Caring is Sharing – Exploiting the Value in Data for Health and Innovation
M. Hägglund et al. (Eds.)

doi:10.3233/SHTI230117

Identifying Relevant FHIR Elements for Data Quality Assessment in the German Core Data Set

Christian DRAEGER[a,1], Erik TUTE[b], Carsten Oliver SCHMIDT[c], Dagmar WALTEMATH[d], Martin BOEKER[e], Alfred WINTER[a] and Matthias LÖBE[a]
[a] *Institute for Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig, Germany*
[b] *Peter L. Reichertz Institute for Medical Informatics (PLRI), Hannover Medical School, Germany*
[c] *Institute for Community Medicine, University Medicine Greifswald, Germany*
[d] *Medical Informatics Laboratory and Data Integration Center, University Medicine Greifswald, Germany*
[e] *Institute of Artificial Intelligence and Informatics in Medicine, Technical University of Munich, Germany*
ORCiD: Christian Draeger 0000-0001-8835-4548, Erik Tute 0000-0002-9176-3913, Carsten Oliver Schmidt 0000-0001-5266-9396, Dagmar Waltemath 0000-0002-5886-5563, Martin Boeker 0000-0003-2972-2042, Alfred Winter 0000-0003-0179-954X, Matthias Löbe 0000-0002-2344-0426

Abstract. The German Medical Informatics Initiative makes clinical routine data available for biomedical research. In total, 37 university hospitals have set up so-called data integration centers to facilitate this data reuse. A standardized set of HL7 FHIR profiles ("MII Core Data Set") defines the common data model across all centers. Regular Projectathons ensure continuous evaluation of the implemented data sharing processes on artificial and real-world clinical use cases. In this context, FHIR continues to rise in popularity for exchanging patient care data. As reusing data from patient care in clinical research requires high trust in the data, data quality assessments are a key point of concern in the data sharing process. To support the setup of data quality assessments within data integration centers, we suggest a process for finding elements of interest from FHIR profiles. We focus on the specific data quality measures defined by Kahn et al.

Keywords. Data Quality, HL7 FHIR, EHRs, Secondary Use

1. Introduction

The Medical Informatics Initiative (MII) makes healthcare data from German hospitals accessible and interoperable for research. Therefore, EHR data is extracted from the source systems in all German university hospitals, transformed into a common data

[1] Corresponding Author: Christian Draeger, Institut für Medizinische Informatik, Statistik und Epidemiologie, Universität Leipzig, Härtelstraße 16-18, 04107 Leipzig, Germany; E-mail: christian.draeger@imise.uni-leipzig.de

model, and made available via the data integration centers (DIC) [1]. The MII agreed to use HL7 FHIR as their interoperability standard at all 37 DIC sites. A common data model for the so-called MII Core Data Set (CDS) has been defined using FHIR profiles, which allow data to be represented and captured in a structured way. The CDS contains, among others, the profiles *Condition, Observation, Patient,* and *Encounter.* Each profile defines a set of relevant elements, described and published in the FHIR registry *simplifier*[2]. Researchers can request CDS elements from a central portal[3] to receive data from the DICs for research purposes. If a request passes the legal and ethical requirements, that DIC will provide the requested elements in FHIR bundles.

Projectathons are conducted to monitor the progress of the MII by evaluating the data sharing processes and infrastructures. However, the secondary use of healthcare data in real-world settings requires more consideration of data quality (DQ) [2]. Data collected for routine care is not automatically suitable for research purposes in terms of quality and granularity. While quantitative measures can give an indication of the data sets' suitability, a detailed data quality assessment (DQA) is required in the context of a specific use case and for each data provision. Compared with the data collections in clinical trials and epidemiological datasets, it is impossible to influence data collection in routine clinical practice. For example, data collections in clinical routine do not follow a specified sampling frame and standardized examination protocol. Therefore, available tools for data quality assessment in clinical and epidemiological studies are only partially suitable for this application scenario. Additionally, DQA tools developed in the MII [3,4] do not work directly on the FHIR bundles provided by the DIC.

Given this lack of DQA tools for the structured, complex FHIR data stored in the DICs, we (1) analyzed the FHIR profiles established in the CDS in an effort to identify elements with the highest impact toward a comprehensive DQ analysis, and (2) aligned them with the Kahn et al. [5] terminology framework. The Kahn framework was chosen as the base terminology for DQ discussions within the MII. Following the MII decision, we used the Kahn framework to identify relevant FHIR elements for each of Kahn's definitions, taking ongoing implementations for current Projectathons into account.

