A Formal Model for the FAIR4Health Information Architecture

Mona PERBIX [a,1], Matthias LÖBE [a], Sebastian STÄUBERT [a], A. Anil SINACI [b],
Mert GENCTURK [b], Miriam QUINTERO [c], Alicia MARTINEZ-GARCIA [d],
Celia ALVAREZ-ROMERO [d], Carlos L. PARRA-CALDERON [d] and Alfred WINTER [a]

[a] *Institute of Medical Informatics, Statistics and Epidemiology (IMISE),*
Leipzig University, Germany
[b] *SRDC Software Research Development and Consultancy Corporation,*
Ankara, Turkey
[c] *Atos Research and Innovation (ARI), Madrid, Spain*
[d] *Institute of Biomedicine of Seville (IBIS), Virgen del Rocío University Hospital,*
University of Seville, Seville, Spain

Abstract. In the EU project FAIR4Health, a ETL pipeline for the FAIRification of structured health data as well as an agent-based, distributed query platform for the analysis of research hypotheses and the training of machine learning models were developed. The system has been successfully tested in two clinical use cases with patient data from five university hospitals. Currently, the solution is also being considered for use in other hospitals. However, configuring the system and deploying it in the local IT architecture is non-trivial and meets with understandable concerns about security. This paper presents a model for describing the information architecture based on a formal approach, the 3LGM metamodel. The model was evaluated by the developers. As a result, the clear separation of tasks and the software components that implement them as well as the rich description of interactions via interfaces were positively emphasized.

Keywords. Health Information Systems, Organizational Models, Systems Analysis

1. Introduction

Medical research nowadays is often an effort involving partners from different sites working together using distributed information systems. Because of the innovative nature, researchers often do not use standard solutions, but develop rather novel and feature-rich, yet little-tested information systems. In distributed scenarios, the documentation of the systems or their configuration is of great importance due to the high demands on the security of sensitive patient data. The FAIR principles were published in 2016 to make data more findable, accessible, interoperable, and reusable [1]. The FAIR4Health project deals with the application of the FAIR principles in health research [2]. In particular, the objective is to improve the share and reuse of research data. Within the project a complex information system was developed, which is composed of different components that have to be orchestrated together. The complete

---

[1] Corresponding Author, Mona Perbix, IMISE, Leipzig University, Härtelstraße 16-18, 04107 Leipzig, Germany; E-mail: mp32poho@studserv.uni-leipzig.de.

source code is available distributed in twelve repositories[2]. Currently, the FAIR4Health project's extensive documentation consists of descriptive documents and illustrative figures. This enables a quick start and overview of the individual components. However, the relationships and dependencies between the components are not detailed enough and are not trivial executed in individual cases, which might make it difficult to keep track of the data flows and the scope of communication between components.

2.   Methods

For reasons mentioned above, it would be advantageous to express the information architecture in a proven modeling methodology. In the context of this work, the architecture is modeled according to the Three-Layer Graph-based Meta-model (3LGM^2) [3]. The model is used to document and explain the entire information architecture. It is intended to help support the integration of the FAIR4Health infrastructure in new sites and to facilitate security concerns related to access to patient data. An information architecture in 3LGM^2 is divided into three layers [4]. The domain layer ("task layer") includes enterprise functions and entity types which the information system to be modeled provides. The second layer is the logical tool layer ("software layer") which describes application components supporting enterprise functions. The physical tool layer ("hardware layer") focuses on physical data processing systems. In addition, the 3LGM^2 model is based on the FAIR4Health project deliverables served as the fundamental basis for the creation of the model.

To assess the potential benefits, a structured evaluation of the model in terms of accuracy and usefulness was conducted using an anonymous survey. For the creation of the questionnaire the internet-based application REDCap was used [5]. The survey was tailored to the following audiences who work closely with the FAIR4Health information architecture:

- *Developers* who developed the system
- *System administrators* who want to install the system at their data site
- *Users* who use the system for research

3.   Results

3.1.   Formal model

Because the FAIR4Health information system is distributed, it can be divided into two parts. On the one hand, the *agent* is a set of tools that are installed locally at the site of the data holding organization to apply the FAIRification process [6]. On the other hand, the *platform* enables the user to work on top of the FAIRified datasets and execute so-called privacy-preserving distributed data mining (PPDDM) algorithms. It manages the distributed agents from a central position.

