Interoperability Is a Process – The Data Sharing Framework

Hauke HUND[a,1], Reto WETTSTEIN[b], Maximilian KURSCHEIDT[a], Simon T SCHWEIZER[a], Christoph ZILSKE[a] and Christian FEGELER[a]
[a] *GECKO Institute, Heilbronn University of Applied Sciences, Heilbronn, Germany*
[b] *Institute for Medical Informatics, Heidelberg University Hospital, Heidelberg, Germany*

Abstract. Common syntax and data semantics are core components of healthcare interoperability standards. However, interoperable data exchange processes are also needed to enable the integration of existing systems between organizations. While solutions for healthcare delivery processes are available and have been widely adopted, support for processes targeting bio-medical research is limited. Our *Data Sharing Framework* creates a platform to implement research processes like cohort size estimation, reviews and approvals of research proposals, consent checks, record linkage, pseudonymization and data sharing across organizations. The described framework implements a distributed business process engine for executing BPMN 2.0 processes with synchronization and data exchange using FHIR R4 resources. Our reference implementation has been rolled out to 38 organizations across three research consortia in Germany and is available as open source under the Apache 2.0 license.

Keywords. Interoperability, Process, Framework, Open Source, FHIR, BPMN

1. Introduction

Interoperability between information systems has commonly been understood as a problem that needs to be solved using a common syntax and data semantics. While widely adopted standards like HL7 v2, HL7 FHIR and openEHR [1-4] mostly aim to solve interoperability for health care delivery processes, usually focusing on a single subject, a standard for supporting bio-medical research processes across organizational boundaries has not been established. These processes include feasibility studies (cross-organizational cohort size estimations), reviews and approvals of research proposals, consent checks, record linkage and pseudonymization, as well as the transfer of data and biomaterial between organizations and at times the return of research results or incidental findings.

With this paper, we will report on recent developments and discuss lessons learned from nationwide deployments in Germany of our open source Data Sharing Framework (DSF). The DSF was first released in April 2020 as an open source reference implementation for executing arbitrary data exchange processes across organizations using HL7 FHIR resources and APIs to communicate information and the Business Process Model and Notation (BPMN 2.0) standard to document and implement executable processes [5].

[1] Corresponding Author: Hauke Hund, GECKO Institute, Heilbronn University of Applied Sciences, Max-Planck-Straße 39, 74081 Heilbronn, Germany; E-mail: hauke.hund@hs-heilbronn.de.

2. Methods

Requirements for the DSF were defined by conducting expert interviews with representatives from academia and industry as well as comparing the network and EHR architectures of member organizations of the HiGHmed research consortium [6]. In addition, published health-IT standards were evaluated and compared against our requirements.

The DSF was implemented using Java 11 with parsing and validation of FHIR resources performed using the HAPI FHIR library[2] and a business process engine based on the open source Camunda Engine version 7[3].

To manage the rollout of the DSF at research organizations across Germany and to enable authentication and communication between partners, an onboarding procedure supported by regularly held hackathon events was established.

3. Results

A number of requirements and limitations were identified, including no external access to EHR systems and pseudonymized or anonymized export of data with cross-organizational record linkage needed in order to comply with the European General Data Protection Regulation (GDPR) as well as support for integration into heterogeneous network- and system architectures of research organizations and healthcare providers.

The DSF was implemented using a FHIR server accessible by external communication partners, acting as a “mail box”; and a private business process engine (BPE) to integrate and access local and remote systems while executing BPMN 2.0 processes.

The DSF is implemented as open source software under the Apache 2.0 license and is available on GitHub[4]. Figure 1 shows how the DSF components communicate with each other for an example scenario with three independent organizations.

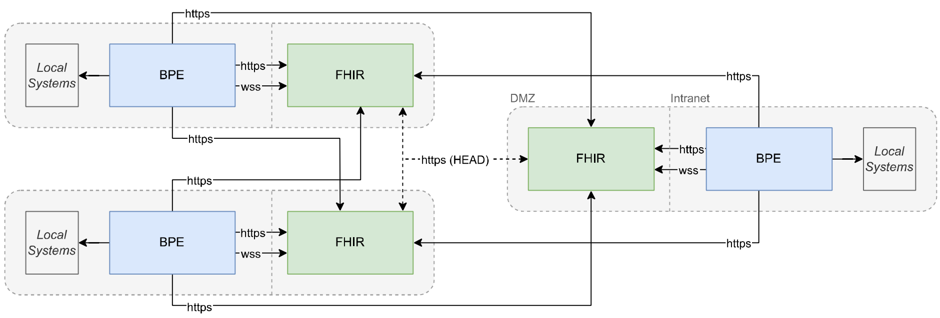

Figure 1. DSF architecture for three example organizations.

