ICT for Health Science Research
A. Shabo (Shvo) et al. (Eds.)

doi:10.3233/978-1-61499-959-1-146

A Report on Archetype Modelling in a Nationwide Data Infrastructure Project

Antje WULFF[a,1], Kim Katrin SOMMER[a], Sarah BALLOUT[a], HiGHmed CONSORTIUM, Birger HAARBRANDT[a] and Matthias GIETZELT[a]
[a] *Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover, Germany*

Abstract. *Background:* The nationwide data infrastructure project HiGHmed strives for achieving semantic interoperability through the use of openEHR archetypes. Therefore, a knowledge governance framework defining collaborative modelling processes has been established. For long-sustained success and the creation of high-quality archetypes, continuous monitoring is vital. *Objectives:* To present an update on archetype modelling and governance framework establishment in HiGHmed. *Methods:* Qualitative and quantitative analyses of the progress in establishing modelling groups, roles and users, realizing modelling workflows, and modelling archetypes. *Results:* Currently, 25 modellers and 17 domain experts are participating. 79 archetypes have been identified, from which 69 are pre-existing and internationally published; completion rates of review rounds are satisfying but improvable. *Conclusions:* The governance framework is valuable to make the activities manageable and to accelerate modelling. Combined with highly engaged data stewards and clinicians, a reasonable number of archetypes have already been developed.

Keywords. Semantic Interoperability, Clinical Governance, openEHR

1. Introduction

HiGHmed is one of four nationwide data infrastructure projects funded by the German Federal Ministry of Education and Research (BMBF) in the context of the Medical Informatics Initiative. The consortium strives for facilitating secondary use of data by establishing a shared information governance framework, data integration centres and *"an interoperable, open health data platform"* [1]. The infrastructure will enable cross-institutional data access, analytics and sharing to improve patient care and enhance clinical research. Semantic interoperability, semantic traceability and clinically-led data modelling are considered core principles. The openEHR approach is adopted as semantic modelling approach, relying on a reference model, which defines the separation of technical and domain content (two-level modelling [2]). Clinical concepts are modelled with machine-readable and computable archetypes holding different items describing the concept's characteristics. By reusing, nesting and constraining archetypes in templates, different use cases can be covered. In HiGHmed, three clinical use cases (infection control, cardiology and oncology) were chosen to demonstrate the

[1] Corresponding author: Antje Wulff, Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Carl-Neuberg-Str. 1, 30625 Hannover, Germany, E-mail: Antje.Wulff@plri.de

feasibility of the open platform approach [1]. Consequently, both use case-specific and general content need to be captured and modelled in archetypes. To reach *"a semantically-enriched, interoperable and harmonized representation of data across institutions"* [3], we already published a clinical knowledge governance framework, which has to be managed in terms of its establishment and compliance. This is a matter of special importance because the first year of such projects is crucial for long-term success in establishing governance, involving clinicians [4] and creating high-quality archetypes. Pederson et al. underline this by presenting challenges and lessons learned about openEHR implementation in Norway [5]. By tracking their openEHR modelling progress, Min et al. were able to demonstrate the feasibility of modelling electronic health records (EHRs) in China [6]. We pursue a similar management by means of qualitative and quantitative analyses, presenting the involvement of modelling participants and clinicians, the realization of modelling workflows, the advantages and bottlenecks of IT support, and the number of modelled and reviewed archetypes.

2. Methods

By using a prototypical clinical data repository for archetype creation and discussion, we are able to track the number of users, the involvement of data stewards, domain experts and clinicians as well as other statistics. With respects to our workflow management (see [3]), first, the specific use case need was defined in different meetings with clinicians and domain experts. Here, requirements of the scope and the content of use case-specific clinical concepts were collected. This included a qualitative analysis of existing health information systems, clinical processes and practices, as well as other standards and literature. As a result, each use case defined a minimal data set comprising data items, which at least have to be modelled to reach the use case goals. In close cooperation with clinical experts, the data items were organized into categories representing general domain concepts, and mapped onto archetype classes representing the cycle of clinical decision-making (*Instruction*, *Action*, *Evaluation*, and *Observation* [7]). Furthermore, we checked the existence of archetypes from international repositories. This phase is time-consuming and has to be monitored continuously. However, for facilitating archetype sharing and semantic interoperability, existing archetypes should be reused as frequently as possible. When reusing archetypes in Germany, a translation is needed, which is another process in need of management.

Finally, archetype content and translation validation by domain experts is needed to achieve consensus and high-quality archetypes. Therefore, review rounds are conducted by data stewards. The archetypes undergo this iterative process until the domain experts give their approval. During this step, the archetype states (draft, under review, published, rejected) are tracked. When published, both the original and the translated archetypes will be made accessible within the global openEHR community.

