Exploring Complexity in Health: An Interdisciplinary Systems Approach
A. Hoerbst et al. (Eds.)

doi:10.3233/978-1-61499-678-1-162

Metadata Repository for Improved Data Sharing and Reuse Based on HL7 FHIR

Hannes ULRICH[a,1], Ann-Kristin KOCK[b], Petra DUHM-HARBECK[b], Jens K. HABERMANN[c,d] and Josef INGENERF[a,b]
[a] *Institute of Medical Informatics, University of Lübeck, Germany*
[b] *IT for Clinical Research, Lübeck, University of Lübeck, Germany*
[c] *Interdisciplinary Center for Biobanking-Lübeck, University of Lübeck, Germany*
[d] *Section for Translational Surgical Oncology and Biobanking, Department of Surgery, University of Lübeck & University Clinical Center Schleswig-Holstein, Campus Lübeck, Germany*

Abstract. Unreconciled data structures and formats are a common obstacle to the urgently required sharing and reuse of data within healthcare and medical research. Within the North German Tumor Bank of Colorectal Cancer, clinical and sample data, based on a harmonized data set, is collected and can be pooled by using a hospital-integrated Research Data Management System supporting biobank and study management. Adding further partners who are not using the core data set requires manual adaptations and mapping of data elements. Facing this manual intervention and focusing the reuse of heterogeneous healthcare instance data (*value level*) and data elements (*metadata level*), a metadata repository has been developed. The metadata repository is an ISO 11179-3 conformant server application built for annotating and mediating data elements. The implemented architecture includes the translation of metadata information about data elements into the FHIR standard using the FHIR Data Element resource with the ISO 11179 Data Element Extensions. The FHIR-based processing allows exchange of data elements with clinical and research IT systems as well as with other metadata systems. With increasingly annotated and harmonized data elements, data quality and integration can be improved for successfully enabling data analytics and decision support.

Keywords RDMS, HL7 FHIR, MDR, Data Curation

1. Introduction

The North German Tumor Bank of Colorectal Cancer (ColoNet) founded in 2010 by the university clinics of Lübeck, Rostock, Greifswald and Hamburg is a transregional tumor bank with the aim to collect high quality samples and corresponding clinical data [1]. To guarantee a reliable pooling and data integration the partners harmonized a minimal data set with 33 data elements with each one to 27 values; e. g. age, TNM and sample types. This ColoNet core data set was the basis for the development of local databases at each partner site, where the clinical data is stored separately from sample and laboratory data in a pseudonymized way. To enable further data processing the clinical annotations and biospecimen characteristics are routinely aggregated in a cen-

[1] Corresponding Author: Hannes Ulrich, Institute of Medical Informatics, University of Lübeck, Ratzeburger Allee 160, 23562 Lübeck, Germany; hannes.ulrich@student.uni-luebeck.de

tral integrated Research Data Management System (RDMS). This data pool is uploaded into an i2b2 database ('Informatics for Integrating Biology and the Bedside'), serving as a query tool (http://www.northgermantumorbank-crc.de/) for the consortium and collaborators The integration of further partners – being university clinics or non-university clinics – can be more complicate, because the data might not be particularly collected for the ColoNet project. As a result the integration of additional partners makes the process of data pooling more challenging. Moreover, these partners could potentially use their own data elements with value characterizations differing from the ColoNet core dataset, which makes a manual mapping necessary. Facing the problem of managing and matching the enormous data from different clinical sites, a metadata repository (MDR) is a promising approach. But existing metadata repositories vary in turn with respect to its software implementation and have shortcomings when it comes to the exchange of metadata. Intended to bridge these obstacles, a metadata repository has been developed through the interaction with a FHIR Server using the HL7 FHIR resources both as input and output format. The facilitated exchange of data elements should be enabled by combining a metadata repository with FHIR. Smits et al. [2] compared FHIR with the previous HL7 CDA standard which was used in previous projects with Patient Healthcare Monitoring Report as an interlingua for transforming proprietary device data, i.e. for integrating decision support tools seamlessly [3]. Taking into account some standard specific particularities we are nevertheless convinced that HL7 FHIR is a suitable standard for modern web-based technologies that allows representing resources on suitable granularity levels. Besides the transformation of instance data (value level) using resources like "Patient" FHIR also provides metadata level resources like "Questionnaire" and "DataElement". The solution described in this paper is a hybrid architecture consisting of an ISO 11179-3 conformant MDR server application for interactively annotating and mediating data elements and the translation of these data elements into FHIR resources.

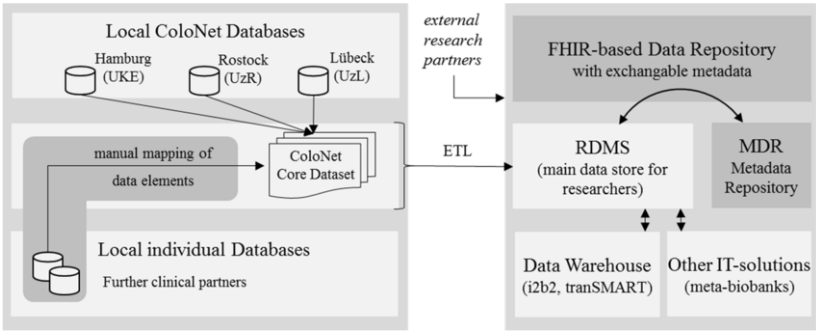

Figure 1. A FHIR server is used for the interaction between the RDMS and the MDR.

