Consent Management Architecture for Secure Data Transactions

Jarkko Hyysalo, Harri Hirvonsalo, Jaakko Sauvola and Samuli Tuoriniemi

Faculty of Information Technology and Electrical Engineering, University of Oulu, Pentti Kaiteran Katu 1, Oulu, Finland

Keywords: Consent Management, Digital Health, Data Intensive Services, Sensitive Transactions, Security Architecture.

Abstract: Digitalization of data intensive services presents several challenges, such as how to safely manage and use the multitude of personal data across various public, private and commercial service providers. Guaranteed privacy is especially critical in sensitive cases like health data management and processing. A key challenge and enabler for efficient data utilization is the need for an adequate consent management framework that meets the General Data Protection Regulation (GDPR). To facilitate sensitive secure data transactions where end-control always resides with the individual, a consent management architecture (CMA) is defined, utilizing the new MyData approach. The proposed CMA enables context-driven authorization of multi-sourced data for safe access by various health services. CMA proof-of-concept and experiences are described and discussed to concretize and evaluate the suggested architecture. Consent management and authorization topics are discussed as a service function of the MyData Operator. The technical APIs required for registering and authorizing data sources and data services via the Operator are demonstrated and analyzed to expedite development of this important area within the research and industrial communities.

1 INTRODUCTION

Digitalization of data-intensive domains with associated analytics and services is of growing economic importance. Enabled through new avenues of innovation, the many benefits range from basic process efficiencies and cost savings to the creation of new service and business models to maximize customer reach. Creating, managing and utilizing large volumes of relevant information generated by multiple sources and entities is required for value-added service delivery, for efficient professional practice or, yet more critically, to make timely decisions in a healthcare setting—for instance, treatment of an illness or execution of clinical practice. Preventive and reactive healthcare is among the most impacted domains now under pressure to break with old conventions on the promise of innovation and increased efficiencies across care chains and personalized services. The health domain has multiple opportunities to create new value by exploiting the sourcing of relevant digital data for useful services and applications.

Digitalization and genuine digital generation of health data that produces a continuous flow of historic and present events as interoperable records offer significant opportunities for progress in this sector—for example, in self-management of personal health information. Developments in digital health have attracted considerable research interest in recent years. Steinbrook (2008) envisioned new personalized health services based on better data sharing with doctors, emergency departments, and family members; including prescription, fitness and diet management services; communication with others with similar health problems; and information about participation in clinical trials. Steinbrook also identified new problems arising from digitalization, such as inaccurate data, risk of losing control of personal data, lack of data verifiability and the uncertain implications of new revenue-based health models involving large corporations such as Google, Microsoft and others. In a review of electronic personal health record systems, Archer et al., (2011) found that dependency on primary care physicians’ electronic medical record systems is hindering progress toward broader adoption of personalized health management. Tucker et al., (2009) identified new opportunities created by cost-effective, massively parallel sequencing for medical
genetics, disease research and personal genomic profiling. In summary, these new opportunities also bring challenges related to the use of new information in a socially responsible manner.

To expedite the promise of digitalization, requirements for a suitable digital health and data-intensive service architecture in a many-to-many actor and owner environment must be identified. While there are many good vertical approaches and implementations, there are very few scalable generic solutions other than commercial one-provider systems. Our approach adopts the open MyData architecture to specify a system having multiple data sources with legal and technical touchpoints in support of new levels of personal data ownership and control (Byström et al., 2015). This approach aligns well with the ongoing work on the European General Data Protection Regulation (GDPR) and takes account of various stakeholders in a) generating personalized data and b) offering value-added services based on the multiple data sources. This novel approach accumulates value from mass data, offering meaningful and highly relevant services, both to individuals and to larger interest groups. Based on the assumption of personal data management, where the person is effectively in charge of their own data, the key issue, both technically and behaviorally, is how the individual’s consent is to be managed, and how this works in practice across multiple data sources, service actors and end users such as healthcare professionals, commercial entities or other individuals or groups. Consent can be defined as permission to process personal data and must always be given by the Data Subject to the Data Controller. Consent must be a freely given, specific, informed and unambiguous indication of the Data Subject’s wishes.

