German Medical Data Sciences 2023 — Science. Close to People.
R. Röhrig et al. (Eds.)

© 2023 The authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/SHTI230716

Towards a Recommendation for Good Health Data Modeling (GHDM) - Results of Expert Interviews

Lena ELGERT[a,1] Jendrik RICHTER [b], Matthias KATZENSTEINER [c], Mareike JOSEPH [a], Sandra HELLMERS [d], Oliver J. BOTT [c], and Klaus-Hendrik WOLF [a]

[a] *Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover, Germany*
[b] *Department for Medical Informatics, University Medical Center Göttingen, Germany*
[c] *University of Applied Sciences and Arts Hannover, Germany*
[d] *Assistance Systems and Medical Device Technology, Carl von Ossietzky University of Oldenburg, Germany*

Abstract. Appropriate data models are essential for the systematic collection, aggregation, and integration of health data and for subsequent analysis. However, recommendations for modeling health data are often not publicly available within specific projects. Therefore, the project Zukunftslabor Gesundheit investigates recommendations for modeling. Expert interviews with five experts were conducted and analyzed using qualitative content analysis. Based on the condensed categories "governance", "modeling" and "standards", the project team generated eight hypotheses for recommendations on health data modeling. In addition, relevant framework conditions such as different roles, international cooperation, education/training and political influence were identified. Although emerging from interviewing a small convenience sample of experts, the results help to plan more extensive data collections and to create recommendations for health data modeling.

Keywords. health information systems, health data, data model, expert interview

1. Introduction

Standards-based modeling of health data is a common way to achieve semantic interoperability and facilitate the integration of patient data from different health systems (1). Appropriate data models are essential for the systematic collection, aggregation, and integration of health data into platforms and for subsequent analysis. Therefore various modeling approaches with different degrees of freedom for health data modeling have been developed during the last years such as OMOP, PCORNet, CEN ISO 13606, HL7-CDA, OpenEHR, CIMI, and HL7's FHIR [1,2].

Extensive specifications exist e.g. for the approaches HL7 FHIR [3], openEHR [4], and OMOP [5]. In addition, however, institutions or projects develop their own project-

[1] Corresponding Author, Lena Elgert, Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Carl-Neuberg-Str. 1, 30625 Hannover, Germany; E-mail: elgert.lena@mh-hannover.de.

or site-specific recommendations and manuals for modeling health data that goes beyond the specifications and corresponds to the respective internally defined governance processes. This is in line with the recommendations for the EU/US eHealth interoperability roadmap which recommends action plans at the appropriate level (local, regional, national, European or global). When creating a roadmap regarding electronic health records and patient summaries, it is recommended to "get back to basics" by working on concrete projects and focus on communication and sharing of experiences. [6] Since the institute's or project's internal recommendation for modeling is often not published, this opportunity to share modeling experiences is missing.

Against this background, the "Zukunftslabor Gesundheit" project aims to research these experiences as a basis to develop and publish recommendations for health data modeling to further improve the quality of health data modeling.

2. Method

2.1. Development of an interview guide

Based on a systematic literature review, the project team developed an interview guide following the collect, review, sort, subsume (SPSS) method of Helfferich [7]. Two authors created an initial collection of questions. In the next step, they checked the questions for openness, prior knowledge, and necessity. Questions with duplicate content, closed questions, and questions with unclear wording were adjusted or removed. The remaining questions were sorted into topic blocks: Warm Up, Composition of the Team, Process of Modeling, Modeling, Review Process, Change Management, Porting of Models, Support (by people, structures, guides), (Inter-)national Development, and Closure. The project team consented on these blocks and subordinate questions.

Qualitative content analysis of a pilot interview revealed the following four different roles of the potential interviewees: 1. clinical data steward (expert for content-related issues concerning the data of the respective area), 2. technical data steward (expert for the technical implementation in terms of modeling), 3. data integration expert (expert for data integration and data transformation, ETL expert), 4. lead/manager (expert for management and control of projects for modeling of health data).

The project team adapted the questionnaire to each of these roles individually by omitting less relevant or redundant questions. For example, manager received the question "What were the advantages and disadvantages of using these modeling standards?" to provide a deeper understanding why a standard is (not) used in a specific project but the clinical data stewards did not. The topic blocks remained the same for all roles.