2. Methods

The recent 6th Projectathon set out to evaluate the data-sharing process, starting with a data request and finishing with a distribution of the required CDS elements as described in the introduction. With this goal in mind, the research question was formulated as: which value of the laboratory parameter NT-proBNP is a suitable marker for the diagnosis of cardiological diseases such as atrial fibrillation, taking age and gender into account? Apart from general patient information, only one laboratory measurement was used.

We first analyzed the CDS with regard to relevant elements for a DQA and then implemented DQ measures for all elements requested in the 6th Projectathon. Kahn et al. [5] defined a comprehensive DQ framework for the secondary use of EHR data. The framework is defined using three categories of data quality. The first category, *Conformance*, "focuses on DQ features that describe the compliance of the representation of data against internal or external formatting, relational, or computational

[2] https://simplifier.net/organization/koordinationsstellemii

[3] https://forschen-fuer-gesundheit.de/

definitions" [5]. The second category, *Completeness*, "focuses on features that describe the frequencies of data attributes present in a data set without reference to data values". The third category, *Plausibility*, "focuses on features that describe the believability or truthfulness of data values". These three main categories are further divided into subcategories, and lastly into definitions represented by one letter. The framework differentiates between the internal context, *Verification*, which considers "how data values match expectations with respect to metadata constraints, system assumptions, and local knowledge", and the external context, *Validation*, which considers "the alignment of data values with respect to relevant external benchmarks". Our analysis addresses *Verification* only. The Kahn framework gives detailed definitions for DQ measures in each category. For each of the proposed definitions, we analyzed which CDS elements would be required to implement a DQ measure. The resulting formalizations are independent of any DQA tool implementation. Therefore, we refer to them as data quality indicators (DQIs).

All implementations for DQA in the 6th Projectathon were based on those DQIs. The R-Tool *firecracker*[4] was used to flatten the data as was required by the analysis scripts. We used the R-Package *dataquieR* [3] to perform our DQA. First, we generated the metadata in the format required by *dataquieR*. Secondly, we wrote an R-Script that utilizes the methods provided by *dataquieR* to perform the actual DQA. This R-Script is called during the data-flattening process, and automatically generates a DQ-Report, which now includes results for each of the initially formalized DQIs.

3. Results

Table 1 shows the formalized DQIs for each of the definitions established by Kahn et al. For each definition, we describe relevant elements from the CDS profiles. We also list which DQIs were evaluated during the 6th Projectathon.

The Kahn *Conformance* category was of particular interest for the CDS, as running a FHIR validator should ensure full conformance with the CDS profiles. But especially for this category, processes in the hospital can result in DQ concerns. In some hospitals, the FHIR *Observation* resource is created as soon as a measurement is requested at the hospital laboratory. The same *Observation* resource should later be updated with the measurement value. If this is not possible, a new *Observation* is created instead. This conveys what had happened in the hospital and will be in accordance with the profile during FHIR validation, but violates *Relational Conformance (b),* which addresses such duplicated resources. This requires further investigation in appropriate provenance concepts specifically for ETL processes and associated data sets at the DICs [6].

A surprising problem occurred in the *Plausibility* category. The CDS captures laboratory measurements in *Observations* occurring during an *Encounter*. While the time at which a laboratory measurement was taken is undisputable, the end date of an *Encounter* is not. Patient discharge, recurring patients, outpatient care and billing necessities influence an *Encounter's* start and end time. As such processes can vary between hospitals, violations of *Temporal Plausibility (b)*, which addresses the correct order of sequences, might be more indicative of process differences, than of errors during the laboratory measurement or data capture for *Observations*.

[4] https://github.com/POLAR-fhiR/fhircrackr

Table 1. Applicable CDS profiles and elements for each definition from the Kahn Framework. Definitions are represented by one letter, according to the table provided by Kahn et al. [5, p. 7-8] (verification context).