The present paper focuses on the concept of consent and consent management architecture (CMA) in a distributed, multi-actor, multi-source data environment as in Poikola et al.’s (2015) MyData approach. We propose a CMA as a basis for enabling secure transactions; for authorizing data access to services; and for health-related personal data management and processing. We believe that CMA is a crucial enabler in sharing and utilizing personal data in health information systems and services. In response to the research question How should a consent management framework be designed and built following GDPR?, we address the implications of sourcing such data and how service-provider management and processing of personal data might accommodate the individual as key consent giver. To address this question, we propose a novel consent-based authorization architecture, which is influenced by User Managed Access (UMA) protocol. In addition, we describe proof-of-concept implementation to evaluate the potential of our CMA framework and to prove its value. To the best of our knowledge, no such implementation of a general purpose consent management framework and architecture exists in the literature or among commercial solutions that meet the GDPR.

The paper is structured as follows. Section 2 discusses related work. Section 3 outlines the proposed consent management architecture for secure transactions. Section 4 describes the implementation of our architecture framework and describes our experiences. Section 5 discusses the results, and Section 6 presents conclusions.

2 RELATED WORK

Efficient management of Electronic Health Records (EHRs) is essential for the creation of new value for healthcare processes and treatments. However, within the healthcare domain, given all its legacy systems, actors and stakeholders, EHRs are seen as a complex and sensitive issue (Jin et al., 2011). The most common access control model used in healthcare domain is Role-Based Access Control (RBAC), however, it seems that in most cases, only the healthcare professionals, general practitioners, IT and pharmacists can access the data (Ferreira et al., 2007), which is not in line with the GDPR and MyData principles.

The challenges for health-related personal data management and processing include also issues of interoperability; compatibility of user devices with third party devices, external sensors or measurement devices; and other issues of usability (Milenković et al., 2006; Liu et al., 2011a; Liu et al., 2011b; Jovanov and Milenković, 2011). Limited standards have also been identified as problematic (Liu et al., 2011b), including existing wireless communication standards that fail to provide a secure and trusted communications infrastructure between devices, sensors, actuators, health providers, network providers and so on. Another issue relates to the requirements of emerging body area networks—for example, interference problems or security concerns (Pantelopoulos and Bourbakis, 2010). There are also severe security-related concerns relating to lack of confidentiality (Pantelopoulos and Bourbakis, 2010; Liu et al., 2011b).

Al Ameen et al., (2012) elaborated further on security concerns, separating these into two
categories: system security (administrative level, physical level, technical level) and information security (data encryption, data integrity, authentication and freshness protection). One major unanswered question concerns who should manage the data (Milenković et al., 2006; Liu et al., 2011b; Jovanov and Milenković, 2011), and privacy and consent-related concerns were also common (cf. Liu et al., 2011b; Al Ameen et al., 2012).

The forthcoming GDPR presents new challenges. Traung (2012) argued that considerable work is required to achieve the goals of GDPR. For example, by virtue of its length, complexity and vague rules, GDPR leaves room for interpretation in terms of its implementation (Kuner, 2012; Koops, 2014), and the text is too complicated to be understood by data subjects and data controllers (Blume, 2014). As GDPR is a long and ambitious text (De Hert and Papakonstantinou, 2012), we have elected to focus on the consent management issues. This is because we consider the consent management framework to be a crucial enabler for sharing and utilizing personal data in health information systems and services—but, as noted above, there is at present no such framework. For example, in the U.S. healthcare industry, Electronic Medical Records (EMRs) have been widely adopted, but consents are still managed primarily by pen and paper or scanned electronic documents (Yu et al., 2014). According to Kaye et al. (2015), the situation in Europe is similarly paper-based. We therefore argue the need for a simple general purpose consent management framework to enable movement toward full digitalization and genuine digital generation and utilization of health data.