All websocket and HTTP connections between the DSF components (locally and across organizations) are encrypted using the TLS 1.3 protocol with mutual authentication using X.509 certificates. To identify known communication partners, thumbprints of the client certificates are stored as *extensions* in *Organization* resources in the DSF FHIR server, forming an allow-list.

---

[2] https://github.com/hapifhir/hapi-fhir (accessed 2023 March 8).

[3] https://github.com/camunda/camunda-bpm-platform (accessed 2023 March 8).

[4] https://github.com/highmed/highmed-dsf (accessed 2023 March 8).

3.1. FHIR Server

A subset of FHIR R4 resources was identified as necessary to create the DSF. To implement a list of known communication partners and their roles in research consortia like data providers, trusted third parties and others; *Organization*, *Endpoint* and *OrganizationAffiliation* resources are used to form an allow-list. This list allows Organizations to deploy multiple DSF instances to separate roles in different research consortia.

ActivityDefinition resources were chosen to document executable processes and extended with information to configure which processes can be executed by what organization or organization role. To start or continue the execution of BPMN processes, FHIR *Task* resources are used. *Task* resources specify the process and version to execute or continue as well as the requesting- and executing organization. If a *Task* resource is used to continue the execution of an existing process instance, additional identifiers (i.e. *business-key, correlation-key)* are specified. *Task input* and *output* parameters are used to communicate information into and out of process executions.

If a dataset needs to be transferred between organizations, a data extracting process can store *Binary*, *Bundle* or *MeasureReport* resources in the local DSF FHIR server before starting a download process at an external organization using a *Task* resource with an absolute literal reference to the dataset as an *input* parameter.

The DSF FHIR server supports *Questionnaire* and *QuestionnaireResponse* resources to support *user tasks* within BPMN processes. *CodeSystem*, *NamingSystem*, *StructureDefinition* and *ValueSet* are implemented to support the validation of FHIR resources. Other resources have been added to facilitate data sharing processes e.g. *DocumentReference*, *Group*, *Library*, *Measure* and *ResearchStudy*. The *Subscription* resource is used to implement the FHIR subscription mechanism with websockets.

3.2. Business Process Engine

Within the DSF, processes are modeled as executable BPMN 2.0 files and packaged into Jar file process plugins alongside executable code and FHIR metadata resources. The graphical models facilitate high-level discussions with stakeholders and are well suited to document and execute distributed processes since communication relationships between organizations can be clearly represented using BPMN *message flow*. Sending a message between processes using a *message send event* or *message send task* is implemented by creating a *Task* resource at a DSF FHIR server. When a new *Task* resource is available with the status *"requested"*, a *Subscription* is triggered, informing the BPE of a request to start or continue a process instance via an open websocket connection.

When a process reaches a *user task,* a corresponding *QuestionnaireResponse* resource with status *"in-progress"* is created by the BPE in the DSF FHIR server. After the user has answered the questionnaire via the provided user interface or an external tool, thus changing the status of the *QuestionnaireResponse* to *"completed",* a second *Subscription* is used to inform the BPE that automatic process execution can continue.

3.3. Process Plugins

A number of process plugins have been implemented to validate the reference implementation, including sharing of openEHR data with Bloom-Filter based record-linkage at a central organization [7] and feasibility studies using secure multi-party computation without a central organization [8]. Other processes to share routine FHIR data within the

German M*edical Informatics Initiative* (MII)[5] and COVID-19 research data within the German *Network University Medicine* (NUM) [9] are being used in production across university hospitals in Germany.

3.4. On-Boarding and Deployment

After acquiring the necessary X.509 certificates and installing the DSF, a research organization can request inclusion in the allow-list of a consortium by transferring the client certificate thumbprint, the URL of the FHIR server and other metadata describing the organization to a managing institution. To validate the setup, the new organization is first added to the allow-list with disabled roles. After testing communications with a simple ping-pong process, the new organization is enabled and can fully participate.

Across the research consortia HiGHmed, MII and NUM 38 organizations currently have at least one installation of the DSF (38 production, 24 for testing).

4. Discussion

Developing and deploying data sharing applications for bio-medical research based on routine data is not only a problem to solve but also a process to manage. Changing data handling requirements, evolving data models and organizations taking on different roles over time necessitate a data sharing environment that can manage change. In addition, the existing heterogeneous network- and system architectures at German university hospitals and research institutions require a platform that can be adapted.

