Concept and Implementation of Data Usage Proposal Process Based on International Standards in SMITH

Marcel KLÖTGEN[a,1], Eric FIEGE[a] and Salima HOUTA[a]

[a] *Fraunhofer Institute for Software- and Systems-Engineering ISST*

Abstract. Accessing secondary-use healthcare data in Germany requires contracting with each organization that acts as a data provider. The SMITH Service Platform offers a central access point for scientists, facilitating contracting as part of an integrated data use and access process with several Data Integration Centers (DIC) at once. Process support is realized by a central Business Process Engine (BPE), which manages process definition and process control, combined with a central IHE infrastructure. The use of IHE XDS and IHE XDW profiles enables the exchange of process instance information with multiple distributed visualization and user interaction tools for provided user tasks based on international standards. User task information include structured forms for submitting instructions and results as task input and output for the users, and are synchronized between the shared process instance and the BPE. A reference user interface is also provided with the SMITH Marketplace. In the future, further standardization efforts regarding the structured forms and the use of the IHE XDW profile should be pursued.

Keywords. medical informatics, health information interoperability, workflow, data science

1. Introduction

1.1. Background

In Germany, medical research and big data analytics are increasingly important for healthcare. There is high potential in reusing, linking and analyzing data from different healthcare-related sources in order to discover new or optimize existing and possibly highly variable treatment methods, pharmaceuticals or technologies (1). Yet, there is no coordinating access and contracting point for scientists regarding medical research in Germany that manages or facilitates the proposal, evaluation and contracting process with multiple healthcare organizations or hospitals nor an integration and aggregation of different data sources and contracting parties. Instead, a scientist, or data consumer, usually contacts each healthcare organization, or data provider, and starts dedicated contracting and evaluation processes based on a *data usage proposal*, taking into account different legal frameworks, technologies, processes and information models. Moreover, a scientist, or data consumer, would have to handle each provided data set differently.

---

[1] Corresponding author: Marcel Klötgen, email: marcel.kloetgen@isst.fraunhofer.de

Smart Medical Information Technology for Healthcare (SMITH) is one of four consortia of the Medical Informatics Initiative (MI-I) (2), specifically trying to facilitate the secondary-use of data gained from patient care and striving for a centralized and overarching concept for data ownership and data sharing by establishing synchronized Data Integration Centers (DIC) at each partner hospital (3), (4). In order to enable the provisioning and sharing of secondary-use data as well as to ensure processual and semantic interoperability with a potentially growing number of partners in the near future, the SMITH consortium implies the use of international standards. Therefore, the SMITH architecture is based on IHE Profiles and HL7 FHIR (3).

1.2.   Requirements

The representation of information from patient care in a common information model largely facilitates the reuse of data for research purposes (5). This allows for an overarching platform that supports shared use cases, such as a centrally managed *data use and access* process for secondary-use data. The *data use and access* process consists of a sequence of tasks; a task describes the actions to be performed in a certain environment by a human or service performer. A centralized platform enables data consumers to use a single point of interaction instead of multiple, distributed and often-undefined interfaces. A common *data use and access* process connects and integrates potential data providers and simplifies the complexity of previously fragmented tasks regarding a *data usage proposal*. It has to fulfill certain requirements:

- *Scalability*: The platform must be able to address and support different requirements of data providers such as different legal frameworks, ethical assumptions, or other local requirements. The technical framework must therefore allow for different implementations and executions of the same task.
- *Process interoperability*: The platform must support the integration of multiple software tools for visualization and execution of tasks. The representation of processes and tasks must therefore be accessible and interoperable for all software components to ensure reliable execution.
- *Information interoperability*: The platform must provide and share information that enable users to execute and fulfill a task. All instructions and information necessary to fulfill a task must be shared and displayed and new information raised by a performer must be provided to the platform.
- *Transparency:* Users, especially the data consumer, must be able to track the status of the process instance and user tasks. Performers must have access to information previously provided by other users.

This work focuses on the description of a system architecture model for the centralized platform, called SMITH Service Platform (SSP), and the concept and implementation of a common *data use and access* process that handles *data usage proposals* and contracting for all involved participants, yet explicitly excluding the actual data delivery. The *data use and access* process is therefore comprised of a sequence of user tasks, which represent the necessary actions to be performed by humans.