2.2. Interview conduction and content analysis

A convenience sample of five experts with experience in health data modeling was recruited. The sample represented all identified roles and consisted of four women and one man. Semi-standardized interviews were conducted via videoconference with durations between 1 h 20 min and 2 hours. With interviewees' consent, all interviews were recorded. One author conducted the interviews and two varying authors transcribed the answers of the interviewees in parallel in a shared document as paraphrases. The

project team did a final proofreading in a circular procedure after one author reviewed the document against the recorded videos and completed the transcript.

The project team did a qualitative content analysis by deductively-inductively generating main categories for the individual questions based on an extended literature research, team members' prior knowledge of modeling, and the interview material.

Each team member was responsible for creating main categories in one section of the interview and presented the main categories to the team for discussion and inter-communication validation. Afterwards the project team members noted subcategories and examples in a shared document. The notes including possible key categories were discussed and agreed upon in a team meeting. Subsequently the project team classified the results into the categories *"*governance*"*, *"*modeling*"*, and *"*standards*"* and derived hypotheses for recommendations in modeling health data.

3. Results

Comprehensive findings about the interviews and modeling framework are presented here followed by the hypotheses for recommendations on modeling health data.

3.1. Overarching findings

The pilot interview reveals four different roles of modeling experts (1. clinical data steward; 2. technical data steward; 3. data integration expert; 4. lead/manager).

The interviewed experts named structuring, standardization, interoperability, and exchange as objectives of their health data modeling projects. On the one hand, the exchange should address the exchange of data in structured form, which is often not supported by the primary systems. On the other hand, the exchange should operate across sites between different facilities.

Cross-sectoral data exchange has not yet taken place in the experts' projects, but is planned in some cases. These plans concern, for example, the use of outpatient data in an inpatient setting.

The interviews express the experts' desire to align national developments in health data modeling more closely with international standards and initiatives. The interviewees state that with regard to national and international reusability, the comprehensibility of a model and the usability of the structured, standardized data models are relevant. They are currently experiencing an increasing and very valuable international exchange with an open, engaged community, but one that can still be expanded. This exchange can also help with continuous improvement of the team.

Support for beginners in modeling and advanced modelers is available via exchange with others, documentation, communication, and various trainings. Within the experts' project, learning takes place through regular discussion of problems in conference calls and scheduled meetings. An internal project modeling wiki and modeling manual, chat channels, forums, workshops, a knowledge repository and an user manual for it represent further exchange and learning opportunities. The latter also across projects. The knowledge repository also documents examples that can be used to improve one's own modeling skills. Training courses for new employees are held on a needs-oriented basis during induction. Existing implementations also serve as learning examples. A generally applicable and, above all, generally accessible modeling guide or training course does

not yet exist. However, the interviewees state that this would be useful and an open to-do for the German-speaking region.

For future development, the experts also see an increased need for the deployment of technical and clinical data stewards. These make an important contribution in the modeling process by having an overview of the data from its creation to its use and subsequent use.

Research, industry and/or policy support or hinder standardization of health data modeling. Thus, vendor lock-in on the part of manufacturers/economy often seems to be an obstacle towards open standards. Political influence and will for standardization are important for nationwide implementation. This is also true, because modeling has a high demand for interdisciplinary human resources. Political coordination in the sense of a stabilization and centralization of modeling and a detachment of modeling from individual projects is necessary according to the interviewees. This includes coordination across consortia of the German Medical Informatics Initiative as well as nationwide.

The project team identified the categories “capturing the "real" world”, “implementing the modeling standard”, and “governance” particularly important for health data modeling. The coordination, definition, and adherence to governance processes is fundamental to the success of a modeling project. In addition to capturing and mapping clinical concepts including relevant contextual conditions, data models have to be maintained after their initial creation or beyond the end of a project. This concerns the selection and application of modeling tools such as a knowledge repository.

3.2. Hypotheses for recommendations on modeling health data

The results of the interviews are assigned to the categories “governance”, “modeling”, and “standards” and result in the hypotheses for recommendations on modeling health data below. It should be noted, that not all five interviewees made statements on all points, but the statements often complement each other.

Hypothesis 8, for example, is based on statements of two experts. One expert states "not [being] aware of any superior approach, but methodologically the exchange of all participants is very important". However, this is independent of the approach. Another expert states, "There is no superior approach. It has to be chosen depending on the application purpose. For our own purposes we have a combination of standard XX and standard YY. The goal is to collaborate and use the standards together by developing technologies to do so (rather than always developing new standards to bring standards together)." These individual statements are assigned to the category standards and result directly in hypothesis 8.