3. Results

3.1. Modelling groups, roles and users

The established modelling groups comprise 25 modellers - so-called data stewards - and 17 domain experts (see Fig. 1 (A)), from which all data stewards are involved in

biweekly meetings. As defined in the HiGHmed clinical knowledge governance framework [3], a superordinate *HiGHmed modelling group*, comprising the leading three modellers of each use case and specific modelling experts, was constituted and meets for strategic decisions and archetype publication. At each of the three HiGHmed core sites, *local data stewards* are responsible for requirement analysis, archetype creation and modification. Because of the great importance of the 17 *domain experts*, the data stewards are not only located at the Medical Data Integration Centres (MeDIC), but also directly in the participating medical departments. Locally, each leading data steward defines the members of a *use case modelling group* and meets with their domain experts and IT architects. However, not every data steward has to work exclusively for one use case.

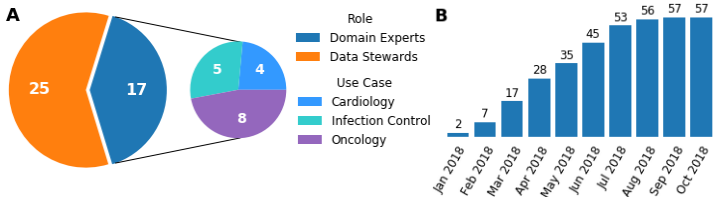

Figure 1. (A) Current number of data stewards and domain experts in HiGHmed; **(B)** Current users per months of the prototypical HiGHmed archetype repository

As proposed in [3], the data stewards successfully use open source modelling tools as the Archetype Editor and the Template Designer. In terms of archetype governance, an archetype repository (Clinical Knowledge Manager, CKM) is in use for collaborative authoring, commenting and reviewing of archetypes. Due to the ongoing recruitment, the number of registered users was steadily increasing during the first half year. Today, recruitment is almost finished, and many clinicians are registered (Fig.1 (B)).

3.2. Modelling workflows

Archetype modelling started simultaneously in January 2018 but the current progress varies between the use cases. An initial data set is defined for all use cases. However, some aspects still await final agreements. A possible reason for delays in data set definitions and requirements analysis is the uncertainty about the amount, the source and the structure of data stored. Nevertheless, to avoid a further delay in modelling, the data stewards started to map the defined data sets onto clinical concepts, and consequently, onto openEHR archetypes. According to the HiGHmed milestones, the modelling efforts for all three use cases will be finished by the end of 2019. Currently, 79 archetypes have been identified, from which 69 are internationally published archetypes (see https://openehr.org/ckm), and ten archetypes are newly created or modified (see https://ckm.highmed.org) (Fig. 2 (A.1)). Local modifications of international archetypes are required to adapt the archetypes to the health care culture and the definition of health-related concepts in Germany. However, we aim at preserving the original structure of the archetype as much as possible, e.g. by specialisation and enhancement of data items or by using the provided slots for inclusion of specific local *Cluster* archetypes. The translation and review processes for 41 reused archetypes are currently ongoing. Up to now, 48.8% of the archetypes are in review round 1, 24.4% in review round 2 and 7.3% in review round 3 (Fig. 2 (A.2).

Overall, 32 first and 7 second review rounds have already been completed. Because the translating data steward decides on adequate reviewers and not every review invitation is followed, the percentage of both the invited reviewers and the completed reviews differ (see Fig. 2 (B.1 and B.2)). Overall, the completion percentage was 36.5% for review round 1 and 51.1% for review round 2.

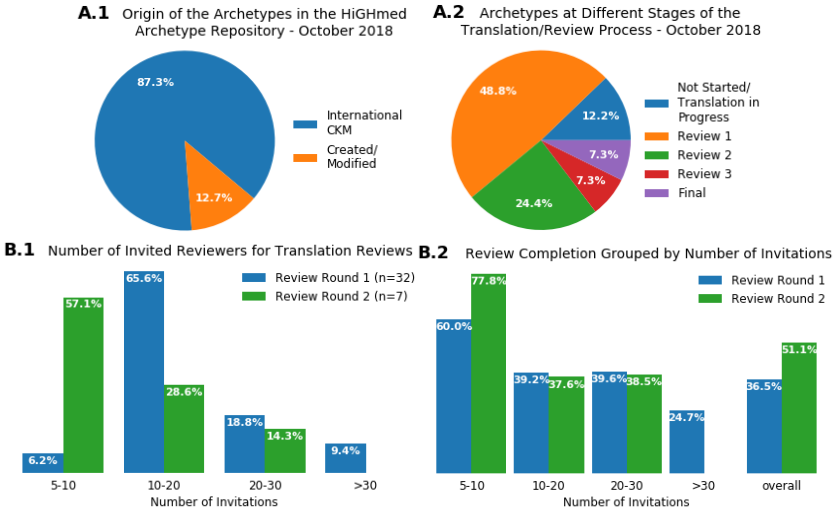

Figure 2. (A.1): Current proportion of reused and created archetypes in HiGHmed; **(A.2)**: Current proportion of archetype states; **(B.1)** Percentage of invited reviewers for translation review round 1 and 2; **(B.1)** Percentage of accepted and completed reviews