2. Methods

In addition to the RDMS, a FHIR server is staged. It allows for the data elements from the ColoNet core dataset, or any other data elements used in Case Report Forms (CRF), to be transformed into FHIR resource data elements so that the CRF itself will be represented as FHIR resource questionnaire. This gathering of CRFs questions is the first step to maintain control of all the study parameters. The use of FHIR makes data elements comparable und reusable, the quality of CRFs can be increased and therefore

simplifies the their creation.. However, using the HL7 FHIR standard is not sufficient for enabling semantic interoperability because there is still a lot of variability for expressing the same content, e. g. sex: (m, f) vs. gender: (m, f, unknown) or weight: real number [kg] vs. weight: (underweight, normal weight, overweight). A system controlling the data elements is supposed to manage individual data items and to annotate them. Therefore, it needs to be able to combine structurally different data elements and to interrelate them to each other. The combination of the introduced standards ISO 11179 and HL7 FHIR shall achieve the aim of improving the management of clinical studies. Furthermore the FHIR-based storage of metadata allows for the exchange of data elements with clinical and research IT systems as well as with other MDR systems, see Fig. 2. Especially with a focus on clinical studies where single CRFs can comprise hundreds of data elements, a FHIR-based metadata repository has been proved a feasible solution [4].

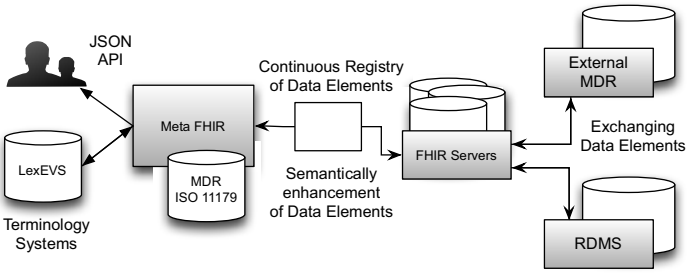

Figure 2. Overview of the implemented architecture

The ISO 11179-3 conformant MDR is represented by a PostgreSQL database. The wide heterogeneity and complex structure of data elements led to problems registering all CRF data without changing the ISO 11179 central base model [5]. But by creating FHIR data elements from the CRFs, it was possible to register them in the metadata repository without changing the central data model. This was done utilizing the connection between the ISO 11179 data element and the FHIR resource 'data element' in combination with the existing FHIR 11179 extensions, see Fig. 3. At startup the system requests all data elements from known FHIR servers and reviews if the data elements are already stored. Before registering unknown data elements, the MDR validates the conformance of data elements to reduce input of poor quality; this task is repeated in a fixed interval. To add new FHIR servers, only their endpoints have to be registered. All data elements shall be associated with a data element concept and a value domain to achieve the ISO 11179 categorization.

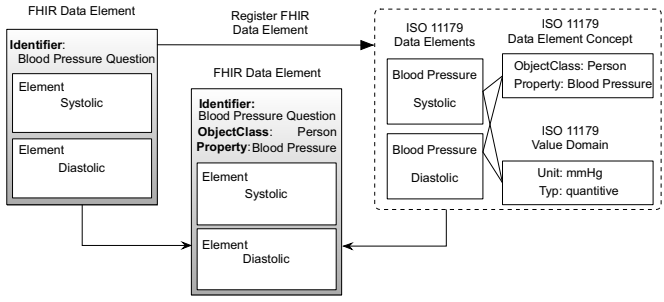

Figure 3. Interrelation between the FHIR and the ISO 11179 data elements using particular parts of the ISO 11179 data element extension by example of a published data element from the MDM Portal [6].

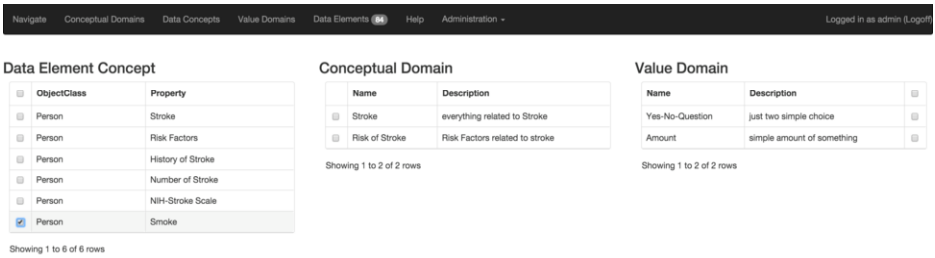

Figure 4 Screenshot of the implemented system showing the interrelation of the available Data Element Concepts, Conceptual Domains and Value Domains, filled with published data elements [6]

Therefore a graphical user interface was implemented, see Fig. 4. Since a metadata repository requires well-maintained data elements, the application provides functions to improve their quality. Using the LexEVS terminology system of the National Cancer Institute, data elements can be enriched with further medical and semantic information. To assist the user while categorizing, a proposal system was implemented. A similarity rate is calculated using the Levenshtein algorithm and is further influenced by semantic codes. On basis of this rating, data element concepts are suggested. Besides the categorization interface, administrative functionality like role and user management was integrated. For the implementation *play* was used, a Java- based web service framework supporting the REST architecture. It offers an internal web server and a built-in object relational mapper. The application follows the common model-view-controller pattern, which ensures functional modularity. The GUI was implemented with HTML and expanded with JavaScript and CSS using the popular frameworks jQuery and Bootstrap.