Hypotheses for recommendations on modeling health data

1. Adherence to (governance) processes and alignment represent key points of functional modeling. If the governance framework with clearly described processes and distribution of roles for modeling is adhered to, this contributes to target-oriented modeling. It is important that all steps in the modeling process are followed. Also preparatory steps must not be neglected. Here it is important that the relevant domain knowledge is first fully queried, the data set is

sufficiently tuned and the structure and scope of the data in or from the primary systems are well understood.

2. Knowledge of generic modeling as well as comprehensive knowledge of the reference model is required for the implementation of modeling.

3. Domain experts should have an overall understanding of the modeling process, especially including an understanding of the consequences and costs of changes when documentation or structures in primary systems change. Currently, this often only emerges during collaboration in the course of the project.

4. Avoid narrowly defined modeling with a strong orientation to the primary system, as well as unstructured, non-standardized modeling or frequent revisions in the sense of very frequent or detailed changes.

5. In the case of a major revision of the data structures, it should be considered in each case to what extent the effort of a renewed complete data integration into new structures is worthwhile compared to the advantages of updated data models (costs/benefits).

6. Multiple rounds of reviews are recommended to ensure the comprehensibility and usability of the developed models. During the modeling of clinical concept models, three more technically oriented review rounds and a final technically oriented review round with the domain experts are recommended. Between the review rounds, there should be a period of time for revision, literature research or conceptual rethinking for new insights and ideas.

7. Modeling must strive for extensive technical functionality while maintaining a high level of comprehensibility and readability for users. Approaches that support both aspects at an appropriate degree are preferable.

8. There is no superior modeling approach/standard. It depends on the requirements and the goals of modeling, which approach, or combinations are suitable. The goal should be to use existing standards in combination with each other through the development of appropriate technologies, rather than developing new standards to meet open needs.

4. Discussion

The expert interviews reveal important aspects of modeling and of the framework for modeling. The results help the Zukunftslabor Gesundheit project team to plan a more extensive data collection and to create recommendations for health data modeling.

The different roles “clinical data steward”, “technical data steward”, “data integration expert”, and “lead/manager” have emerged. These are of interest for further surveys regarding group-specific evaluations. Variances in responses are to be expected and will be further examined. In this interview study, a characterization of the interviewees beyond role and gender as well as the presentation of individual statements

was not planned, since the identification of persons of the small sample within a well-networked community should be avoided.

Interviewees had no previous experience with cross-sector data exchange. Therefore, it can be assumed that relevant aspects of data modeling for this purpose are not or not yet sufficiently present in the results. These aspects may involve challenges of cross-sectoral data sharing such as data sharing regulations, data exchange capabilities, and cross-sector data integration [8].

The advantages of international cooperation and connection to and compliance with international standards became clear. These results are consistent with the findings of a study on the reuse of archetypes in the openEHR standard. Here, a significant reuse of archetypes was shown in the case of multilingual implementations [9].

Internal education, training and continuing education in large-scale projects seem to work well, supported by numerous tools and in different formats. However, cross-project and publicly accessible training based on experience in the form of modeling guidelines or courses is missing so far. This would be helpful especially for smaller projects. Furthermore, this aspect is important with regard to an increasing demand for well-trained modelers.

Standardization can be promoted by politics and industry. The German Medical Informatics Initiative is a good example of political support [10]. Nationwide coordination of the various modeling initiatives should be a next step. On the industry side, customer retention is an important goal of manufacturers. A change in vendor strategy toward open standards would further drive modeling. Standards such as DICOM have gained acceptance, for example, by being included in calls for proposals. This would also be an opportunity for (political) influence on the standardization of modeling with its high demand for interdisciplinary human resources.

Based on the condensed categories of governance, modeling and standards, the project team was able to generate eight hypotheses for recommendations on health data modeling. These are mainly at the practical modeling level. Due to the limited scope of this congress paper, it is not possible to embed all hypotheses in the current literature. Exemplarily this is shown for hypothesis 8. Hypothesis 8 states that there is no superior modeling approach and recommends a target-oriented selection and combination of modeling approaches. This hypothesis can be supported by Oemig et al. who postulate "[...] there are sufficient standards with specific focus and individual advantages and disadvantages. For the implementation of the partial aspects of the interoperability problem to be worked on, we recommend [...] to choose combinations." [11] One such combination can be the use of openEHR-based servers with a FHIR-Broker. With regard to the growing international use of openEHR-based platforms as an alternative to classical information system architectures, Haarbrandt and Wulff state that openEHR complements the interface-oriented exchange of FHIR with its robust and scalable patient record architecture with regard to the development of clinical application systems and research registries. [12]

The eight hypotheses and other findings emerged from interviewing a small convenience sample of experts working on similar projects. This has to be mentioned as a limitation. With a nationwide online survey, the previous results and hypotheses are to be tested, extended, and further discussed against the background of the current literature in a next step.

