Automated regulatory constraints and data governance for healthcare

Milen G. Kebede, Giovanni Sileno, Tom Van Engers
Complex Cyber Infrastructure group, University of Amsterdam, Amsterdam, The Netherlands
m.g.kebede@uva.nl

Keywords— consent-aware research, data governance, data sharing, healthcare, semantic web

Motivation

Motivation The EU general data protection regulation (GDPR) has undoubtedly raised the bar for the standard of data protection in the current big data era. Several non-European countries are taking this regulation as a model for their own legislation. From a engineering perspective, in order to share healthcare data between public and private organizations, we need to implement functions that operationalize (at least to an adequate extent) data protection directives as the GDPR, as well as ad-hoc normative artifacts signed by the parties involved–typically data-sharing agreements between institutes, consents given by patients. The GDPR itself states that consent is one of the grounds on which the processing of personal data is considered legitimate (Art. 7); it also ensures transparency of processing by demanding data controllers to provide clear and concise information notices to the data subject (Art. 12), etc. More in general, the GDPR data protection principles touch upon concerns such as fairness, lawfulness, transparency, purpose limitation, data minimisation, data quality, security, integrity, and confidentiality.

Even though data-sharing is a sensitive subject for medical and research purposes and should be given caution (patient data is a prototypical example of personal data), in practical terms it is not the primary concern. Research institutes in healthcare primarily aims to improve and find innovative cures or insights. For this reason, compliance to data sharing regulation becomes a burden more than a duty. Consequently, even if the privacy requirements that the GDPR presents are meant to empower the data subject for the control over their data, there remains a gap between the intent of the GDPR and current data sharing systems in medical and research institutes [1]. Consent and other forms of agreements between stakeholders are for the most part still managed manually, or mapped to simple computational patterns that do not fully capture the interactions expressed in regulatory sources. First, knowing that regulations will be applied by the infrastructures, registry maintainers will be less conservative to data sharing [3]. Second, the provision of consent-related services will empower patients in maintaining ownership of their data, and to get transparent information about its usage. Third, automating regulatory constraints will reduce the burden of healthcare workers, so that they can focus on medical research. In the long run, it will also support the standardization of data sharing practices among institution.

Purpose

The main objective in this research [7] concerns how the processing of regulatory and organizational requirements, consent, medical necessity and agreement-based access-to-data can be automatized within a data-sharing infrastructure, preventing and responding to breaches by embedded compliance measures. The intended outcome is an open authorization and governance solution, i.e. able to trace back any decision regarding consent, agreements and regulations. The following questions are currently investigated:

1. What are the key concepts and patterns for modeling consent, data sharing agreements, relevant regulations for an effective and compliant processing of data in healthcare?
2. What are the infrastructure and organizational requirements necessary to develop and maintain a legally-aware data sharing infrastructure in healthcare?

The first question builds upon an existing literature (on the challenges) on the representation of norms. The second question concerns both operational problems (e.g. execution of inferential tasks) problems, and informational problems (e.g. adequate interfaces for acquisition and explanation).

Use Case

The use case associated to the project is the Diffuse Intrinsic Pontine Gliomas (DIPG) registry managed by the Princess Maxima Center for pediatric oncology. The registry contains data from more than 700 children in Europe and from 33 countries. Data consist of MRI images, clinical features (history, symptoms, physical examination), treatment and outcome data, as well as biological data from tumor biopsies taken at diagnosis and/or from autopsies. DIPGs are largely fatal, rare brainstem tumors in children that need innovation in treatment. For this reason, there is an urgent need to share data among institutions that reside in the same country as well as institutions across countries. This use case provides the perfect environment where to test proof of concepts concerning the three main axes of the research: representation, execution/inference, and interface.

Initial plan of work

On the representational axis, we need to semantically define the concepts of informed consent and other forms of agreement, so as to describe the actors involved in the process of data sharing, their duties and powers. Recently, there has been a lot of work done in relation to compliance to the GDPR; most of these works focus on formal ontologies [2]. Semantic web technology is in principle the most suitable technology, also for its wide use for data-sharing infrastructures (e.g. FAIR data points or hubs), but it is debated whether it is suited for normative reasoning [5].

