German Medical Data Sciences 2024
R. Röhrig et al. (Eds.)

doi:10.3233/SHTI240850

Metadata Driven Integration of Clinical Data for Secondary Use in FHIR - A Pilot Study at the UKSH

Friedrich SIMON[a,1], Jan SCHLADETZKY[a], Stefanie MACKE[a], Torge ABLAß[a], Josef INGENERF[a,b], and Ann-Kristin KOCK-SCHOPPENHAUER[a]
[a]*IT Center for Clinical Research, Lübeck, Universität zu Lübeck, Lübeck, Germany*
[b]*Institute of Medical Informatics, Universität zu Lübeck, Lübeck, Germany*
ORCDiD-ID: Friedrich Simon – https://orcid.org/0009-0002-9640-4212

Abstract. Introduction The reuse of clinical data from clinical routine is a topic of research within the field of medical informatics under the term secondary use. In order to ensure the correct use and interpretation of data, there is a need for context information of data collection and a general understanding of the data. The use of metadata as an effective method of defining and maintaining context is well-established, particularly in the field of clinical trials. The objectives of this paper is to examine a method for integrating routine clinical data using metadata. **Methods** To this end, clinical forms extracted from a hospital information system will be converted into the FHIR format. A particular focus is placed on the consistent use of a metadata repository (MDR). **Results** A metadata-based approach using an MDR system was developed to simplify data integration and mapping of structured forms into FHIR resources, while offering many advantages in terms of flexibility and data quality. This facilitated the management and configuration of logic and definitions in one place, enabling the reusability and secondary use of data. **Discussion** This work allows the transfer of data elements without loss of detail and simplifies integration with target formats. The approach is adaptable for other ETL processes and eliminates the need for formatting concerns in the target profile.

Keywords. Data Integration, FHIR, Metadata, MDR, Secondary Use

1. Introduction

Integrating different healthcare systems and applications to enable seamless data exchange or to facilitate secondary use is a major challenge in the healthcare sector. This includes a reasonable amount of interoperability, the ability to communicate and exchange data, among the Hospital Information System (HIS), clinical subsystems, and external data sources [1].

In the field of medical informatics, where forms for data collection vary in size and dimension, the Entity-Attribute-Value (EAV) data model provides an innovative solution for handling dynamic and diverse data structures [2]. To integrate data out of this dynamic data model, metadata can play a crucial role in organizing, describing and interpreting data sets [3]. When used to integrate healthcare data, metadata can provide

[1] Corresponding Author: Friedrich Simon, Universität zu Lübeck, IT Center for Clincial Research, Lübeck, Ratzeburger Allee 160, 23562 Lübeck, E-Mail: fr.simon@uni-luebeck.de.

a deeper understanding of patient data, clinical codes and other relevant information [4]. Metadata is more common in the field of medical science mainly in the fields of clinical trials data [5,6].

This paper examines the potential of a metadata repository (MDR) to facilitate and simplify the transfer of data from a primary clinical system into standardized formats (e.g. a FHIR profile). The objective of this research is not only to fulfill legal obligations [7] through the subsequent documentation of metadata, but also to provide a solid basis for future research approaches and improved cross-institutional patient care. In view of the growing volume of data and the increasing importance of data interoperability, this work will examine the potential for deeper data integration through the use of metadata. The hypothesis is that with this prototypical implementation at the University Hospital Schleswig-Holstein (UKSH) metadata can be used to integrate more information such as individual values rather than the result.

2. Methods

2.1. Obtaining of Metadata Information

Essential to the project is the extraction and definition of relevant metadata. A typical HIS stores most relevant clinical information by importing data from connected subsystems, like laboratory or intensive care or direct data collection with forms. We focused on the site-specific forms in this work, as this is where the greatest need for integration exists. In this work the HIS (Orbis Dedalus) is the source of all necessary information, including data elements and contextual data. In principal the HIS used offers a proprietary metadata description, which can be exported as a XML file. However, this export is only available for a subset of forms, so queries had to be created for the retrieval of metadata directly from the database. This is in line with the findings of the literature review by Ulrich et al. [8], which highlighted the lack of standardization and availability of metadata. The forms utilized in the HIS are designed from data elements and form elements (e.g. dividing lines, headings or images). For the data elements and the forms, the underlying tables could be identified in the HIS's EAV database schema, and queries could be developed to export metadata in a standard manner. The query expects the form name as input and, if applicable, a specific version of the form. The result of the query is required separately for each form and serves as the basis for the following definition of the metadata. No further information on the form elements could be identified in the database as part of the research work, although the form elements can serve a major role in the structure and contain additional information, especially with regards to the context of the data collection. Nevertheless, context information such as the sequence, the group affiliation and unique IDs to identify data elements are helpful when creating definitions in the MDR and the clear addressability of form elements provides a solid basis for subsequent further manual definition by experts.

2.2. Definition of Metadata and Corresponding Relations

It is evident that crucial metadata, which is partly missing in the source system, must be defined manually in the MDR by clinical and IT experts. The MDR by Kairos GmbH used in this work provides a graphical user interface for this purpose, thereby facilitating data entry for non-IT personnel. All data elements can be addressed in the MDR and extended with corresponding attributes, such as semantic metadata, which is used to specify the context.

