

Executing Distributed Healthcare and Research Processes – The HiGHmed Data Sharing Framework

Hauke HUND^{a,1}, Reto WETTSTEIN^b, Christian M. HEIDT^a and Christian FEGELER^a

^a*GECKO Institute, Heilbronn University of Applied Sciences, Germany*

^b*Department Medical Information Systems, Heidelberg University Hospital, Germany*

Abstract. Several standards and frameworks have been described in existing literature and technical manuals that contribute to solving the interoperability problem. Their data models usually focus on clinical data and only support healthcare delivery processes. Research processes including cross organizational cohort size estimation, approvals and reviews of research proposals, consent checks, record linkage and pseudonymization need to be supported within the HiGHmed medical informatics consortium. The open source HiGHmed Data Sharing Framework implements a distributed business process engine for executing arbitrary biomedical research and healthcare processes modeled and executed using BPMN 2.0 while exchanging information using FHIR R4 resources. The proposed reference implementation is currently being rolled out to eight university hospitals in Germany as well as a trusted third party and available open source under the Apache 2.0 license.

Keywords. Data Sharing, Interoperability, Biomedical Research, Process, Framework, Open Source, FHIR, BPMN

1. Introduction

1.1. Background

Several standards and frameworks have been described that contribute to solving the interoperability problem on a syntactic and semantic level, focusing on a shared understanding of context and meaning of terms and other data points. While standards like HL7 v2, HL7 FHIR and openEHR are being adopted in the healthcare industry [1-4], their data models and use cases usually focus on a single subject and typically only support processes related to healthcare delivery.

Within the German Medical Informatics Initiative and especially within the HiGHmed consortium sharing routine clinical data for biomedical research at university hospitals is one of the primary goals [5,6]. For this, not only syntactic and semantic challenges must be solved but also issues regarding the harmonization of data exchange processes. These include feasibility studies (cross-organizational cohort size estimations), approvals and reviews of research proposals, consent checks, record linkage and

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

pseudonymization, as well as the transfer of data and biomaterial between organizations and at times personalized return of research results or incidental findings.

The aforementioned semantic interoperability frameworks form an important base for data exchange in the research context but on their own cannot solve the interoperability problem; an important challenge being the heterogeneity of process execution in multi organizational networks. With this paper, we propose a framework for implementing standardized, shared processes using Business Process Model and Notation (BPMN 2.0) as well as communication between actors using HL7 Fast Healthcare Interoperability Resources (FHIR R4).

1.2. Requirements

Based on the general architecture of independent Medical Data Integration Centers (MeDICs) at eight university hospitals located across Germany and a Trusted Third Party (TTP) at Heilbronn University of Applied Sciences, a distributed systems and communication architecture was chosen for data exchange and distributed data analysis within the HiGHmed consortium.

A number of technical requirements were identified by expert interviews within the consortium, starting from the general need to exchange medical data and perform distributed analyses: Communication between organizations should be peer-to-peer and based on standardized HTTP protocols and data models. Processes need to be modeled and executed based on a standardized process notation. Furthermore, the resulting architecture should scale beyond the current members of the HiGHmed consortium and should also be usable for other healthcare and research projects unrelated to the consortium. Additionally end-users, like researchers and data scientists, should only be required to communicate with systems and services at their local organization, removing the need for distributed identity management.

As a non-technical requirement, compliance with the EU General Data Protection Regulation (GDPR) as well as national and state regulations was identified. The solution should therefore employ security by design and minimal use of protected personal data.

In order to validate and test the resulting data exchange specification, a reference implementation needs to be provided and readily available to MeDICs and TTPs.

2. State of the art

Within the IHE community² two profiles have been proposed that feature functionalities related to the requirements described above: “The Cross-Enterprise Document Workflow (XDW) Profile enables participants in a multiorganizational environment to manage and track the tasks related to patient-centric workflows [...]” [7]. Also “Retrieve Process for Execution (RPE) [...] was proposed, defining] a profile for collaborative workflow or collaborative process management [...]”. As such, “RPE defines the transport mechanism for process definitions, but does not provide the definitions per se” [8]. Both profiles are currently in trial implementation state, meaning commercial implementations are not readily available.

² IHE: Integrating the Healthcare Enterprise, a non-profit organization coordinating the development of healthcare IT standards.