Table 1 sets out the key issues and requirements for consent management as it relates to GDPR.

A generic, human-centered framework for consent management is of paramount importance, as GDPR requires that consent should be freely given, for a specified purpose, informed and unambiguous in confirming the data subject’s agreement to processing of their personal data.

Derived from the literature, the requirements are addressed within our consent management framework and indicate how to build a general purpose consent management architecture that aligns with the GDPR.

### 3 CMA OVERVIEW

The GDPR is in later stages of preparation and will be deployed for personal data-related information

<table>
<thead>
<tr>
<th>TOPIC</th>
<th>ISSUE</th>
<th>REQUIREMENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROLES</td>
<td>Vague and outdated delineation of roles (De Hert and</td>
<td>The framework shall identify and define the roles and</td>
</tr>
<tr>
<td></td>
<td>Papakonstantinou, 2012) and responsibilities (Blume, 2014)</td>
<td>responsibilities of important actors.</td>
</tr>
<tr>
<td>RULES</td>
<td>Vague rules for further processing of collected data beyond the</td>
<td>The framework shall define the consent management</td>
</tr>
<tr>
<td></td>
<td>originally specified purpose (De Hert and Papakonstantinou, 2012)</td>
<td>mechanisms required for further processing of data.</td>
</tr>
<tr>
<td>WITHDRAWAL</td>
<td>Individual right to withdraw consent</td>
<td>The framework shall provide mechanisms for consent</td>
</tr>
<tr>
<td>PRIVACY POLICIES</td>
<td>Different privacy policy requirements (De Hert and</td>
<td>withdrawal.</td>
</tr>
<tr>
<td></td>
<td>Papakonstantinou, 2012) and different laws (Blume, 2014) in different</td>
<td>The framework shall have the necessary readiness and</td>
</tr>
<tr>
<td></td>
<td>EU member states</td>
<td>capability to adapt to different privacy policies.</td>
</tr>
<tr>
<td>DATA USAGE</td>
<td>Lack of awareness of when and how one’s own data are</td>
<td>The framework shall provide transparent processing</td>
</tr>
<tr>
<td></td>
<td>being processed (De Hert and Papakonstantinou, 2012) Data</td>
<td>of a person’s own data.</td>
</tr>
<tr>
<td></td>
<td>controller’s obligation to delete information as soon as it is</td>
<td>The framework shall enable data removal.</td>
</tr>
<tr>
<td></td>
<td>no longer needed (De Hert and Papakonstantinou, 2012; Blume, 2014)</td>
<td>The framework shall enable data portability.</td>
</tr>
<tr>
<td>PORTABILITY RIGHTS</td>
<td>Individual’s right to obtain a copy of their profile as</td>
<td>The framework shall enable interoperability between</td>
</tr>
<tr>
<td></td>
<td>uploaded to internet platforms in a suitable format for further</td>
<td>systems.</td>
</tr>
<tr>
<td></td>
<td>personal processing and use (De Hert and Papakonstantinou, 2012;</td>
<td>The framework shall guarantee the individual’s right to</td>
</tr>
<tr>
<td></td>
<td>Voss, 2014)</td>
<td>manage and control personal data.</td>
</tr>
<tr>
<td>INTEROPERABILITY</td>
<td>No mention of system interoperability in EU law (De Hert and</td>
<td>The framework shall provide comprehensible policies</td>
</tr>
<tr>
<td></td>
<td>Papakonstantinou, 2012)</td>
<td>(Gnesi et al., 2014).</td>
</tr>
<tr>
<td>MANAGEMENT OF INDIVIDUAL data</td>
<td>Very limited individual capability to manage one’s own data</td>
<td>The framework shall guarantee the individual’s right</td>
</tr>
<tr>
<td></td>
<td>(Tene and Wolf, 2014; Gnesi et al., 2014)</td>
<td>to manage and control personal data.</td>
</tr>
<tr>
<td>PRIVACY POLICIES</td>
<td>Low level of individual expertise to analyze and manage complex</td>
<td>The framework shall provide comprehensible policies</td>
</tr>
<tr>
<td></td>
<td>privacy policies (Tene and Wolf, 2014; Gnesi et al., 2014)</td>
<td>(Gnesi et al., 2014).</td>
</tr>
</tbody>
</table>
management during 2016-2018. To address the lack of a general purpose consent management framework, we propose a person-managed data rights architecture framework, following the approach and key principles of MyData (Poikola et al., 2015; Byström et al., 2015) and the key functional requirements of the GDPR. The top level requirements state that a) individuals should have the right and the practical means to manage their data and privacy according to GDPR attributes; b) data should be easy to access and use; c) there should be a way to convert data from single entities into a meaningful resource that can be used to create new services; d) in support of an open business environment, the shared data infrastructure should enable the coordinated management of personal data, should ensure interoperability and should make it easier for different entities (e.g. companies, public officials and public-private service partnerships) to comply with tightening data protection regulations; and e) individuals should be able to change service providers and to control their data manager(s).

