Applying the RDA Data Maturity Model on the Core Dataset of the German Medical Informatics Initiative

Lea Michaelis[1,*], Esther Thea Inau[1], Michael Rusongoza Muzoora[2], Judith A.H. Wodke[1], Thomas Ganslandt[3], Sylvia Thun[2] and Dagmar Waltemath[1]

[1]*Department of Medical Informatics, University Medicine Greifswald, Greifswald, Germany*
[2]*Berlin Institute of Health at Charité - Universitätsmedizin Berlin, Berlin, Germany*
[3]*Chair of Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany*

Abstract

The German Medical Informatics Initiative authored a modularised Core Dataset which enables the sharing of clinical routine data across German hospital sites. In this work, we apply the RDA Data Maturity Model to run a FAIR assessment on one module of the Core Dataset. We summarise the results of this assessment and present the lessons learnt. We anticipate that the assessment will help to design a FAIRification workflow for the Core Dataset and to estimate the overall resources required for that task.

Keywords

Medical Informatics Initiative, FAIR assessment, FAIR Data Maturity Model, FAIR-IMPACT

1. Introduction

Among the main tasks of the German Medical Informatics Initiative (MII) is the development of a Core Dataset (CDS)[1] to facilitate research on clinical routine data across German hospitals [1]. The current version of the CDS is divided into seven basic modules (Person, Treatment Case, Consent, Diagnosis, Procedure, Laboratory test results and Medication) and ten extension modules. Since May 2023 the German National Portal for Medical Research Data [2] accepts data use proposals based on these CDS items.

In parallel to these community-driven developments in medical informatics, stakeholders in the biomedical and health research domains have repeated their call for the findability, accessibility, interoperability and reusability (FAIR) of health data and the implementation of data stewardship [3]. Adherence to the FAIR guiding principles for data stewardship [4] saves time and effort in data exploration and reuse, and it improves the quality of data sets, metadata sets, and data dictionaries. For example, the FAIR principles require substantial metadata for all data items. These metadata can then help with integrating disparate data sources, finding data items, or with comparing data items with respect to semantic similarities.

SWAT4HCLS 2024: The 15th International Conference on Semantic Web Applications and Tools for Health Care and Life Sciences, February 26–29, 2024, Leiden, The Netherlands
lea.michaelis@med.uni-greifswald.de (L. Michaelis)
0000-0001-9691-2677 (L. Michaelis); 0000-0002-8950-2239 (E. T. Inau); 0000-0002-1384-1509 (M. R. Muzoora); 0009-0009-9712-060X (J. A.H. Wodke); 0000-0001-6864-8936 (T. Ganslandt); 0000-0002-3346-6806 (S. Thun); 0000-0002-5886-5563 (D. Waltemath)

[1]https://www.medizininformatik-initiative.de/en/medical-informatics-initiatives-core-data-set

FAIR-IMPACT is a project conducted by the European Open Science Cloud in 2023 which aims to facilitate the implementation of FAIR-enabling practices, tools, and services across scientific communities [5]. We participated in the "FAIR-IMPACT FAIRness assessment challenge" 2023 to run a baseline assessment of the FAIRness of the MII CDS. As the MII CDS is being used for data requests via the German national Research Data Portal, we were interested to get a clearer picture of the current status of implementation. We hypothesise that the presented guide to FAIRification of one exemplary CDS module, the basic module *Person*[2], will add value to the CDS as a whole and that it will contribute to a more structured guidance for health data FAIRification.

2. Methods

Over a three-month period, we participated in three FAIR-IMPACT online workshops (3–4 hours each). The first workshop introduced different FAIR assessment tools and methods. Further on, strategies to identify relevant semantic artefacts and/or datasets were elaborated upon. We used the FAIR Data Maturity Model (FDMM) [6] for our collaborative FAIR assessment of the MII CDS module *Person*. By discussing and exchanging with the MII CDS team, we obtained scores and results for the FDMM and, in consequence, the baseline FAIR assessment results. The resulting discussion in the second workshop also revolved around best practices and avoidable pitfalls after applying the respective FAIRification methods and tools. The third workshop provided the opportunity to present our improvements and changes made and to receive feedback for our FAIRification process. FAIRification is an iterative and gradual process. For clarifications, resolving ambiguities, and identification of possible revision points, we collaborated and continuously aligned with the MII CDS teams.

3. Results

When conducting this work, the MII CDS modules were available as FHIR-compliant profiles on Simplifier.net. We analysed the version of 2022-02-11, which has developed since[3]. HL7 FHIR (Health Level 7 Fast Health Interoperability Resources) [7] is the de facto standard for interoperable health data. An implementation guide[4] with details about a profile, supported FHIR search parameters and related extension profiles, standard operating procedures, code systems, and value sets is available for each MII CDS FHIR profile.