On the execution axis, we observed that most electronic health records use access control methods to give access to legitimate users. Therefore, after capturing the regulatory constraints, we must put in place the right access control method to ensure confidentiality and integrity. Several types of access control methods are used in literature [8]; the role based access control (RBAC) model is deemed to be the most common [9]. However, RBAC might exhibits limitations in an environment where roles are changing dynamically. Since our goal is to empower patients for the control of their data; we need therefore to think of a sort of consent-based access control model, where patients are in principle able to communicate in real-time their granting or denying access (under certain conditions) to their data. We need therefore to understand to what extent the current standards in access-control can be integrated with the normative reasoning constructs explored on the representational axis. Second, adequate monitoring needs to be put in place for the occurrence of breaches; it remains to be understood what kind of technology is currently used in the healthcare sector, and how it can be automatically controlled on the basis of regulatory directives.

For the interface axis, we need to think about how users (e.g. data subjects and controllers) can interact with our system. We plan to use human-centered approaches to visualize and present out the functioning and the outcome of the automated data sharing system. Some challenges we expect concerning this are that consent and agreement forms can be quite extensive and complex. Specifying patients' preferences impacts consent and how it is given [6]. Factors like trust relationship, harm threshold, balance of risk and benefits, transparency of data exchange and access control have been identified as factors influencing the patient's decision making [4]. We can argue that capturing consents and agreements in a way that is readable and comprehensive to the data subject can improve their willingness to share data for research. Therefore we need to understand how the two interfaces (acquisition and explanation) can be tailored, using available technologies, with the content captured through the work on the representational axis, to achieve a satisfactory result for the user.

Conclusion

The objective of this abstract is to show existing practical problems and research gaps concerning data-sharing infrastructures for medical research. It presented part of the directions we are currently investigating in order to solve these issues. This work is part of the Enabling personalised interventions (EPI) project that aims to empower patients and providers through self management , shared management, and personalisation across the full health spectrum [7].

References

[1] Gerl, A., Bennani, N., Kosch, H., Brunie, L. (2018). LPL, towards a GDPR-compliant privacy language: Formal definition and usage. In Transactions on Large-Scale Data- and Knowledge-Centered Systems XXXVII (Vol. 10940 LNCS, pp. 41–80).

[2] Li, M., Samavi, R. (2018). DSAP: Data Sharing Agreement Privacy Ontology. Proceeding of the 11th Semantic Web Applications and Tools for Healthcare and Life Sciences (SWAT4HCLS).

[3] Pandit, H. J., Debruyne, C., O'Sullivan, D., Lewis, D. (2019). GConsent - A consent ontology based on the GDPR. European Semantic Web Conference (ESWC 2019), 11503 LNCS, 270–282.

[4] Lisa A Moon. 2017. Factors influencing health data sharing preferences of consumers: A critical review. Health Policy and Technology.

[5] SONG Wei, ZHANG Ming. Concise Guide to the Semantic Web [M]. Beijing: Higher Education Press, 2004, PP: 56-60

[6] Patil, Sunil, et al. "Public preferences for electronic health data storage, access, and sharing—evidence from a pan-European survey." Journal of the American Medical Informatics Association 23.6 (2016): 1096-1106.

[7] EPI project https://delaat.net/epi/

[8] Narayanan, Hema Andal Jayaprakash, and Mehmet Hadi Güneş. "Ensuring access control in cloud provisioned healthcare systems." 2011 IEEE Consumer Communications and Networking Conference (CCNC). IEEE, 2011.

[9] Calvillo-Arbizu, Jorge, Isabel Román-Martínez, and Laura M. Roa-Romero. "Standardized access control mechanisms for protecting ISO 13606-based electronic health record systems." IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI). IEEE, 2014.

[10] Palmirani, Monica, et al. "Legal Ontology for Modelling GDPR Concepts and Norms." JURIX. 2018.