Kahn Definition	Description	CDS Profile (CDS Elements)	Projec-tathon
Value Conformance (a)	The value consists only of digits and dot.	Observation (NT-proBNP.value)	yes
Value Conformance (b)	Gender consists only of the allowed categorical values (HL7 Administrative Gender).	Patient (Patient.gender)	yes
Relational Conformance (a)	The references are resolvable in both directions (if any), Conditions are determined by the Encounter.	Condition (Condition.id) Encounter (Encounter.id)	yes
Relational Conformance (b)	Do two Encounters with the same start and end date exist with different IDs?	Encounter (Encounter.startdate, Encounter.enddate) *	yes
Relational Conformance (c)	An element from any older MII Core Data Set profile which had been changed since a previous version.	any CDS Profile with different Elements in a previous version	no
Computational Conformance (a)	The calculation of the age from the date of birth by us is equal to the calculated age in the DIZ.	Patient (Patient.birthdate)	no
Completeness (a)	Gender should not be empty.	Patient (Patient.gender)	yes
Completeness (b)	Patients with multiple encounters should have a certain chronic diagnosis in each medical case.	Condition (Condition.id)* Encounter(Encounter.id)*	no
Uniqueness Plausibility (a)	Each individual Encounter should only be assigned to one unique patient.	Patient (Patient.id) * Encounter (Encounter.id)	yes
Temporal Plausibility (a)	The Encounter's end date should not occur before the Encounter's start date.	Encounter (Encounter.startdate, Encounter.enddate)	yes
Temporal Plausibility (b)	The date of the NT-proBNP measurement is between the corresponding Encounter's start and end date.	Encounter (Encounter.startdate, Encounter.enddate) Observation (NT-proBNP.date)	yes
Temporal Plausibility (c)	Chronic diagnoses of a patient gain diagnostic confidence over time.	Condition (Diagnosis.verificationStatus.code)*	no
Atemporal Plausibility (a)	Hard Limits: NT-proBNP < 10,000 pg/ml	Observation (NT-proBNP.value)	yes
Atemporal Plausibility (b)	Difference in value of the same measurement obtained by different instruments.	Observation (NT-proBNP.value)*	no
Atemporal Plausibility (c)	Different hard limits for male and female patients.	Observation (NT-proBNP.value) Patient (Patient.gender)	yes
Atemporal Plausibility (d)	Difference in values for NT-proBNP measurement of a patient in case of multiple measurements.	Observation (NT-proBNP.value)*	no

* required multiple times

4. Discussion and Conclusions

Our results show that well-formalized DQIs will not only find erroneous data points but also have the potential to highlight issues in the underlying data-capturing processes. While a comprehensive DQA setup can reveal differences in those underlying processes, what is admissible depends on the specific use case.

The use case of the 6th Projectathon allowed us to implement such a comprehensive DQA setup based on our initially formalized DQIs for ten of the sixteen DQ measures defined by Kahn et al. While the limited data required in that use case prevented further implementations, we used the FHIR profiles of the CDS to formalize DQIs for all sixteen definitions. We hope that this benefits the DQA efforts of future use cases built on the CDS. Moreover, our efforts highlight the benefits of formalizing DQIs independently of the requirements from a specific DQA tool. Their natural language descriptions allow the inclusion of domain experts with detailed knowledge of the processes at data capture during a clinical routine, the ETL processes at the DICs, or the CDS FHIR profiles, but with limited programming skills. The independent conceptualization offers more flexibility in the later implementation. Also, due to this tool agnostic nature these DQIs are potentially relevant towards further improving already established DQA processes and pre-existing tools deployed at any point of the clinical data life cycle. Especially if such processes and tools had been designed for FHIR data, as is the case with FHIR CQL, which sees some use outside of the MII already. Lastly, the conceptual model can be updated independently of the DQA tool.

For the MII, if those DQIs can be stored in an interoperable, reusable way, they could be beneficial for initial DQA conducted directly at the DICs. In the future, we hope that such a *FAIRification* of independent DQIs can also help to address the currently omitted *Validation* context. Finally, this will enable gold-standard references to be built on the extensive data stored at DICs.

Acknowledgments: Research was supported by BMBF grants SMITH (01ZZ1803A), POLAR (01ZZ1910A) and DFG grants WI 1605/10-2, NFDI 13/1.

References

[1] Gehring S, Eulenfeld R. German Medical Informatics initiative: Unlocking data for Research and Health Care. Methods of Information in Medicine. 2018;57(S 01). doi:10.3414/me18-13-0001.

[2] Kohane IS, Aronow BJ, Avillach P, et al. What every reader should know about studies using electronic health record data but may be afraid to ask. Journal of Medical Internet Research. 2021;23(3). doi:10.2196/22219.

[3] Schmidt CO, Struckmann S, Enzenbach C, et al. Facilitating Harmonized Data Quality Assessments. A data quality framework for observational health research data collections with software implementations in R. BMC Medical Research Methodology. 2021;21(1). doi:10.1186/s12874-021-01252-7.

[4] Kapsner LA, Mang JM, Mate S, et al. Linking a consortium-wide data quality assessment tool with the MIRACUM Metadata Repository. Applied Clinical Informatics. 2021;12(04):826–35. doi:10.1055/s-0041-1733847.

[5] Kahn MG, Callahan TJ, Barnard J, et al. A Harmonized Data Quality Assessment Terminology and framework for the secondary use of electronic health record data. eGEMs. 2016;4(1):18. doi:10.13063/2327-9214.1244.

[6] Gierend K, Krüger F, Waltemath D, et al. Approaches and criteria for provenance in biomedical data sets and workflows: Protocol for a scoping review. JMIR Research Protocols. 2021;10(11). doi:10.2196/31750.