The CMA framework is based on the definition of user accounts as i) interoperable and ii) standardized by account model, managed control and data flows through API practices. User accounts can be held and managed by one or more trusted operators. The key concepts are MyData Operator and MyData Account. MyData Operator enables digital consent management and provides MyData Accounts and related account management services. MyData Account will host all service contracts and consent receipts, defining access means and rights to Data Sinks and Data Sources. MyData Account is the interface that enables individuals to manage their personal data across multiple data sources, providers and individuals.

Other key roles include Account Owner, Data Source and Data Sink. Actors can take one or more operational roles; typically, the same organization would act as Data Source and Data Sink. Account Owners are individuals who create and use MyData Accounts to link services and authorize data flows with consents. Data Sources provide the data for services that use the data (Data Sinks).

The key requirement is to offer logical paths enabling the data owner (an individual) to control their personal data as an unambiguously defined entity during the multiple creation, storing and processing events by viable Source(s) and Sink(s). User accounts can be provided either by organizations acting on behalf of the account holder or by account holder(s), who can set up their own user account services (as when hosting a personal web service or e-mail server), offering the accounts as a service.

By virtue of the open architecture and consent model, developers can build access and services to this framework using public programming interfaces and libraries.

The CMA separates flows of data according to purpose and usage rights. Consent permissions (e.g., Protection API and Authorization API and the concept of Resource Set) from UMA protocol specifications, but does not conform to UMA protocol flow.

Table 2: Summary of APIs with providers, utilizers and descriptions.

<table>
<thead>
<tr>
<th>NAME</th>
<th>PROVIDER</th>
<th>UTILIZER</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Protection API</td>
<td>Operator</td>
<td>Data Source</td>
<td>- Introspection: Verifying authorization status of data request</td>
</tr>
<tr>
<td>Authorization API</td>
<td>Operator</td>
<td>Data Source</td>
<td>- Refreshes expired proof of authorization when consent is still valid</td>
</tr>
<tr>
<td>Data Sink API</td>
<td>Data Sink</td>
<td>Operator</td>
<td>- Operator-delivered consent and proof-of-authorization</td>
</tr>
<tr>
<td>Data Source API</td>
<td>Data Source</td>
<td>Operator</td>
<td>- Consent management in general (CRUD)</td>
</tr>
<tr>
<td>Data API</td>
<td>Data Source</td>
<td>Data Sink</td>
<td>- Operator-delivered consent</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Consent management in general (CRUD)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>- Sink request data from source with resource identifier</td>
</tr>
</tbody>
</table>
protection, authorization and control) are managed by means of a separated set of interfaces and programming instances, allowing new consent operators, services and applications to evolve. The actual personal data flow used by the services is always executed as a transaction between Source(s) and Sink(s) entities, with eligible permissions by operator.