While XDW is used in patient centric workflows, RPE can be used to coordinate workflows across healthcare and research systems. Both employ use-case specific content profiles. For RPE the Clinical Research Process Content (CRPC) has been proposed to “[...] automate the sharing of information among systems during the clinical research process” [9].

To the best of our knowledge, open and standards-based systems to execute and coordinate processes in a healthcare research setting across organizations that cover all our requirements are not readily available.

The proposed reference implementation uses the current stable release of HL7 FHIR – Version 4.0.1³ for exchange of information between organizations and uses executable processes modeled using BPMN in Version 2.0⁴.

The HL7 FHIR standard defines information resources within the healthcare and biomedical research domain as well as HTTP REST Interfaces to access these resource instances. Additionally, FHIR specifies controlled vocabularies and the handling of existing external classifications and terminologies. Furthermore, the standard defines mechanisms to validate resources against use-case specific profiles, enabling syntactic and semantic interoperability of resources across organizations.

The BPMN 2.0 standard maintained by the Object Management Group (OMG) is a graphical notation stored as XML to model business processes. BPMN enables modelling and execution of processes across organizational boundaries. Different organizations are typically represented as so-called pools and information exchange between organizations can be modelled with message-events depicted as envelopes. As such, modelling and executing BPMN enables process interoperability between organizations.

3. Concept

The HiGHmed Data Sharing Framework (DSF) as published in Version 0.2.1 on GitHub⁵ differentiates two kinds of organizations: Medical Data Integration Centers (MeDICs) and Trusted Third Parties (TTPs). Both organizations can use the same reference applications and general network setup.

The DSF contains two server applications: A FHIR webservice endpoint, implementing a necessary subset of the FHIR R4 specification, and a Business Process Engine (BPE) to execute BPMN 2.0 processes and access local services and resources.

While the FHIR endpoint is deployed in a demilitarized network zone (DMZ) and accessible by authenticated and authorized external organizations, the BPE is deployed inside the MeDIC/TTP internal network and not accessible from the outside or the local FHIR endpoint. In order to communicate information from external organizations to the internal BPE, a FHIR websocket subscription is used. The secure websocket connection (WSS) is initialized by the BPE and represented as an arrow in Figure 1.

In addition to the websocket connection, the BPE communicates via HTTPS (TLS 1.3) with local and remote FHIR endpoint webservices. End-users like researchers do not interact directly with DSF services, but rather only the DSF applications act as client and server within the DSF network and communicate with each other. Organizations are considered users by the distributed network of DSF installations. Authentication of

³ <https://hl7.org/fhir/R4>

⁴ <https://www.omg.org/spec/BPMN/2.0>

⁵ <https://github.com/highmed/highmed-dsf>

organizations is implemented using mutual TLS authentication with X.509 client and server certificates. Currently certificates signed by the DFN-PKI⁶ are used. Access to the FHIR endpoint is granted if the client certificate is known and part of an allow list.

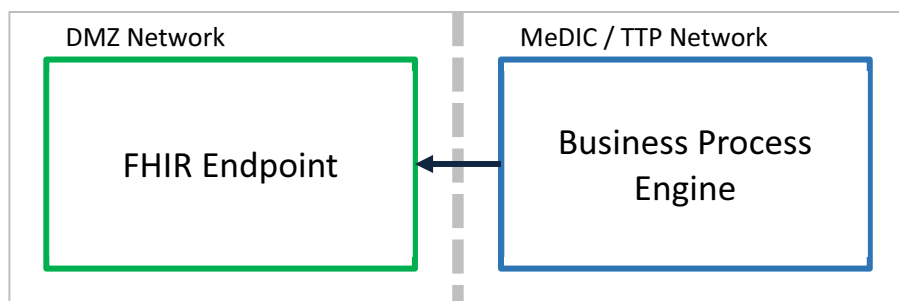


Figure 1. MeDIC / TTP network setup with FHIR endpoint and business process engine.

In order to implement the feasibility study, record-linkage and pseudonymization use-cases, a subset of FHIR R4 resources was identified and implemented within the FHIR endpoint application: ActivityDefinition, Binary, Bundle, CodeSystem, Endpoint, Group, NamingSystem, Organization, Practitioner, PractitionerRole, ResearchStudy, StructureDefinition, Subscription, Task and ValueSet are currently supported.

