MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

doi:10.3233/SHTI220024

FAIRness for FHIR: Towards Making Health Datasets FAIR Using HL7 FHIR

Alicia MARTÍNEZ-GARCÍA[a], Giorgio CANGIOLI[b], Catherine CHRONAKI[b], Matthias LÖBE[c], Oya BEYAN[d,e,f], Anthony JUEHNE[d], Carlos Luis PARRA-CALDERÓN[a]

[a] Group for Research and Innovation in Biomedical Informatics, Biomedical Engineering, and Health Economy. Institute of Biomedicine of Seville, IBiS/"Virgen del Rocío" University Hospital /CSIC/University of Seville, Seville, Spain
[b] HL7 Foundation, Brussels, Belgium
[c] Institute for Medical Informatics, Statistics and Epidemiology (IMISE), Leipzig University, Leipzig, Germany
[d] Research Data Alliance, Reproducible Health Data Services WG
[e] University of Cologne, Faculty of Medicine and University Hospital Cologne, Institute for Medical Informatics, Germany
[f] Fraunhofer FIT, Sankt Augustin, Germany

Abstract

Medical data science aims to facilitate knowledge discovery assisting in data, algorithms, and results analysis. The FAIR principles aim to guide scientific data management and stewardship, and are relevant to all digital health ecosystem stakeholders. The FAIR4Health project aims to facilitate and encourage the health research community to reuse datasets derived from publicly funded research initiatives using the FAIR principles. The 'FAIRness for FHIR' project aims to provide guidance on how HL7 FHIR could be utilized as a common data model to support the health datasets FAIRification process. This first expected result is an HL7 FHIR Implementation Guide (IG) called FHIR4FAIR, covering how FHIR can be used to cover FAIRification in different scenarios. This IG aims to provide practical underpinnings for the FAIR4Health FAIRification workflow as a domain-specific extension of the GoFAIR process, while simplifying curation, advancing interoperability, and providing insights into a roadmap for health datasets FAIR certification.

Keywords:

Reference Standards, Health Information Interoperability, Guideline.

Introduction

Medical data science aims to facilitate knowledge discovery assisting humans and machines in analysis of algorithms, tasks, and results.

The FAIR (Findable, Accessible, Interoperable, Reusable) principles focus on guiding scientific data management and stewardship, and therefore are relevant to all stakeholders in the digital health ecosystem. The FAIR principles first published in 2016 [1], are part of the European Open Science strategy, aiming to advance Research and Innovation in the European Union. They are also part of the United States NIH Data Science strategy.

The European Open Science Cloud (EOSC) is an initiative of the European Commission to facilitate European scientists' access to scientific data, computing platforms and computing services. GoFAIR [2] is a bottom-up international approach for the practical implementation of EOSC in the global Internet of FAIR Data & Services.

HL7 Fast Healthcare Interoperability Resources (FHIR) [3] is a next generation standards framework created by HL7, which provides APIs (Application Programming Interfaces) based on data formats or structures (known as "Resources") for accessing and exchanging EHRs.

The FAIR4Health European project [4] aims to facilitate and encourage the health research community to FAIRify, that is, to augment, share and reuse datasets derived from publicly funded research initiatives, demonstrating the potential impact of the FAIR4Health strategy on health outcomes and health research. Thus, FAIR4Health project set out to stimulate the safe sharing and reuse of health datasets developed in health research projects. The possibility of using datasets from electronic health record (EHR) systems related to healthcare provision is another important consideration. Reusing EHR data for clinical research generates numerous benefits [5]. The FAIR4Health project uses the GoFAIR approach as the basis to define a specific workflow taking into account the specific requirements of medical data.

The Research Data Alliance (RDA) was launched as a community-driven initiative in 2013 by the European Commission, the National Science Foundation and National Institute of Standards and Technology of the United States, and the Department of Innovation of Australia with the goal of building the social and technical infrastructure to enable open sharing and re-use of research data [6]. RDA generated intense activity in developing a framework and components for effective use of the FAIR principles promoted by the scientific community. Especially relevant is the RDA FAIR Data Maturity Model which has been employed in FAIRness for FAIR to assess health data sets [7].

A FAIR digital object is conceptually represented by [8]:

- A Persistent Unique Identifier.
- Data Elements that represent the actual health data (single association between two concepts, images, raw data, etc.). They are practically, although not technically, distinct from metadata.
- Metadata providing the context of the data. "A Data Object should minimally contain basic machine actionable metadata that allows it to be distinguished from other data objects" and it "should be sufficiently rich that a machine or a human user, upon discovery, can make an informed choice about whether or not it is appropriate to use that Data Object in the context of their analysis".
- Provenance describing entities and processes involved in producing and delivering a specific resource. Provenance is a key for FAIR data.