2.   State of the art

Several recent projects and institutions have produced results and inspiration for the idea of a cross-organizational and coordinated *data usage proposal* management. One of these institutions is the LIFE Research Center for Civilization Diseases of the University Leipzig (6). The LIFE Research Center for Civilization Diseases acts as a data provider for secondary-use data gained from patient care of the university hospital Leipzig. The LIFE data sharing approach incorporates the OAIS reference model (7) and covers the needs for organizational structures and IT infrastructure for the management of *data usage proposals* and data access. An organizational entity, the Data and Metadata Transfer Group, acts as a coordinator for the *data use and access* process, while the Use & Access Board evaluates current *data usage proposals* based on a document template and provided as hard copy. For the most part, the *data use and access* process is based on a state model of the *data usage proposal*. Thus, LIFE is able to provide its managed data sets, yet lacking the capabilities to coordinate cross-organizational data provision for different data providers.

ProSkive represents a similar approach for managing *data usage proposals* for secondary-use data and biomaterial currently in development, delivering a web interface based on independent microservices for initiating and tracking project proposals (8). However, no indication is given about scalability or process interoperability.

Integrating the Healthcare Enterprises (IHE) aims at improving interoperability by defining which established standards should be used for the most common use cases of the healthcare sector. The IHE Cross Enterprise Document Workflow (XDW) profile provides a structure for the exchange of process information (9), yet only focusing on user tasks of a process instance rather than process control and process definitions, which are defined in separate profile supplements as workflow definitions. This demands sharing and mutual commitment to a process definition by all performers, since the workflow definition supplements do not provide machine-readable process models.

3.   Concept

3.1.   Data Use and Access process

A cross-consortia working group of the MI-I defined a process model that describes the relevant user tasks of a *data use and access* process, similar to practice-approved solutions (6). The support of a *data use and access* process involves several actors with different tasks and process artifacts bearing relevant information:

- *Data Consumer*: A person or organization, e.g. a scientist, that requests secondary-use data matching certain selection criteria by creating and providing a *data usage proposal* in the context of a data use project. A data consumer can be internal or external regarding a DIC.
- *Data Use Coordinator*: A trusted person or organization that represents and consolidates all participating DICs to the data consumer and is responsible for checking formal parameters of exchanged information and contracts.
- *Data Use and Access Committee*: Being part of each DIC, this group of persons is responsible for assessing *data usage proposals* and deciding whether the DIC will participate in the data use project and provide the requested data sets.

According to this process model, the data consumer creates and provides a description of the data use project based on a common *data usage proposal* form, explaining the relevant usage criteria and why the data sets are requested. Afterwards, the data use coordinator checks whether the *data usage proposal* fulfills all necessary formal aspects. If accepted, the data use and access committee of each participating DIC individually evaluates the *data usage proposal* based on individually relevant legal frameworks, ethical assumptions or other particular requirements and conditions. Each data use and access committee decides separately whether to accept the *data usage proposal*. If accepted by at least one data use and access committee, the approving participants agree to and sign a legal contract authorizing the provision of the requested secondary-use data sets only from the approving DICs. In case of rejection, the data consumer can resubmit a revised *data usage proposal* which must be evaluated again.

3.2.   SMITH Service Platform

The SMITH Service Platform realizes an overarching and centralized platform, offering the means to integrate and connect all DICs and thus offer services that apply to all DICs. The Data Sharing Services macro component realizes a central connection point for DICs on the one hand, and interfaces for use case or actor-specific applications on the other hand. The interface for connecting a DIC is based on international standards and is therefore scalable, assuming that the DIC implements the DIC reference architecture described by the SMITH consortium (3). Equally, the interface for use case-specific applications, the Marketplace Interface, is also based on international standards, allowing for the extension of use cases and functionality through interoperable applications in the future. The Marketplace Interface therefore defines interactions for the provision and retrieval of information relevant for process control. These interactions are based on the IHE profiles Cross-Enterprise Document Sharing (XDS) and Document Metadata Subscription (DSUB).