4. Discussion

In our work, we present a concise overview of the progress on archetype modelling in a nationwide data infrastructure project. Although the HiGHmed project is still in its first year, the defined processes and workflows for archetype modelling, as published in [3], have been established successfully. The number of data stewards and archetypes available in our repository is steadily increasing. Another promising approach in HiGHmed is the reuse of international archetypes. However, we are in an early stage of modelling and are aware that some enhancements with local content will be needed. This means that the number of specific clusters, which can be used within archetype slots to capture local content, will increase in near future, especially, with an increasing number of clinicians engaged in content review rounds. For example, at first glance, the use of the international laboratory test result archetype seemed appropriate for representing microbiology laboratory test results because all defined items from the minimal data sets are available. However, since microbiology experts are involved, the creation of a specific archetype capturing more specific microbiology items is intensively discussed. Here, our concept of positioning data stewards directly within the clinic has been proven valuable to ensure expert involvement. Furthermore, content-related issues can be solved even before conducting content-specific review rounds. With respects to the archetype state statistics, we were not able to publish our

archetypes yet. This is due to the fact that the content reviews rounds have not finished. Furthermore, the participation in review rounds is satisfying but improvable (completion rates between 36.5% and 51.1%). Consequently, we did not upload our translations or the new archetypes to the international space yet. Nevertheless, we are in close cooperation with international modellers from the openEHR foundation and try to involve international experts from the beginning and also try to contribute to international and European discussions. From our experience, this requires high engagement, courage and motivation of data stewards. Our division of modelling groups and tasks works well here. A very time-consuming task, which has not been foreseen and not covered with any statistics, is agreeing on terminology codes and standardized, structured lists. Therefore, data stewards have extensively worked with original data lists originating from each institution and internationally available terminologies and standards (e.g. LOINC, or WHO lists for laboratory-related names), and try to agree on suitable mappings. This task is a forthcoming critical step, potentially slowing down the modelling process. The final percentage of archetypes needed cannot be determined because some archetypes might still change or some more, yet unknown, data items and archetypes will be needed. However, the major part of the minimal data sets is largely covered.

Overall, with our work, we share our progress on archetype modelling as well as our experiences and lessons learned. Furthermore, we provide a first step towards an extensive evaluation of archetype modelling and the benefits of using a clinical knowledge governance framework. Within our context, our framework already helped to make the modelling efforts manageable and to accelerate the modelling process. In combination with highly engaged data stewards and clinicians, we were able to create a good stack of archetypes within the first year of the project.

Acknowledgements

We want to thank all data stewards and clinicians for their engagement and support. The research for this paper was done within the project “HiGHmed” as part of the German Medical Informatics-Initiative (MI-I). The project is funded by the German Federal Ministry of Education and Research (BMBF, grant id: 01ZZ1802C).

References

[1] B. Haarbrandt, B. Schreiweis, S. Rey, et al, HiGHmed - An Open Platform Approach to Enhance Care and Research across Institutional Boundaries, *Methods Inf Med* **57** (S01) (2018), e66–e81.

[2] T. Beale, Archetypes: constraint-based domain models for future-proof information systems, *Eleventh OOPSLA workshop on behavioral semantics: serving the customer*: Northeastern University, Seattle, Washington, Boston, 2002, 16–32.

[3] A. Wulff, B. Haarbrandt B, M. Marschollek et al, Clinical Knowledge Governance Framework for Nationwide Data Infrastructure Projects, *Stud Health Technol Inform*. **248** (2018), 196–203.

[4] S. Ljosland Bakke, National governance of archetypes in Norway, *Stud Health Technol Inform*, **216** (2015), 1091.

[5] R. Pedersen, G. Conceição, L.M. Ruiz, Implementation of OpenEHR in Combination with Clinical Terminologies: Experiences from Norway, *Int J Adv Life Sciences (IJALS)*. **9** (3&4) (2017), 82–91.

[6] L. Min, Q. Tian, X. Lu, et al, Modeling EHR with the openEHR approach: An exploratory study in China, *BMC Med Inform Decis Mak.* **18** (1) (2018), 75.

[7] OpenEHR Foundation, *openEHR Architecture Overview*, http://www.openehr.org/releases/BASE/Release-1.0.3/docs/architecture_overview/architecture_overview.html, Last access: 21.01.2019.