The definition of relations is the central step towards successful mapping in metadata-driven data integration. All relevant metadata about data elements and the corresponding context information from the source system is used as source. The FHIR resource composition has been chosen as target format. It allows inclusion of detailed information through references to other resources in the sections element, thus enabling the modeling of medical information as e.g. observation resources for scores. The target profile only consists of a text field within the MDR, which is filled via the script of the relation, see Figure 1. The script can best be imagined as a template for the target format, as it inserts the instance data from the source at the appropriate places and extends it accordingly with semantic metadata. The relations between the data elements defined by the IT experts, as well as the categorization and metadata, are used to fill the corresponding target format.

Figure 1. Communication of the All-in-one method with the MDR, exemplified by a data element for Hypertensive Disorder, from the CHA2DS2-VASc score form. In the initial step, the request is transmitted to the MDR, accompanied by the source and target profile and the instance data it contains. The MDR then executes the prior defined transformation script on the data and returns the result.

2.3. Transformation of Instance Data

The MDR's ability to reach the target format is achieved through a pipeline, which uses the specification of the source and target format defined in the MDR, as well as the instance data to be transformed. A database query, based on patient id and form name was developed to export instance data from the HIS. This query retrieves the corresponding entries for a patient and form out of the EAV tables. The ability to address data elements uniquely, makes it possible to assign them during instance data export, regardless of the parameter naming the result of this inquiry can be directly used for transformation into a target format.

When calling the MDR's REST interface, the defined relations between the profiles are used to translate the data from the source profile to the target profile. The request is sent via POST command to the endpoint rest/v1/relations/convert with further information in the body of the message. The body contains the following parameters srcProfileCode, srcProfileVersion, trgProfileCode, trgProfileVersion and values (list of instance data to be translated). The result of a valid request to the MDR is an XML response with the following information: trgProfileCode, trgProfileVersion, values (list of all values calculated in the target profile under the corresponding code), usedRelation and logMessages. Defining n:1 relations is necessary to enable this functionality.

3. Results

A metadata-based approach to facilitate data integration was developed using a MDR. The mapping of the structured forms into the FHIR resources was achieved exclusively within a MDR system. In addition to the functionality of mapping relationships between definitions, the MDR used also offers the possibility of translating instance data via a REST interface. This functionality enables the consolidation of metadata, mapping, transformation rules and ultimately the transformation of instance data within one system. As a result, data integration processes were simplified and logic and definitions could be managed and configured at one point. All elements of the form could be mapped in the MDR, and a proposal for mapping as a composition with reference to an observation could be mapped as a relation. Finally, the definition in the MDR was used to translate instance data using the described method. In the current expansion stage, this process has been implemented for two locally defined forms (CHA2DS2-VASc Score and fall protocol). as a FHIR composition with a link to an observation. The observation was implemented according to the FHIR profile of the MII (ICU score). The fall protocol was implemented as a composition with corresponding section headings from the Orbis form. The MDR provides the capability to visualize profiles, definitions, and relations via a graphical user interface. The transformation via the MDR included in the pipeline facilitates the re-usability, the retrievability, and thus the secondary use of data.

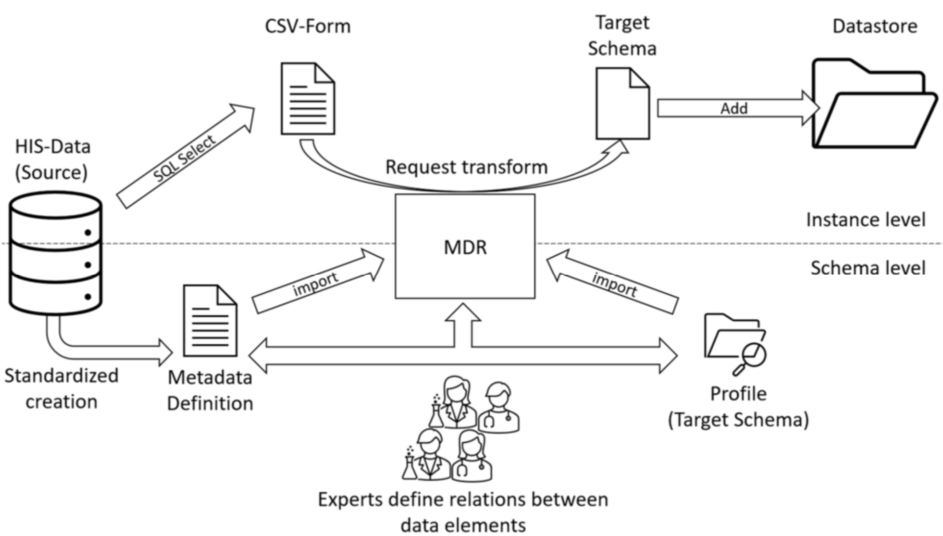

Figure 2. Process of metadata-based data integration, starting with an SQL database of a HIS as the source system and ending with the transfer to a target system (e.g. a FHIR server). The lower part shows the standardized methods, which are used to generate and import metadata definitions for both the source data and the target format into the MDR. The transformation process is visualized in the upper part of the figure, with the data being selectively retrieved from the database and translated into the target format by the MDR before being integrated into the target system.