The SMITH Marketplace acts as a central access and contracting point for all users that are involved in the *data use and access* process. It supports all tasks of the the *data use and access* process by providing a tailored web interface.
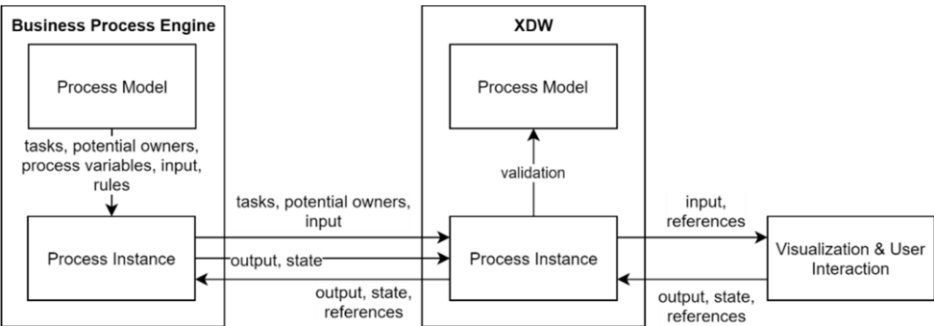
3.3.   Process control

figure 1: *process synchronization*

The distributed execution and visualization of user tasks requires information sharing about the process instance. This information must describe the process instance and each user task including status, potential owners, and input and output parameters. The central SMITH Service Platform thus manages the shared information about the process instance, where it is accessible for participating organizations and software tools.

The description of the process instance is based on the IHE profile Cross Enterprise Document Workflow (XDW) (9) and thus on an international standard, allowing for an easy mounting of different software tools. XDW enables users to coordinate and manage their tasks regarding a common process or workflow in a distributed environment. Since XDW is patient-centered, patient binding surrogates must be used in order to apply to the multi-patient scope of *data use and access*. Therefore, a virtual patient representing the use case (*data use and access*) is shared with all systems and used as the patient reference for the XDW document content and metadata.

The visualization and interaction is based on the process instance information provided as a XDW document. The SMITH Marketplace, incorporating the current user and authorization context, provides the visualization and interaction for user tasks. Therefore, all active XDW documents are requested and the active user tasks are displayed. Each user task contains input and output parameters as references to documents in the same environment. In addition to human-readable information, structured form documents are provided with a user task in order to automatically visualize a task-specific web form with instructions, gathering of required information and results, and provision of documentation. When a user task is finished, its structured form containing all gained information is provided as an output parameter of the user task and can be used for display purposes of subsequent user tasks.

Process control incorporates both a process model and process instance information. Since XDW only provides process instance information and misses the definition of process flow, rules and decisions, or shared process variables, additional process model and control management is required. A Business Process Engine (BPE) will therefore manage and provide the defined *data use and access* process model. If a process instance is started, its process flow is also managed by the BPE. The necessary synchronization of both the process instance managed by the BPE and the process instance based on XDW covers the alignment of user task definitions, potential owners, input parameters, output parameters, and state information. The SMITH Marketplace provides and manages the BPE and the necessary implementation of synchronization rules between both process instances.

4.   Implementation

Authorized users access the SMITH Marketplace web application. Depending on their role, users can perform different functionalities and use a work list showing all active tasks they have been assigned to. Scientists, assuming the role of the data consumer, manage their own data use projects and can start a new *data use and access* process instance, addressing only user selected DICs. Represented as a task of the scientist's own work list, the *data usage proposal* is compiled using a structured and formalized web form. Once this task is completed, the data use coordinator users receive a task on their own work list, asking them to approve the *data usage proposal* through a web form and additional documents. Afterwards, the data use and access committee users of each