Declarations

Conflict of Interest: The authors declare, that there is no conflict of interest.

Author contributions: KHW: direction of the project; LE, MJ, JR, MK, SH: conception of the work, data aquisition and interpretation; LE, JR, MK: writing the manuscript; KHW, OB substantial revising of the manuscript. All authors approved the manuscript in the submitted version and take responsibility for the scientific integrity of the work.

Acknowledgement: Funded by the Lower Saxony Ministry of Science and Culture under grant number ZN3491 within the Lower Saxony "Vorab" of the Volkswagen Foundation and supported by the Center for Digital Innovations (ZDIN).

References

[1] Gamal A, Barakat S, Rezk A. Standardized electronic health record data modeling and persistence: A comparative review. Journal of biomedical informatics 2021; 114:103670.
[2] Rosenbloom ST, Carroll RJ, Warner JL, Matheny ME, Denny JC. Representing Knowledge Consistently Across Health Systems. Yearbook of medical informatics 2017; 26(1):139–47.
[3] HL7 FHIR Foundation. HL7 FHIR Release 5: Welcome to FHIR [Internet]. 2023 [cited 2023 May 26]. Available from: https://www.hl7.org/fhir/.
[4] openEHR International. openEHR Specifications: Open industry specifications, models and software for e-health [Internet]. 2023 [cited 2023 May 26]. Available from: https://specifications.openehr.org/
[5] Observational Health Data Sciences and Informatics (OHDSI). Standardized Data: The OMOP Common Data Model [Internet]. 2023 [cited 2023 May 26]. Available from: https://www.ohdsi.org/data-standardization/
[6] Chronaki C, Stegwee R, Gille H, Cangioli G, Mykkänen J, Bergman V, Rasmussen J. TRILLIUM II Reinforcing the Bridges and Scaling up EU/US Cooperation on Patient Summary:
WP5 D5.1 v2019-08-02 Recommendations for the EU/US eHealth interoperability roadmap: Open Innovation in digital Health: the case of the international patient summary- WP5-HL7 [Internet]. 2019 [updated 2019 Aug 2; cited 2023 May 26]. Available from: https://ec.europa.eu/research/participants/documents/downloadPublic/U0Y0TXNJbnV2T0JQTklIQlZobTRjU2d3Y3FqdmlISEFNdVpDQWpxWGFlZVRsZ0F4dEkyNEFBPT0=/attachment/VFEyQTQ4M3ptUWQzNGtsYk9kdDZOUXpnTG45UVNLejU=
[7] Helfferich C. Die Qualität qualitativer Daten: Manual für die Durchführung qualitativer Interviews. 4th ed. Wiesbaden: VS Verlag; 2011. 214 p. (Lehrbuch). ger.
[8] Walker DM, Hefner JL, DePuccio MJ, Garner JA, Headings A, Joseph JJ, Clark A. Approaches for overcoming barriers to cross-sector data sharing. Am J Manag Care. 2022;28(1):11–6.
[9] Leslie H. openEHR Archetype Use and Reuse Within Multilingual Clinical Data Sets: Case Study. J Med Internet Res. 2020;22(11):e23361.
[10] TMF – Technologie- und Methodenplattform für die vernetzte medizinische Forschung e.V. Medizin Informatik Initiative [Internet]. Berlin. 2023 [cited 2023 Mar 8]. Available from: URL: https://www.medizininformatik-initiative.de/de/start.
[11] Oemig F, Helmer A, Birkle M, Blobel B. Architekturvergleich verschiedener Lösungsansätze für die Interoperabilität in der Medizin: Ein Überblick über das bereits Vorhandene. Available from: https://e-health-com.de/fileadmin/user_upload/dateien/Downloads/EHC_5_2018_Serie_Standards__7_Langversion.pdf
[12] Haarbrandt B, Wulff A. Plattform-Architekturen: Offen in Europa. EHEALTHCOM [Internet]. 2018 [cited 2022 Jun 22];(5):14–8. Available from: https://e-health-com.de/fileadmin/user_upload/dateien/ePaper/EHC_5_2018_ePaper.pdf.