Our FAIR assessment addressed all four aspects of FAIRness separately using the RDA FDMM. The MII CDS reaches an exceptionally high base value with 75% of indicators fulfilled (F:7/7, A:12/12, I:6/12, R:6/10). This is due to the fact that the CDS specifies the national standard format for digital biomedical research data exchange, i.e., reusability is one if not the primary aim, and that experts from medical and technical domains define the specifications together.

[2]https://www.medizininformatik-initiative.de/Kerndatensatz/Modul_Person/IGMIIKDSModulPerson.html
[3] https://art-decor.org/art-decor/decor-datasets--mide-?id=&effectiveDate=&conceptId=&conceptEffectiveDate=
[4]https://simplifier.net/guide/modul-person-ig-1.0-en?version=current

Findability: Rich metadata is provided with each of the FHIR profiles[3]. FHIR profiles contain globally unique object identifiers (IDs) which are available at resource level[5]. ID persistency and discoverability via search engines can still be improved.

Accessibility: The implementation guide enfolds metadata which contains information for data access. Data and metadata can be accessed manually[6]. As the implementation guide provides HTTPS protocols, the data identifier resolves to a digital object and metadata and data are automatically accessible through a standardised free access protocol. One integral part of accessibility is the aspect of authentication and authorization [4]. FHIR does not define any security related functionality, but define exchange protocols and content models that need to be used with various security protocols defined elsewhere [7].

Another integral part of accessibility is that metadata should be guaranteed to remain available after the respective data is no longer available [4]. Since the technical capabilities of FHIR do not guarantee the fulfilment of this requirement, arrangements have been made for the implementation guide to be found in both ArtDecor and Simplifier.net so as to ensure the longevity of the metadata[8].

Interoperability: The FHIR implementation is a machine-understandable knowledge representation [8]. Both the data and metadata are standardised and provided in JavaScript Object Notation (JSON) and Extensible Markup Language (XML) formats, which facilitates data exchange among heterogeneous systems. The semantic quality of FHIR plays a critical role for the interoperability of the CDS items. Generally, FHIR supports the Resource Description Framework (RDF). CDS data items are cross-linked with terms from medical terminologies such as ICD-10, LOINC, OPS codes and SNOMED CT. Terms from these terminologies are resolvable using globally unique and persistent identifiers and thoroughly documented for purposes of findability and accessibility [9, 10]. However, not all the data items have been sufficiently annotated. Furthermore, it is recommendable to track the provenance of semantic enrichments and cross-links [11].

Reusability: FHIR is a globally recognized health information standard that can be used to represent human and machine-readable data and metadata. Thus, the implementation of the MII CDS as FHIR profiles ensures that both the data and metadata comply with relevant community standards as advocated for in the FAIR data principles. FHIR resources include a rich set of attributes describing most relevant data and metadata, which can further be extended to cover additional requirements [12]. The accompanying implementation guide contains a wealth of accurate and relevant attributes for purposes of enhanced data reusability. The FHIR standard offers different means to encode license information and the conditions under which data can be reused. As of now the CDS module *Person* only contains human-readable information related to the conditions under which data can be reused. The metadata furthermore contains versioning

[5]https://www.medizininformatik-initiative.de/fhir/core/modul-person/StructureDefinition/Patient
[6]https://www.medizininformatik-initiative.de/Kerndatensatz/Modul_Person/IGMIIKDSModulPerson.html
[7]https://build.fhir.org/security.html
[8]https://simplifier.net/,https://art-decor.org/

information. Detailed provenance information is missing and should be added, e.g., by using the Provenance resource[9].

4. Discussion and Conclusion

The MII CDS specifies the most common data elements shared across German university hospitals. The reuse of harmonised data contributes to the interoperability of German medical data integration centers and to the reusability of data sets. The CDS development is a community project under the umbrella of an Interoperability Task Force within the MII. We expect that achieving consensus on the matters arising during the retrospective FAIRification as well as implementing decisions will require considerable time and effort. In return, this investment will increase certainty of the future data machine-actionability for analytics and enhance data findability and, in consequence, reusability of clinical routine data [3].

We appreciate the lessons learned from conducting this work, for example that embarking the FAIRification journey for the entire MII CDS requires a deep understanding of the processes by which the CDS is defined, adapted, and expanded.