participating DIC receive a task on their own work list, equally requesting them to assess the *data usage proposal* by selecting the evaluation result and providing an explanation in a structured and formalized web form. Then, contracting is carried out in a similar way.
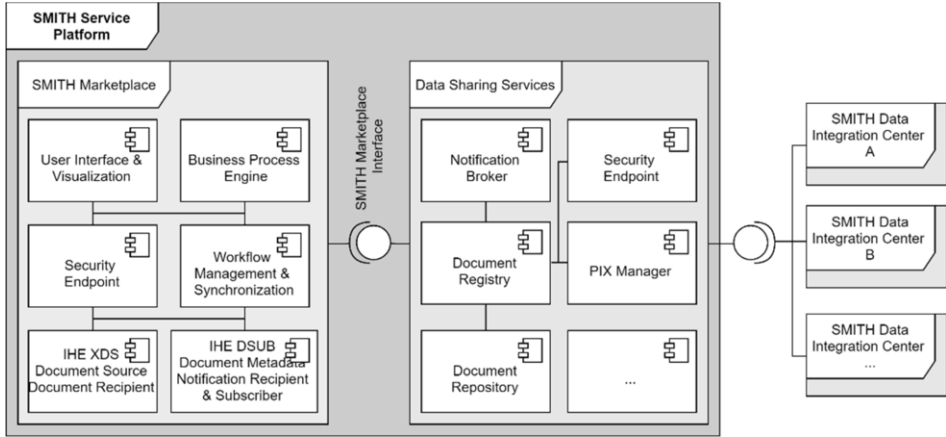

figure 2: *SMITH systems & Service Platform components*

Since XDW defines the content of a document describing a process instance rather than an information exchange workflow, the underlying infrastructure is based on XDS. Process instance information are shared as a XDW document based on a XDS infrastructure. The Data Sharing Services represent an appropriate Affinity Domain for the exchange of process information rather than personal or healthcare-related information. The SMITH Marketplace and all connected DICs must provide Document Source, Document Consumer and Document Metadata Subscriber and Document Metadata Notification Recipient actors, preferably, in order to participate in the information exchange. The BPE is part of the SMITH Marketplace and interacts closely with the Document Registry and Document Repository of the Data Sharing Services.

The use of XDW always accompanies the use of a workflow definition, which is not yet available for the *data use and access* process. For this reason, the SMITH Marketplace does not use a standardized workflow definition. However, it realizes the *data use and access* process based on the open source BPE Camunda. The combination of Camunda and XDW enables the mapping of generic processes. The Workflow Management & Synchronization component provides a synchronization between Camunda workflow instances and XDW documents. A listener ensures that the respective task is attached to the appropriate XDW document when it is executed. Likewise, results that are provided by other participants in the XDW document are returned to the BPE using DSUB subscriptions.

The SMITH Marketplace is realized with Vaadin and deployed on a Wildfly server. Both Camunda and the SMITH Marketplace are used as Docker containers.

5. Lessons learned (Discussion)

This work represents only a small contribution to the long-term goal of constantly optimizing patient care. Common and standardized processes and tools incorporating different data providers and scientists let expect an increase of accessibility and quality regarding secondary-use data, thus fertilizing medical research in the future.

With a novel approach for the implementation of a *data use and access* process based on international standards and the combination of an open and shareable process instance and central process control, the requirements of scalability and process interoperability have been met. The definition and inclusion of input and output parameters of a user task based on structured forms addresses the requirement of information interoperability, enabling automatic visualization and support for user interaction through the SMITH Marketplace and other external tools. Thus, a novel concept for cross-organizational collaboration in a common *data use and access* process could be provided.

The approach yields a new process definition for *data use and access* based on the XDW profile, and is therefore a candidate for standardization as a novel XDW-WD profile. Besides, the synchronization of user task instances with process control and a machine-readable process model through a central BPE renders the formal provision of a blueprint workflow definition optional. Future improvements must address the patient-centered approach and the resulting patient surrogates regarding XDS and XDW profiles. Moreover, the structured forms used for user task visualization should be replaced with suitable standardized content structures, such as HL7 FHIR Questionnaire.

The establishment of new organizational entities covering responsibilities and personnel of data use and access committees and the development of the SMITH infrastructure including the integration of the SMITH Service Platform with all DICs are still ongoing tasks. First practical experiences are expected in the near future. However, co-design principles are used to identify and address the requirements of users such as data use and access committee members to be implemented with the SMITH Marketplace.