A set of specialized access rules is implemented to allow access to FHIR resources. These access rules either allow access based on a local or remote organization, or based on the state of the requested resource. While generally Task resources can be read, created, updated or deleted by external organizations, only read access is granted for all other resources.

Organization and Endpoint resources are used to implement a list of known organizations and associated FHIR endpoints, including URLs of the FHIR servers. To identify known client certificates, a certificate thumbprint extension on the Organization resource is used to implement the allow list.

Within the DSF, processes are modeled as executable BPMN 2.0 files. These graphical models support communication with stakeholders and only require minimal modifications to get processes into production and maintained. BPMN models are well suited to document distributed processes since communication relationships between organizations can be clearly represented using message events.

Receiving a message event within BPMN results in either the start of a new process instance or the continuation of an existing one. These message events translate well to FHIR Task resources. When one organization wants to communicate with another, the initiating organization creates a Task resource with status *requested* at the endpoint of the recipient. The BPE is informed via the Subscription that a new Task resource was created which leads to a process being started or continued, also resulting in a state change of the Task resource to *in-progress*. When a process finishes or fails, the Task resource is updated to represent the *completed* or *failed* state.

Input fields of the Task resources are used to communicate parameters needed to start or continue a process. These parameters can either be simple values or references to

⁶ DFN-PKI: Certificate authority of the German Research Network (Deutsches Forschungsnetz, DFN). <https://www.pki.dfn.de/ueberblick-dfn-pki/>

existing complex resources stored at a FHIR endpoint within the network of organizations. Thus, Task resources can either relate to patient centric healthcare processes or be used to execute research processes with a patient cohort context or no context at all, for instance while testing communication links.

In order to implement feasibility studies and data sharing between organizations including record-linkage and pseudonymization, medical data sets and identifying patient information need to be accessed. The MeDIC BPE communicates during the execution of these processes with medical data-repositories like openEHR and a master-patient index, while the TTP BPE accesses record-linkage and pseudonymization services to execute TTP-specific process steps.

4. Implementation

The HiGHmed DSF was created as a Java reference implementation open source under the Apache 2.0 license. Releases are published as Docker containers via the project’s GitHub repository⁵. The FHIR endpoint application server as well as the BPE application server are provided as Docker images. Additional Docker images are provided for Apache httpd reverse proxies used to terminate incoming TLS (HTTPS) connections. For persistent storage of FHIR resources and BPMN process states, standard PostgreSQL database Docker images are used.

Processes to update resources, test communications between organizations and perform feasibility studies are implemented. Tools and algorithms to execute record-linkage and generate pseudonyms are also part of the current 0.2.1 release of the DSF.

Examples and configurations are provided to start a single MeDIC or TTP instance via Docker containers (docker-compose) as well as three MeDICs and a TTP via four virtual machines each running Docker containers to test processes involving multiple organizations.

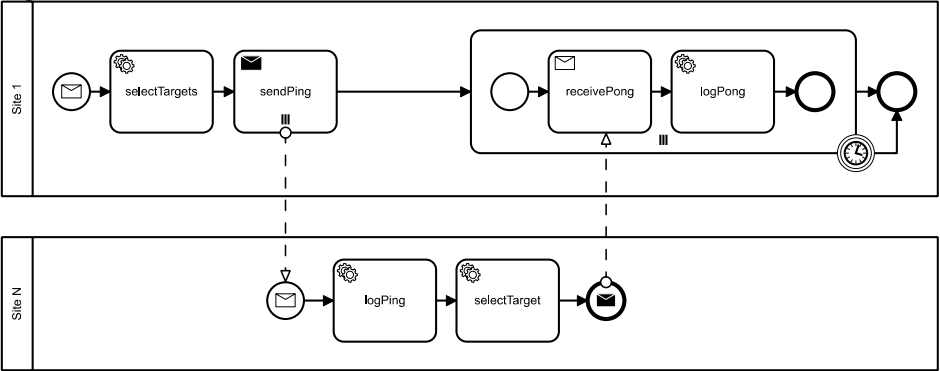


Figure 2. Implemented ping (top) and pong (bottom) processes to test communication between organizations.