This work is a classic example of the importance of enthusiastic collaboration, as clarifying the details of the assorted parameters required numerous consultations with the data owners. It is imperative to train the different cooperating stakeholders on what it actually means to strive for "FAIRness" and to support them in the FAIRification process.

The results of our FAIR assessment have led us to identify gaps and areas for improvement, which is now the basis for future FAIRification of the MII CDS that we plan to implement together with the CDS community. In the first step of this implementation, we will adapt existing information by providing data and metadata in the required human and machine actionable formats. This will increase the fraction of fulfilled indicators further by 12%. It is important to understand that although the initial FAIR assessment score may not be as high as expected, the goal is to gradually and iteratively work towards improving this score.

Data FAIRification is critical in the wake of the current reproducibility crisis of publicly funded research [13]. This work is a demonstration of the effectiveness and necessity of streamlined and collaborative FAIRification processes across further medical data initiatives. We anticipate that the results of this work will motivate stakeholders beyond the MII to take on the FAIRification journey for purposes of increased data reusability as a result of better data transparency.

We highly recommend that FAIRification is planned for in the infancy stages of the data lifecycle as this requires considerably less effort than compared to retrospective FAIRification. The value that FAIR data provides is unquestionable.

5. Acknowledgments

We appreciate the guidance provided by the FAIR tutors at the FAIR IMPACT programme and the feedback from Danny Ammon and Hans-Ulrich Prokosch. This work was partially conducted with funding from the German Federal Ministry of Education and Research (MIRACUM, FKZ

[9] https://fhir-ru.github.io/provenance.html

01ZZ1801A, 01ZZ1801M and 01ZZ2019), NFDI4Health – the National Research Data Infrastructure for Personal Health Data (DFG-funded project 442326535), FAIR4Rare (01VSF22026), NAPKON (01KX2121) and the FAIR-IMPACT.

References

[1] S. C. Semler, F. Wissing, R. Heyder, German medical informatics initiative, Methods of information in medicine 57 (2018) e50–e56.

[2] H.-U. Prokosch, M. Gebhardt, J. Gruendner, P. Kleinert, K. Buckow, L. Rosenau, S. C. Semler, Towards a national portal for medical research data (fdpg): Vision, status, and lessons learned, Caring is Sharing–Exploiting the Value in Data for Health and Innovation (2023) 307.

[3] E. T. Inau, J. Sack, D. Waltemath, A. A. Zeleke, Initiatives, concepts, and implementation practices of the findable, accessible, interoperable, and reusable data principles in health data stewardship: Scoping review, Journal of Medical Internet Research 25 (2023) e45013.

[4] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al., The fair guiding principles for scientific data management and stewardship, Scientific data 3 (2016) 1–9.

[5] J. Parland-von Essen, I. Dillo, Fair-impact, Research Ideas and Outcomes (2022). doi:10.3897/rio.8.e96144.

[6] C. Bahim, C. Casorrán-Amilburu, M. Dekkers, E. Herczog, N. Loozen, K. Repanas, K. Russell, S. Stall, The fair data maturity model: An approach to harmonise fair assessments. (2020).

[7] M. Ayaz, M. F. Pasha, M. Y. Alzahrani, R. Budiarto, D. Stiawan, The Fast Health Interoperability Resources (FHIR) standard: systematic literature review of implementations, applications, challenges and opportunities, JMIR medical informatics 9 (2021) e21929.

[8] C. Chronaki, G. Cangioli, A. M. Garcia, P. van Damme, C. L. P. Calderon, FAIRness for HL7 FHIR: supporting interoperability of health datasets (2021).

[9] S. C. Semler, Loinc: Origin, development of and perspectives for medical research and biobanking–20 years on the way to implementation in germany, Journal of Laboratory Medicine 43 (2019) 359–382.

[10] L. Michaelis, R. A. Poyraz, M. R. Muzoora, K. Gierend, A. Bartschke, D. Waltemath, S. Thun, How FAIR Is NUM?–Lessons Learnt from a FAIR Survey Within the German Network University Medicine (NUM), Technical Report, EasyChair, 2022.

[11] K. Gierend, F. Krüger, S. Genehr, F. Hartmann, F. Siegel, D. Waltemath, T. Ganslandt, A. A. Zeleke, Capturing provenance information for biomedical data and workflows: A scoping review (2023).

[12] N. Hong, K. Wang, S. Wu, F. Shen, L. Yao, G. Jiang, An interactive visualization tool for hl7 fhir specification browsing and profiling, Journal of healthcare informatics research 3 (2019) 329–344.

[13] L. Koesten, P. Vougiouklis, E. Simperl, P. Groth, Dataset reuse: Toward translating principles to practice, Patterns 1 (2020).