In the logical tool layer, the application components of the FAIR4Health information architecture and their communication with the help of services are presented, see Figure 1. The communication relationships are consistent directional arrows between calling

---

[2] https://github.com/orgs/fair4health/repositories

and providing interfaces that are part of an application system. Moreover the application components belonging to the agent and platform are separated by color and location. Each of the components is described and categorized in great detail in the 3LGM^2 tool [https://3lgm2.de/en/], relationships are typed and interface definitions are annotated with the protocols, standards, persistence methods and workflow elements used (not visible in the graphical export).
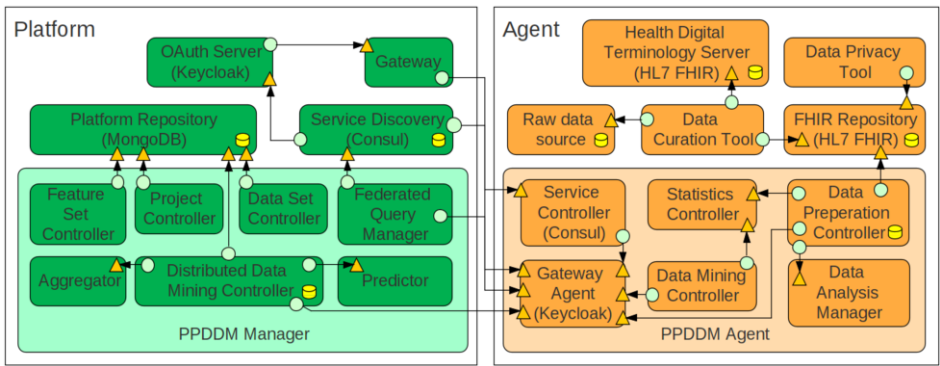
Figure 1. 3LGM^2 model (layer view) for the FAIR4Health information architecture. Application components belonging to the agent are colored orange and application components belonging to the platform are colored green.

With the help of the inter-layer relationships, application systems can be assigned to the individual enterprise functions of the FAIR4Health workflow they execute, see Figure 2.

Figure 2. Excerpt from the matrix view of the Inter-layer relationships for the FAIR4Health information architecture Application components are on the columns and functions are on the rows.

Because the FAIR4Health information architecture is inherently quite complex and 3LGM^2 adds another layer of complexity, the model was evaluated by project participants in terms of its usefulness.

3.2.   Evaluation of the formal model

Five participants from the FAIR4Health project team were asked to respond to the survey. Of these, three identified themselves as developers, one as a system administrator, and one primarily as a user. The usefulness of features offered by the 3LGM^2 model is shown in Table 1 in relation to the target groups. Useful rated functions differ between the target groups. The developers in particular find the separation by layers very useful for their work, since the 3LGM^2 model combines different architecture diagrams. According to the majority of participants, the model actually enables a deeper

understanding of the interaction of different components and the FAIR4Health information architecture in depth. In this context, a detailed representation of the model was regarded more significant by the respondents than an easy-to-grasp graphical illustration.

Table 1. Useful functions of the formal model used per target group

Administrator	User	Developer
Clear understanding of the relationships between the elements.	Clear understanding of elements through a metamodel.	Clear understanding of the relationships between the elements.
	Different views.	Helpful separation by layers. Better general overview.

4. Discussion and Outlook

The FAIR4Health information architecture was modeled with the 3LGM^2 meta model to provide a more formal description. The evaluation showed that the 3LGM^2 model added value in terms of consistent modeling elements and separation by layers. However due to the small number of participants, the results of the evaluation are not representative in the community. It is planned to integrate and test the FAIR4Health information system at the University hospital in Leipzig. The 3LGM^2 model of the information architecture serves here as a reference model for the integration. The created model and key images are available through the Leipzig Health Atlas research data management system (LHA id: 8C6DG2JAKM-3[3]).

Acknowledgements

References

[1] Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016; 3:160018. doi: 10.1038/sdata.2016.18.

[2] Alvarez-Romero C, Martínez-García A, Sinaci AA, Gencturk M, Méndez E, Hernández-Pérez T et al. FAIR4Health: Findable, Accessible, Interoperable and Reusable data to foster Health Research. Open Res Europe 2022; 2:34. doi: 10.12688/openreseurope.14349.1.

[3] Wendt T, Häber A, Brigl B, Winter A. Modeling Hospital Information Systems (Part 2): using the 3LGM2 tool for modeling patient record management. Methods Inf Med 2004; 43(3):256–67.

[4] Winter A, Brigl B, Wendt T. Modeling hospital information systems (Part 1): The revised three-layer graph-based meta model 3LGM2. Methods Inf. Med. 2003;42(5):544–51.

[5] Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009; 42(2):377–81. doi: 10.1016/j.jbi.2008.08.010.

[6] Sinaci AA, Núñez-Benjumea FJ, Gencturk M, Jauer M-L, Deserno T, Chronaki C et al. From Raw Data to FAIR Data: The FAIRification Workflow for Health Research. Methods Inf Med 2020; 59(S 01):e21-e32. doi: 10.1055/s-0040-1713684.

---

[3] https://www.health-atlas.de/lha/8C6DG2JAKM-3