Figure 2 depicts the ping and pong processes implemented in the DSF for testing basic communications between organizations. “Site 1”, by creating a Task resource in the local FHIR endpoint, starts the execution of the “ping” process within the local BPE. While executing, the BPE creates Task resources at all remote FHIR endpoints to start the pong process in their BPEs. As part of the remote execution of the pong process and

while “Site 1” is waiting, every remote site creates a Task resource back at “Site 1” to allow the originating ping process to continue and finish.

In order to relate all outgoing and incoming Task resource to the originating single ping process instance, a business-key is used and transferred to the remote site and back. Furthermore, a unique correlation-key is used to track every bilateral communication relationship.

An example Task resource used to start the pong process at “Site N” (referenced as Organization_2) is shown in Figure 3. The illustrated resource is an abbreviated version missing the properties *id*, *meta* with *versionId*, *profile* and *lastUpdate*, as well as *intent* and *authoredOn*.

```
<Task xmlns="http://hl7.org/fhir">
  <instantiatesUri value="http://highmed.org/bpe/Process/pong/0.2.0"/>
  <status value="requested"/>
  <requester><type value="Organization"/><identifier>
    <system value="http://highmed.org/fhir/NamingSystem/organization-identifier"/>
    <value value="Organization_1"/></identifier></requester>
  <restriction><recipient><type value="Organization"/><identifier>
    <system value="http://highmed.org/fhir/NamingSystem/organization-identifier"/>
    <value value="Organization_2"/></identifier></recipient></restriction>
  <input>
    <type><coding><system value="http://highmed.org/fhir/CodeSystem/bpmn-message"/>
    <code value="message-name"/></coding></type>
    <valueString value="pingMessage"/>
  </input>
  <input>
    <type><coding><system value="http://highmed.org/fhir/CodeSystem/bpmn-message"/>
    <code value="business-key"/></coding></type>
    <valueString value="bae3202f-cc42-4720-828d-dcc0d74f0db6"/>
  </input>
  <input>
    <type><coding><system value="http://highmed.org/fhir/CodeSystem/bpmn-message"/>
    <code value="correlation-key"/></coding></type>
    <valueString value="e194ea45-12a3-4e6a-8aae-a5b25fabce96"/>
  </input>
</Task>
```

Figure 3. Task resource created by Organization_1 and used to start the pong process at Organization_2

All university hospital organizations within the HiGHmed consortium have running installations of the DSF or are in the process of deploying the reference implementation. The implemented processes have successfully been tested between three organizations.

5. Lessons learned (Discussion)

Open and standards-based systems to execute and coordinate processes in a healthcare delivery and research setting across organizations meeting our requirements are not readily available.

We present a framework to implement distributed processes modeled with BPMN 2.0 using FHIR R4 resources to exchange information between organizations. While BPMN message events are used to model communication between organizations, FHIR Task resources are used to start, continue and track the execution of processes. A combination of both standards enables us to implement a syntactic, semantic and process interoperable distributed system.

The proposed architecture and DSF reference implementation will enable the HiGHmed consortium to realize a distributed process execution architecture without the

need for central coordinating organizations or systems. Processes to test communication channels, update resources and execute distributed feasibility studies are implemented in the current release.

Simulated and real life integration tests performed with three MeDICs and a TTP indicate that our requirements are being met. A larger evaluation of the architecture and reference implementation will be performed based on the HiGHmed rollout plan.

The proposed implementation requires every organization to deploy a FHIR end-point accessible from public networks and to run a BPE internally. This setup might not be feasible for smaller organizations limited by network bandwidth or unable to deploy a separated public network. To work around this limitation a Software as a Service (SaaS) delivery model could be used in the future.

While testing the DSF within the HiGHmed consortium we found an incompatibility with TLS interception software. This incompatibility stems from TLSI software typically not being able to correctly intercept TLS 1.3 traffic that uses mutual authentication with client and server certificates [10]. To solve this issue, TLSI had to be deactivated for the DSF applications.

While organizations are authenticated across the network using client and server certificates, end-user accounts only need to be authenticated locally, making distributed user authentication unnecessary. Request for data usage may however still be associated to researchers using Practitioner and PractitionerRole resources.

The framework enables EU GDPR compliant operation by using state of the art transport encryption (TLS 1.3) and authentication (X.509 certificates) as well as minimizing central data storage by the use of our distributed architecture. Since compliance with regulations and laws is very much context and content specific, distributed processes can be individually tailored to different legal contexts, for example: informed consent, treatment contract or legal obligation by the organization.