In the context of the FAIR4Health project, the capability of the HL7 FHIR standard to provide support for the realization of the FAIR principles in healthcare data has been analyzed [9], highlighting how the FAIR digital object conceptual elements map into FHIR resources.

Preliminary results demonstrate how the FAIR digital object conceptual elements can be mapped to the FHIR resources. In this way, the components of FAIR digital objects can be represented in HL7 FHIR, thus increasing their interoperability and reusability. Appropriate FHIR profiles can further facilitate advanced search capabilities based on extensive metadata, and easier access to the health data capitalizing on the data model of HL7 FHIR. Previous experiences have assessed the use of HL7 FHIR to increase the adoption of data repositories related with clinical observational studies [10].

The objective of this paper is to present the performed work and next steps of the 'FAIRness for FHIR' project in the direction of developing a roadmap for FAIR data certification.

FAIRness for FHIR project objectives

The main goals of the 'FAIRness for FHIR' project are:

- Facilitate the collaboration between the FAIR and the FHIR communities.
- Enable cooperative usage of the FHIR standard to consistently implement the FAIR principles.
- Support the assessment and implementation of FAIR health data by using HL7 FHIR.

The FHIR4FAIR IG aims to provide guidance on how HL7 FHIR can be used to support FAIR health data implementation and assessment. Other kinds of health-related artefacts, as clinical guidelines, algorithms, software, models, etc. are out of scope.

The FAIR4FHIR IG addresses two kind of use cases:

- guide researchers who wish to make available under clearly stated terms of use, a FAIR health dataset derived from a data source collected and curated for use in a specific context.
- help researchers and data scientists to search and access previously collected FAIR health datasets to answer specific research questions.

Methods

Taking the FAIR4Health project as a starting point, at the MedInfo 2019 conference in Lyon, MIE 2020 conference, and EU-China 2020 conference, the 'FHIR4FAIR' workshops considered how the HL7 FHIR standard could support the FAIRification process in the case of health datasets, providing input to policy, standards, and research. These workshops, and others planned for the future, aim to build capacity and advance knowledge on how to use HL7 FHIR to implement the FAIR principles while reducing curation time and advancing interoperability and reusability of health datasets. This is very relevant for the researchers, allowing them to leverage not only reuse of health data for research and innovation, but also repeatability of research results..

So, the 'FAIRness for FHIR' project [11] was created, with the short-term aim of developing a new HL7 FHIR Implementation Guide (IG) called FHIR4FAIR [12]. Since January 2021, the 'FAIRness for FHIR' project has participated in the HL7 FHIR connect-a-thons and has advanced its agenda with weekly calls.

To achieve the defined goals, the IG is structured in two main parts: (i) An informative section providing general guidelines on the cooperative usage of FHIR and FAIR, focusing on best practices for satisfying the FAIR principles when publishing FHIR resources. (ii) A set of FHIR conformance resources and examples that provide, for selected case(s), practical examples of how FAIR principles can be realized and assessed with HL7 FHIR.

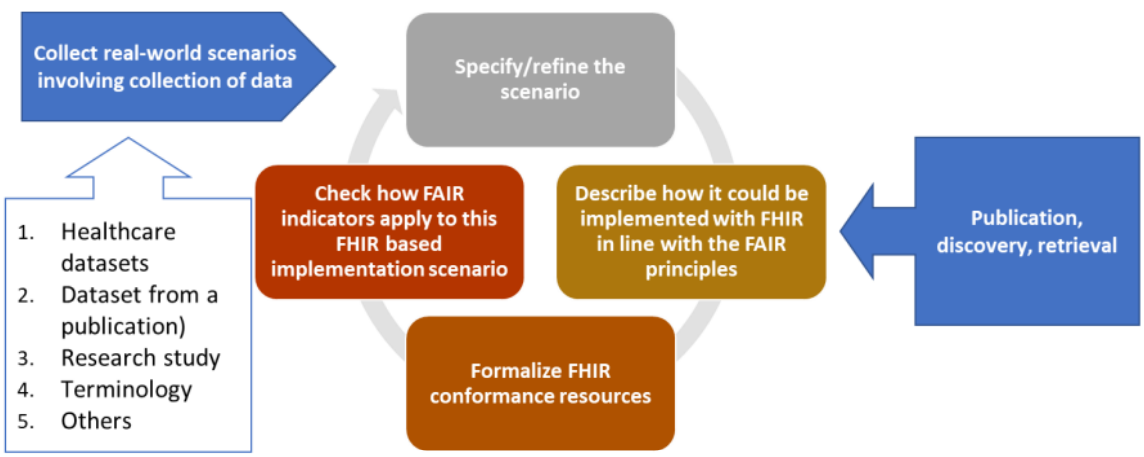

Figure 1– FAIRness for FHIR methodology

The FHIR4FAIR IG has been designed following an incremental, expert-based, iterative, and meet-in-the-middle approach and does not cover all the aspects related to the data FAIRification or all types of health data. Its aim is to promote reuse of existing artefacts, and avoid recreating the FHIR profiles have been already been specified by existing guides (e.g. genomic, lab results, vital sign, etc.).