4. Discussion

When transferring data to subsystems, often only summarized results are transferred and detailed information is discarded. With the presented approach, all data elements of the source form can be retrieved, transformed and used for the target format without a loss of detail. A subsequent comprehensive metadata description of forms must be accurately compiled by IT experts. However, once these have been described in the MDR, it is straightforward to serve other target formats. Definitions and profiles in the MDR can be versioned, reused and customized separately, thereby ensuring sustainability, maintainability and transparency. Moreover, the fact that no instance data is required for the definition of a mapping simplifies integration from a privacy perspective. Maintaining descriptions of forms and data elements in an MDR provides an overview of used definitions and organizational structure within the database. The direct translation of instance data by the MDR into a corresponding target format via a REST interface offers the possibility of adapting this approach in other ETL processes.

5. Conclusion

This approach presents a significant benefit for subsequent uses of not having to concern about formatting in the target profile. Figure 3 below summarises the key aspects in a SWOT analysis.

Strengths	**Opportunities**
• Effective, all form elements addressable • Target profile independent • Centralised logic • Version management • No instance data required for integration	• Sustainability and transparency through access to the MDR profiles for external parties • Integration of further forms
Weaknesses	**Threats**
• Team of experts required • No evaluation of free text elements • Time-consuming MDR configuration	• Dependence on MDR

Figure 3. SWOT analysis of the presented method for data integration.

The acquisition and generation of metadata out of forms via the database is independent of the manufacturer's proprietary exchange format and enables an alternative data source for metadata that offers decisive advantages, e.g. referenced catalogs can be resolved directly in the query and thus exported directly. As important context information cannot be exported to a full extent, it must be manually annotated in the MDR afterwards. A limitation of this work is that free text documentation was not included, even though a significant proportion of medical documentation in the healthcare sector consists of free texts [9]. One disadvantage of this approach is that the information on the individual elements of the target schema that is typically available for MDR is missing and this approach is limited to an n to 1 mapping.

It has been shown that the integration of data with an MDR can be feasible for structured documents, allowing for the acquisition of comprehensive and detailed information in a standardized format. Overall standards should be used wherever possible when creating forms. The work highlights the importance of effective data integration in healthcare and provides important insights for future research in this area.

Declarations

Conflict of Interest: All authors declare that they have no conflicts of interest.

Author contributions: FS conceived the work, acquired data, and interpreted it; JI supervised the work; AKS, JS, SM, TA made substantial contribution to the concept; FS and AKS wrote the manuscript; and JS, SM, TA and JI substantially revised it. All authors approved the manuscript in the submitted version and are responsible for its scientific integrity.

References

[1] Lehne M, Sass J, Essenwanger A, Schepers J, Thun S. Why digital medicine depends on interoperability. Npj Digit Med 2019;2:79. doi:10.1038/s41746-019-0158-1.

[2] Löper D, Klettke M, Bruder I, Heuer A. Enabling flexible integration of healthcare information using the entity-attribute-value storage model. Health Inf Sci Syst 2013;1:9. doi:10.1186/2047-2501-1-9.

[3] Brandt CA, Morse R, Matthews K, Sun K, Deshpande AM, Gadagkar R, et al. Metadata-driven creation of data marts from an EAV-modeled clinical research database. Int J Med Inform 2002;65:225–41. doi:10.1016/s1386-5056(02)00047-3.

[4] Canham S, Ohmann C. A metadata schema for data objects in clinical research. Trials 2016;17:557. doi:10.1186/s13063-016-1686-5.

[5] Dugas M, Neuhaus P, Meidt A, Doods J, Storck M, Bruland P, et al. Portal of medical data models: information infrastructure for medical research and healthcare. Database 2016;2016. doi:10.1093/database/bav121.

[6] Kirsten T, Kiel A, Rühle M, Wagner J. Metadata Management for Data Integration in Medical Sciences. Datenbanksysteme für Business, Technologie und Web (BTW 2017), Gesellschaft für Informatik, Bonn; 2017, p. 175–94.

[7] Bundesgesetzblatt Teil I - Gesetz zur Beschleunigung der Digitalisierung des Gesundheitswesens - Bundesgesetzblatt n.d. https://www.recht.bund.de/bgbl/1/2024/101/VO.html (accessed July 2, 2024).

[8] Ulrich H, Kock-Schoppenhauer A-K, Deppenwiese N, Gött R, Kern J, Lablans M, et al. Understanding the Nature of Metadata: Systematic Review. J Med Internet Res 2022;24:e25440. doi:10.2196/25440.

[9] Kong H-J. Managing Unstructured Big Data in Healthcare System. Healthc Inform Res 2019;25:1. doi:10.4258/hir.2019.25.1.1.