3. Results

A MDR prototype has been implemented with the possibility to gain added value by linking and mapping data elements. It is possible to define relations between individual data elements and to specify them semi-automatically with the help of a suggestion system. The system has a suitable suggestion schema to identify similar data elements. The same data element, defined with different ranges of values, can be assigned in the same conceptual domain. On the one hand, the MDR enables the user to manage all registered data elements and their corresponding metadata. Data elements from routine patient care or CRFs can be gathered in one single system in a structured manner. It allows for comfortable queries within the classified data elements to look for e. g. 'all data elements associated with the property of weight'. Moreover, the system visualizes the relationships between the elements, see Fig. 4. It also allows for browsing predefined data elements to reuse them in clinical trials. The clinical researchers can pick a selection of data elements and build new CRFs. The reuse of data elements can simplify the cooperation among research groups significantly [7].

On the other hand, the presented architecture has several technical advantages: The hybrid architecture allows not only for fast response times to user requests, but also for a minor data preservation and simple extension of the dataset. The system provides the user with a search function across clinical coding systems, e. g. LOINC and ICD-10. The generated ISO 11179 object class and property entities are repatriated to the FHIR server, which constitutes standardized interchange of MDR information.

To prove this increase in efficiency, the system was evaluated with positive results using two methods: The layout and usability of the system was tested against the ISONORM 9241/110 - S. The classification proposals have been tested on 20 data elements extracted from the MDM portal [6] using cross-validation.

4. Conclusion

As patient care and medical research are reliant on timely available high-quality structured data, this paper presents an approach of designing an ISO 11179-3 conformant MDR. The proposed system offers the possibility to categorize and search for questions of clinical studies. Furthermore, the entire dataset can be effectively overviewed, context related data elements can be interrelated and duplicate entries can be identified. With increasingly annotated and harmonized data elements and mappings for mediating data elements (e. g. units), the data quality and integration can be improved in order to enable data analytics and decision support. These abilities should result in fewer errors and less effort required for designing clinical studies.

In addition to aggregating semantic information for enhancing the quality of the internal and external datasets, the use of UMLS codes offers a large number of controlled medical vocabularies. Alongside the manual input, searching automatically through terminology systems results in a quality improvement without straining users.

HL7 FHIR is a rising exchange standard for clinical information. The utilization offers a unitary data model and simple expandability. Technically, only a FHIR server endpoint is required and it is promising to become commonly used.

To present the advantages of the proposed MDR, a system was implemented based on the existing metadata. The design focuses on easy maintenance as well as expandability; open source libraries were used. By designing and presenting this approach, a tool was offered that succeeds in harmonizing the context of clinical forms and core datasets automatically. In the next step, the proposed system will be further evaluated by integrating the MDR and FHIR server into the ColoNet project's framework.

References

[1] Oberlander M, Linnebacher M, ..., Habermann JK, Consortium C: The "North German Tumor Bank of Colorectal Cancer": status report after the first 2 years of support by the German Cancer Aid Foundation. *Langenbecks Archives of Surgery*. 2013; **398**(2):251-8.

[2] Smits M, Kramer E, Harthoorn M, Cornet R: A comparison of two Detailed Clinical Model representations: FHIR and CDA. European *Journal for Biomedical Informatics*. 2015; **11**(2):en7-en17.

[3] Ingenerf J, Kock A-K, Poelker M, Seidl K, Zeplin G, Mersmann S, et al.: Standardizing intensive care device data to enable secondary usages. *Stud Health Technol Inform.* 2012; **180**:619-23.

[4] Choquet R, Maaroufi M, de Carrara A, Messiaen C, Luigi E, Landais P: A methodology for a minimum data set for rare diseases to support national centers of excellence for healthcare and research. *J Am Med Inform Assoc*. 2014; **22**(1):76-85.

[5] Park YR, Yoon YJ, Kim HH, Kim JH: Establishing semantic interoperability of biomedical metadata registries using extended semantic relationships. *Stud Health Technol Inform*. 2013; **192**:618-21.

[6] Dugas M. Metadata Repository for Medical Forms [Internet]. Medical-data-models.org. 2016 [cited 11 March 2016]. Available from: https://medical-data-models.org

[7] Köpcke F, Kraus S, Scholler A, Nau C, Schuttler J, Prokosch HU, et al.: Secondary use of routinely collected patient data in a clinical trial: an evaluation of the effects on patient recruitment and data acquisition. *Int J Med Inform*. 2013; **82**(3):185-92.