The upcoming overarching concepts of the MI-I aim at connecting the solutions of its four consortia through a central coordinating organizational body and infrastructure (Zentrale Antrags- und Registerstelle, ZARS). Thus, all university hospitals in Germany will participate in a common and integrated *data use and access* solution for scientists with a central point of interaction. The technical specifications of the ZARS infrastructure and the role and integration options of the SMITH Service Platform are yet to be developed. Currently, an integration of the existing and technically diverse results is planned, maintaining the open and standards-based interface and process control of the SMITH Service Platform.

6. Conclusion

The union of a Business Process Engine handling process definition and execution rules with the provision and sharing of process instance information based on a standardized content structure defined by the IHE XDW profile and an open infrastructure for connecting distributed software for visualization and user interaction based on IHE XDS allows for a scalable environment and dynamic support for the *data use and access* use case. The SMITH Marketplace represents a reference application for visualization and

user interaction and an example implementation for user task support based on process instance information. The solution facilitates the accessibility of multiple data providers in the context of a data use project and offers a central interaction point for a distributed *data use and access* process.

Acknowledgements

References

1.   Celi LA, Mark RG, Stone DJ, Montgomery RA. "Big Data" in the Intensive Care Unit. Closing the Data Loop. Am J Respir Crit Care Med. 2013;(187(11)):1157–60. doi:10.1164/rccm.201212-2311ED

2.   Semler SC, Wissing F, Heyder R. German Medical Informatics Initiative: A National Approach to Integrating Health Data from Patient Care and Medical Research. Methods of Information in Medicine. 2018;(57). doi:10.3414/ME18-03-0003

3.   Winter A, Stäubert S, Ammon D, Aiche S, Beyan O, Bischoff V, Daumke P, Decker S, Funkat G, Gewehr JE, de Greiff A, Haferkamp S, Hahn U, Henkel A, Kirsten T, Klöss T, Lippert J, Löbe M, Lowitsch V, Maassen O, Maschmann J, Meister S, Mikolajczyk R, Nüchter M, Pletz MW, Rahm E, Riedel M, Saleh K, Schuppert A, Smers S, Stollenwerk A, Uhlig S, Wendt T, Zenker S, Fleig W, Marx G, Scherag A, Löffler M. Smart Medical Information Technology for Healthcare (SMITH): Data Integration based on Interoperability Standards. Methods of Information in Medicine. 2018;(57). doi:10.3414/ME18-02-0004

4.   Smart Medical Information Technology for Healthcare [Internet] [cited 2020 Mar 25]. Available from: https://www.smith.care/konsortium/

5.   Rea S, Pathak J, Savova G, Oniki TA, Westberg L, Beebe CE, Tao C, Parker CG, Haug PJ, Huff SM, Chute CG. Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: The SHARPn project. Journal of Biomedical Informatics. 2012;(Volume 45, Issue 4):763–71. doi:10.1016/j.jbi.2012.01.009

6.   Kirsten T, Wagner J, Kiel A, Rühle M, Löffler M. Selecting, Packaging, and Granting Access for Sharing Study Data: Experiences and Recent Software Developments in the LIFE Study. INFORMATIK 2017. 2017;2017.

7.   The Consultancy Committee for Space Data Systems. REFERENCE MODEL FOR AN OPEN ARCHIVAL INFORMATION SYSTEM (OAIS): RECOMMENDED PRACTICE CCSDS 650.0-M-2 [Internet] [cited 2020 Mar 30]. Available from: https://public.ccsds.org/pubs/650x0m2.pdf

8.   Vormstein P, Schneider T, Götze K, Brucker D, Nabinger R, Kuttruf J, Brandts C, Storf H. ProSkive: An approach for streamlining application procedures for requests for bio-material and clinical data. In: Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie, editor. 62. Jahrestagung der Deutschen Gesellschaft für Medizinische Informatik,

Biometrie und Epidemiologie e.V. (GMDS), Oldenburg, 17.-21.09.2017. Düsseldorf: German Medical Science GMS Publishing House; 2017.

9.   Integrating the Healthcare Enterprise. IHE IT Infrastructure Technical Framework: Cross-Transaction Specifications and Content Specifications [Internet]. 2019 [cited 2020 Mar 30]. Available from: https://www.ihe.net/uploadedFiles/Documents/ITI/IHE_ITI_TF_Vol3.pdf#name ddest=5_4_XDW_Workflow_Content_Module