As illustrated in Figure 1, this architectural arrangement allows for separation of actual user account management with consent services and logged service data flows only between data Source(s) and user(s). In this way, the operator never stores any personal data (e.g., health data) generated by any Source(s) but acts as a trusted consent manager, assigning rights or limits to the use of a present user’s data, on their behalf.

With clearly defined APIs (see Table 2 for summary) and operator(s) registry principles, the operator is able to offer a consolidated view to the present user account holder of granted permissions, permission requests and changed permissions. Source and Sink entities can utilize the generic CMA framework to access the consent service by operator APIs. Transfer of data from Source to Sink happens through Source-provided Data API that should be accessible by a Sink, either directly or through a proxy service, combining interfaces from many Sources (e.g., public officials in a health care region). Issues impacting interoperability such as discovery of compatible Sources and Sinks, trust between networks and data formats have been addressed, but they are beyond the scope of this paper.

4 IMPLEMENTATION OF THE CMA FRAMEWORK

To validate the feasibility of the proposed consent management framework, we have made a proof-of-concept implementation of the operator and minimum viable stub implementations of Source and Sink. APIs of all entities are implemented as REST (Representational State Transfer) APIs.

The framework was implemented using Python language, which was selected because it enables rapid development and prototyping. We made (heavy) use of Flask, a Python framework for building RESTful web services and web applications. We also utilized Flask-extension Flask-RESTful, which eases writing of REST APIs. Flask also provided support for SQLAlchemy, which we used as Object Relational Mapper (ORM). For actual database implementation, we originally chose SQLite, but as the implementation progressed, we found it necessary to switch over to MySQL.

In its current state, our framework enables Account Owner to complete the necessary actions to establish a consented data flow of their data from a Source to a Sink, involving three main steps: Service Connection, MyData Authorization and Data Connection. Each step can be performed only after the previous step is completed.

4.1 Service Connection

In the service connection step, depicted in Figure 2, a service (i.e., a Source or a Sink) is connected to a person’s user account on an operator. During service connection, the person’s account in the service is associated with their user account on the operator. Being connected, operator and service also exchange information needed to ensure secure communication of consent information (e.g., agreeing on cryptographic keys used for signing of consent receipts and proof-of-authorization sent to a Sink).

Figure 2: Service connection.

4.2 MyData Authorization

In the MyData Authorization step, depicted in Figure 3, a Source is given consent to provide a person’s data, and a Sink is given consent to fetch this data from the specific Source. In other words, Source is authorized to provide data, and Sink is authorized to fetch data. In order to unambiguously identify data referenced by these consents, Operator generates a resource set, describing data the user has elected to share to Sink and the resource set identifier, which is used to refer to the previously generated resource set. Operator delivers the resource set and its corresponding ID to Source within Source’s consent. Only the resource set ID is delivered to Sink within Sink’s consent. Sink must use this identifier in data requests it makes to Source. It is worth noting that Source is not
explicitly instructed to provide data to a specific Sink as such. On receiving a request for data, Source is expecting proof that Sink making the request is authorized to request the data in question. Operator delivers this proof-of-authorization to Sink along with Sink’s consent, and Sink must deliver this proof with every data request it makes to Source.

4.3 Data Connection

In the Data Connection step, depicted in Figure 4, Sink makes a request to Source for data, using the resource set identifier delivered by operator in the previous step. This request must contain proof-of-authorization to request the data referenced by resource set identifier. Upon receiving such a request, Source executes an introspection step, in which it queries operator about the validity and extent of the provided proof-of-authorization by delivering it, along with the requested resource set ID, to operator. Operator verifies that the received proof-of-authorization is in fact issued by operator and that the consent associated with this proof-of-authorization is still valid (and not, for example, withdrawn). Operator also verifies that the resource set Sink has requested is the same resource set as it was authorized to have. Based on operator’s response to introspection, Source either rejects and denies data access or accepts Sink’s data request and delivers data to Sink.