The described *Data Sharing Framework* allows for the realization of a distributed process execution environment that can scale beyond a small number of organizations. With configurable organization roles, versioned FHIR resources and executable BPMN processes, we can manage changes to data models, data exchange processes and organizational structures without the need to modify the underlying architecture or platform implementation. Separating the process engine from the FHIR server further allows us to deploy and adapt to local requirements. Configurable process plugins can be implemented with different local environments in mind, or completely separate interoperable plugin implementations based on the same FHIR profiles can be created.

With state-of-the-art transport encryption, authentication and by minimizing central data storage using a distributed architecture, the framework enables EU GDPR compliant operation. Since compliance with regulations and laws is very much context and content-specific, distributed processes can be individually tailored to different legal contexts, e.g. informed consent, treatment contract or legal obligation of the organization.

The DSF was developed as a machine-to-machine middleware serving as a bridge between organizations. Using mutual authentication with X.509 certificates allows us to eliminate the need for identity providers required by protocols such as OpenID Connect or SAML 2.0. Our implementation requires organizations to deploy a FHIR server accessible from public networks and to run a BPE internally. The current rollout has shown that this is feasible for larger organizations like university hospitals but may not be easily deployed at smaller institutions. A Software as a Service (SaaS) delivery model could be used to work around this limitation. Additionally, tooling to support better monitoring of the local and distributed status of the DSF has been identified as missing.

[5] https://www.medizininformatik-initiative.de/en/start (accessed 2023 March 8).

Improvements to the DSF and the creation of a broader development community have been funded by the German government for the upcoming years, making us look forward to reaching out to the European and international health-IT community.

5. Conclusions

Our open source *Data Sharing Framework* implements a distributed business process engine for executing arbitrary bio-medical research and healthcare delivery processes modeled using BPMN 2.0 with information exchange using HL7 FHIR R4. The described reference implementation is available under the Apache 2.0 license and has been rolled out to 38 organizations for three different research consortia in Germany.

Using BPMN to model executable processes allows us to capture requirements and discuss complex data exchange protocols with non-technical stakeholders. In conjunction with FHIR resources and APIs, an interoperable and re-implementable data exchange platform can be created. Thus, enabling syntactic, semantic and process interoperable data exchange.

Acknowledgements

The project is funded by the German Federal Ministry of Education and Research (BMBF, grant ids: 01ZZ1802E and 01ZZ1802A). The authors would like to thank all contributors of the open source implementation, especially Florian Seidel (Charité, Berlin, Germany) for testing and helping with the hackathons.

References

[1] Stanimirović D, Beštek M. Special Topic Interoperability and EHR: Combining openEHR, SNOMED, IHE, and Continua as approaches to interoperability on national eHealth. *Appl Clin Inform. 2017;08(03):810-25*. doi: 10.4338/ACI-2017-01-RA-0011
[2] Ljosland Bakke S. National governance of archetypes in Norway. Stud Health Technol Inform. 2015;216:1091. doi: 10.3233/978-1-61499-564-7-1091
[3] Bender D, Sartipi K. HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. Proceedings of the 26th IEEE International Symposium on Computer-Based Medical Systems. 2013;:326-31
[4] Duda SN, Kennedy N, Conway D, Cheng AC, Nguyen V, Zayas-Cabán T, Harris P. HL7 FHIR-based tools and initiatives to support clinical research: a scoping review. Journal of the American Medical Informatics Association. 2022;29(9):1642-53. doi: 10.1093/jamia/ocac105
[5] Hund H, Wettstein R, Heidt CM, Fegeler C. Executing Distributed Healthcare and Research Processes – The HiGHmed Data Sharing Framework. Stud Health Technol Inform. 2021;278:126-33. doi: 10.3233/SHTI210060
[6] Haarbrandt B, Schreiweis B, Rey S, Sax U, Scheithauer S, Rienhoff O, Knaup-Gregori P, et al. HiGHmed - An Open Platform Approach to Enhance Care and Research across Institutional Boundaries. Methods of Information in Medicine. 2018;57(S 01). doi: 10.3414/ME18-02-00 02
[7] Wettstein R, Hund H, Fegeler C, Heinze O. Data Sharing in Distributed Architectures – Concept and Implementation in HiGHmed. Stud Health Technol Inform. 2021;283:111-18. doi: 10.3233/SHTI210548
[8] Wettstein R, Kussel T, Hund H, Fegeler C, Dugas M, Hamacher K. Secure Multi-Party Computation Based Distributed Feasibility Queries – A HiGHmed Use Case. Stud Health Technol Inform. 2022;296:41-49. doi: 10.3233/SHTI220802
[9] Prokosch H-U, Bahls T, Bialke M, Eils J, et al. The COVID-19 Data Exchange Platform of the German University Medicine. Stud Health Technol Inform. 2022;294:674-78. doi: 10.3233/SHTI220554