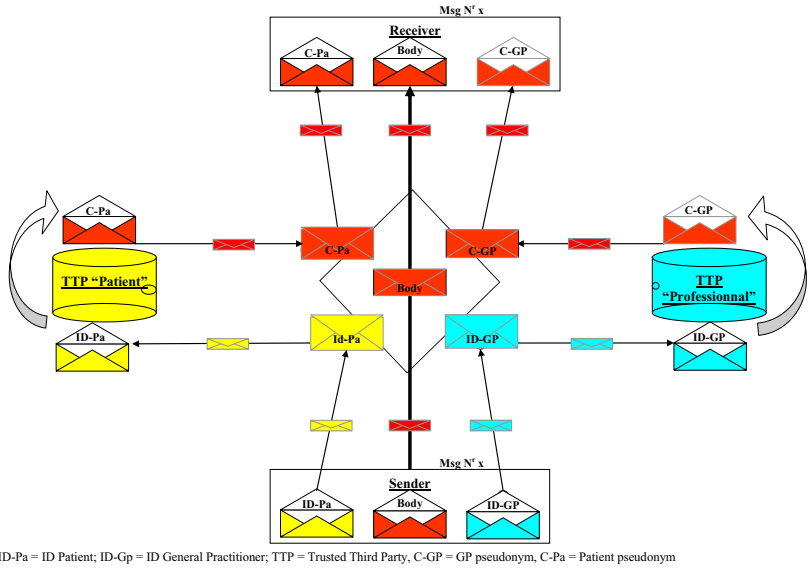

Figure 2. Data flows (simplified)

2.2. The Current Belgian National Architecture

The web application that is currently being set up (data collection 2012) is briefly described hereafter. Using his/her e-ID card and the e-health authentication service, the GP connects to the web application (https connection). He/she enters his/her national ID number and the patient's national ID number. The e-health coding services produce specific pseudonyms. The GP then enters various clinical data for the patient. This data could also be extracted from the GP's EPR and uploaded to the web application. The web application then performs some validation tests (data format, ranges; etc.). After successful controls a message is prepared for the research team that includes the pseudonyms and the clinical data. Using the e-health ciphering service, the message is

then encrypted, using a public key belonging to the MD responsible for the research team. The encrypted message is then sent to the research team. Following analysis, the research team will use the GP's pseudonyms to prepare individual feedback for the GPs. After identification and authentication, the GP may connect to the web application and request feedback. He/she will provide the application with his/her national ID number. Using the e-health coding service, the web application will receive the GP's pseudonym. Based on this pseudonym, the web application will upload the correct feedback, encrypt it (using the e-health ciphering service), and send it to the GP making the request.

In this architecture, the IDs and pseudonyms are clearly separated by the TTP: the research team only has access to the GP and patient pseudonyms; the GPs only have access to the IDs. Only an irreversible algorithm allows an ID to be transformed into a pseudonym. Only this algorithm is stored at the TTP. However, the TTP has access to the clinical data (no end-to-end encryption). To reduce and control this confidentiality hole, the data is used only by programs (no direct human access) within a secure environment (the TTP) during a short period of time (no data storage at TTP), and an oversight committee has been set up.

3. Conclusions

We built a generic confidentiality management model for semi-anonymous primary care health research networks. Even our best solution still has some (fleeting) limited confidentiality holes. The model is sustainable at a national level in Belgium and we think it could also be used in other countries. We used it to measure the gap between our best solution and various proposed nation-wide architectures, in order to come up with the most efficient architecture with an acceptable level of privacy protection.

In the near future, the Belgian architecture that is currently available should be improved to ensure end-to-end encryption. This could be supported by downloading the control and ciphering applications to GPs' computers or by providing the research team with pseudonym-related ciphering certificates. The usability of the web application for other research networks should also be considered.

Acknowledgment

The work was funded by the National Institute for Health and Disability Insurance.

References

[1] S. de Lusignan , C. van Weel, The use of routinely collected computer data for research in primary care: Opportunities and challenges, *Family Practice* **23** (2006), 253–263.
[2] D. Kalra, R. Gertz, P. Singleton, H.M. Inskip, Confidentiality of personal health information used for research. *BMJ* **333** (2006), 196-198.
[3] B. Claerhout, G.J.E. De Moor, Privacy protection for clinical and genomic data. The use of privacy-enhancing techniques in medicine, *Int J Med Inf* **74** (2005), 257-265.
[4] S. Bartholomeeusen, K. Chang-Yeon, C. Faes, R. Mertens, F. Buntinx, The denominator in general practice, a new approach from the Intego database, *Family Practice* **22** (2005), 442-447.

[5] E. De Clercq, H. Vandenberghe, P. Jonckheer, H. Bastiaens, M.F. Lafontaine, V. Van Casteren, Assessment of a Three-Year Experience with a Belgian Primary Care Data Network, In: F.H. Roger France, E. De Clercq, A. Hasman, G. De Moor(Edrs.) E-Health in Belgium and in the Netherlands, 163-169, Stud Health Technol Inform, Volume 93, IOS Press, Amsterdam, 2002.
[6] E. De Clercq, V. Van Casteren, P. Jonckheer, P. Burggraeve, Primary Healthcare Research Network: The Belgian ResoPrim Recommendations. In: K.-P. Adlassnig, B. Blobel, J. Mantas, and I. Masic (Edrs.) *Medical Informatics in a United and Healthy Europe – Proceedings of MIE 2009,* 38-42, Stud Health Technol Inform 150, IOS Press Amsterdam, Berlin Tokyo, Washington D.C., 2009.
[7] G. Goderis, E. De Clercq, V. Van Casteren, N. Bossuyt, S. Moreels, F. Buntinx, S. Bartholomeeusen, ACHIL : Griekse mythe of Belgishe realiteit ? Huisarts 40 (2011), 145-149.
[8] T. Bodenheimer, E.H. Wagner, K. Grumbach, Improving primary care for patients with chronic illness. *JAMA* **288**, 14 (2002), 1775-1779.
[9] A.E. Vlug, J. van der Lei, B.M. Mosseveld, M.A. van Wijk, P.D. van der Linden, M.C. Sturkenboom, et al., Postmarketing surveillance based on electronic patient records: The IPCI project. *Methods of Information in Medicine* **38**, 4-5 (1999),:339–344.
[10] N. Pearson, et al., Collecting morbidity data in general practice: the Somerset morbidity project. *BMJ* **312** (1996), 1517-1520.
[11] K. Helbing, S.Y. Demiroglu, F. Rakebrandt, K. Pommerening, O. Rienhoff, U. Sax, A Data Protection Scheme for Medical Research Networks. Review after Five Years of Operation. *Methods Inf Med* **49**, 6 (2010), 601-607.