5 DISCUSSION

Although relatively simple, the CMA framework still addresses the set requirements, providing a simple, general purpose consent management framework and architecture that conforms to the GDPR. In our framework, the main focus was on consent and how consent can be used. This is somewhat similar to what the UMA protocol defines, but our implementation focuses on use cases in a distributed, multi-actor and multi-sourcing data environment while following requirements imposed by GDPR.

CMA provides one example on how to implement the MyData approach. On our approach, data subjects can manage their own consents easily, following the core idea of MyData—that individuals are in control of their own data. The implementation presented in this paper offers a novel way of addressing the forthcoming GDPR. It provides a model for human-centred personal data management using a consent-based approach realized within a generic consent management architecture framework, with clarifying roles, responsibilities and liabilities in line with GDPR and the MyData approach. This is a practical technical solution that can be applied to the growing need of individuals to control their own data, as well as the need of organizations to meet the requirements of the GDPR.

Returning to the research question, How should a consent management framework be designed and built following GDPR?, we have presented a proof-of-concept implementation of a CMA architecture influenced by UMA specifications. The implementation emphasizes the role of MyData Operator, which helps to introduce the Sources and Sinks to each other, simplifying the authorization process by comparison with the UMA specification. Another key element is MyData Account, a personalized communication interface through which individuals can view, manage and control their consents easily, in a transparent and standardized way. This interface can also be accessed at any time, and standardization is emphasized to enable interoperability. The ability to transfer consents enables individuals to choose and change service providers easily. Improved interoperability also enables easier secondary use of data as suggested by Cresswell et al., (2013).

In summary, the theoretical implications here create new knowledge for organizations developing solutions for personal health information systems and services. This work identifies current challenges
in consent management and describes potential solutions for tackling these challenges. The implications for practice include the experiences and solutions and, more specifically, our framework, which offers a means of realizing a consent management architecture.

Our study also indicates issues and directions for future research, such as the need for a better understanding of consent management mechanisms and architecture. As our consent management architecture is defined at a general level and is not a functional or requirements specification, further work is still needed. Our future work will include more thorough specification and empirical evaluation of our framework, especially with regard to performance and usability, in collaboration with industry and research organizations.

6 CONCLUSIONS

Significant amounts of personal data are currently collected by different applications and services, which can be used for further processing with e.g. monetizable outcomes. To date, individuals have typically had little or no control over how their data are created or used. Privacy laws set strict requirements for collecting, processing and sharing personally identifiable information, and utilizing personal data must begin from free, informed, specific and explicit consent given by the data subject. Domain specific solutions for consent management do exist, but there is strong demand for more generic cross-domain solutions that would enable new and legacy systems to share and use personally identifiable information. Our research question was addressed by means of a novel consent-based authorization architecture, and in addition we presented a proof-of-concept implementation and discussed our experiences during implementation.

This paper makes an important contribution to the secure digitalization of data and personalization of services in the domains of health and information systems. Furthermore, our results enable the creation of new digital health solutions by virtue of a novel privacy-preserving architecture. Organizations will also benefit from practical methods of securing an individual’s consent to the use of personal data, so improving opportunities to utilize those data to provide innovative services. Finally, our implementation confirms that the technologies already exist to build CMA, although some still need further development. Improvements are proposed, such as the centralized operator (MyData Operator), which lends novelty to our approach, and we recommend consideration of MyData Operator’s role in addition to UMA-specified roles to simplify the authorization process.

This paper has its code available at https://github.com/dhrproject.

ACKNOWLEDGEMENTS

This research has been supported by a grant from Tekes—the Finnish Funding Agency for Innovation as part of Digital Health Revolution programme. The multi-disciplinary programme is coordinated and managed by Center for Health and Technology, University of Oulu, Finland.

REFERENCES