Processes to execute distributed analysis other than feasibility studies, as well as data sharing including approvals and reviews of research proposals, consent checks, record linkage and pseudonymization have not been implemented using the DSF, but we expect to be able to implement complex processes necessary to coordinate distributed analysis in the near future. Modeling and executing processes with model notation BPMN 2.0 enables us to document executable processes. This form is also well suited to facilitate communication during the design of these processes between different stakeholders, such as physicians, researchers and medical informatics professionals.

Using the DSF also for healthcare, patient-centric use-case is currently being evaluated within the HiGHmed consortium. Examples being order-entry processes between hospitals and specialized genetics laboratories as well as processes for preparing virtual tumor board conferences.

Currently only one TTP organization is part of the HiGHmed DSF network, but multiple TTPs, being selected based on the project or request, are expected in the future and supported by the framework.

6. Conclusion

The open source HiGHmed Data Sharing Framework implements a distributed business process engine for executing arbitrary biomedical research and healthcare delivery processes modeled using BPMN 2.0 while exchanging information using FHIR R4 resources.

The described reference implementation is currently being rolled out to eight university hospitals in Germany as well as a trusted third party and available open source under the Apache 2.0 license⁵.

The proposed framework should enable us to better integrate heterogeneous data sources, processes and healthcare IT architectures across organizational boundaries, getting us closer to syntactic, semantic and process interoperability.

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 committers that contributed to the open source implementation, especially Florian Seidel (Charité, Berlin) for also helping to test the current release.

Conflict of Interest

The authors state that they have no conflict of interests.

References

- [1] M. Beštek, D. Stanimirović, Special Topic Interoperability and EHR: Combining openEHR, SNOMED, IHE, and Continua as approaches to interoperability on national eHealth. *Appl Clin Inform.* **8** (2017) 810–825. doi:10.4338/ACI-2017-01-RA-0011
- [2] S.B. Ljosland, National governance of archetypes in Norway. *Stud. Health Technol. Inform.* **216** (2015) 1091-1091. doi:10.3233/978-1-61499-564-7-1091
- [3] D. Bender, K. Sartipi, HL7 FHIR: An Agile and RESTful approach to healthcare information exchange. *In Proceedings of the 26th IEEE international symposium on computer-based medical systems* (2013) 326-331.
- [4] M. Lehne, S. Luijten, P. Vom Felde Genannt Imbusch, S. Thun. The Use of FHIR in Digital Health – A Review of the Scientific Literature. *Stud. Health Technol. Inform.* **267** (2019) 52-58. doi:10.3233/SHTI190805
- [5] P. Knaup, T.M. Deserno, H.-U. Prokosch, and U. Sax, Implementation of a National Framework to Promote Health Data Sharing, *Yearb Med Inform.* **27.01** (2018) 302-304.
- [6] B. Haarbrandt et al., HiGHmed - An Open Platform Approach to Enhance Care and Research across Institutional Boundaries, *Methods of Information in Medicine* **57** (2018), e66–e81.
- [7] IHE ITI Technical Committee. Cross-Enterprise Document Workflow (XDW), Revision 2.3, (2014) 5. https://www.ihe.net/uploadedFiles/Documents/ITI/IHE_ITI_Suppl_XDW_Rev2.4_TI_2014-10-13.pdf (accessed July 14, 2020)
- [8] IHE QRPH Technical Committee. Retrieve Process for Execution (RPE), Revision 4.2, (2018) 7. https://www.ihe.net/uploadedFiles/Documents/QRPH/IHE_QRPH_Suppl_RPE.pdf (accessed July 14, 2020)
- [9] IHE QRPH Technical Committee. Clinical Research Process Content (CRPC), Revision 2.1 (2018) 5. https://www.ihe.net/uploadedFiles/Documents/QRPH/IHE_QRPH_Suppl_CRPC.pdf (access March 31, 2020)
- [10] National Security Agency (NSA). Managing Risk From Transport Layer Security Inspection (2019). https://www.waterisac.org/system/files/articles/MANAGING+RISK+FROM+TLS+INSPECTION_20191106.PDF (accessed July 14, 2020)