The specific focus of the IG is to support researchers using FHIR to implement the FAIR principles in the case of health datasets. The design choices are therefore constrained by HL7 FHIR and it may be the case that some FAIR principles might not be fully accomplished. A list of possible limitations, including those related to the mapping of FAIR digital objects to HL7 FHIR, are summarized in Figure 1, which shows the incremental, iterative and meet-in-the-middle approach adopted.

A set of possible domains has been identified and a set of real-world scenarios were selected.

1. For each selected case scenario the following steps are followed (iteratively): The scenario has been specified and refined (e.g. creating several sub-scenarios, while considering different architectures) as needed.
2. A team of experts analyzes how the scenario can be realized using the FHIR standard resources and profiles.
3. FHIR conformance resources are selected, refined, or specified as needed.
4. The FAIR RDA indicators are used to examine the level of FAIRness of the revised dataset [7].

Multiple iterations of steps 1-4, allow to examine different deployment architectures and assess using RDA indicators how the adoption of FHIR resources or profiles impacts the FAIRness of health datasets.

The chosen methodology is currently tested prototypically on different scenarios:

1. **Healthcare datasets**. Data from EHRs or other kinds of healthcare repositories, for example, the healthcare datasets used in FAIR4Health project to validate the developed technological solution [13,14].
2. **Publications**. Data from publications of scientific articles accompanied by health datasets and citations to support the referenced research.
3. **Research studies**. Data from research studies, cohort studies, surveillance programs, and/or clinical trials, involving pseudo-anonymized data.
4. **Other scenarios**. Terminology related scenarios are considered to highlight structured health data information.

Results

Since January2021, the participants of the 'FAIRness for FHIR' project are performing weekly follow-up meetings allowing members of HL7 across the world to collaborate with the members of the FAIR4Health consortium. Following these principles, first draft of the FHIR4FAIR IG has been created [12].

The audience of the FHIR4FAIRIG includes:

- **Researchers**. People who generate, process or use FAIR health data for research.
- **Health Data Providers**. Healthcare providers that populate clinical data warehouses. Clinical study groups, registry operators, epidemiological cohorts' managers, surveillance or Public Health workers that which to publish FAIR health data sets or maintain FAIR repositories.
- **Data Stewards**. Data Stewards are responsible for metadata quality compliance and operationally implement strategic data governance requirements.
- **Technical Implementers.** Vendors of EHR systems, data repositories, and/or Electronic Data Capture (EDC) systems that provide primary data for FAIRification.
- **Government agencies.** Funding institutions that want to ensure the sustainable use and reuse of research project results. Legislators issue guidelines for the use of personal data (GDPR, Data Governance Act). Regulatory bodies provide best practice guidelines that support FAIR.
- **Public**. Citizens who want to use their data for their own purposes or donate data for research and need to provide the right metadata.

The different scenarios considered call for different FHIR resources. Resources considered are the research study, citation, and library resource. Meanwhile, additional effort is invested in considering the metadata applicable for the different cases studies.

The FHIR4FAIR IG is at the moment reviewed and will be formally presented to HL7 standards developers and users in HL7 September 2021 Ballot.

Discussion

The concept of a FAIR digital object is quite wide and can vary in terms of granularity and type of data that should be represented.

In fact, a FAIR digital object can be a single atomic information (e.g. a coded diagnosis) up to a collection of data (e.g. a dataset). Data, moreover, can represent quite different kinds of information: it might be for example a waveform, an image, a condition, a medication, or other kinds of data.

Such variations make the mapping between FAIR digital objects and the FHIR resource covering them complex (see Figure 2).

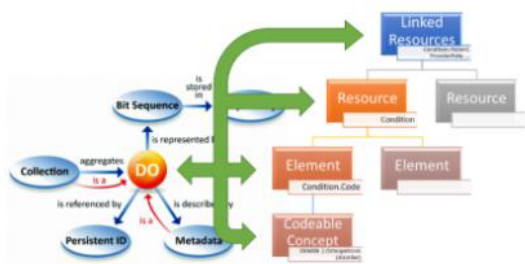

Figure 2– FAIRness for FHIR methodology

To cover the Findable and Reusable ('F' and 'R' from FAIR) characteristics, the data should be described with rich as possible, accessible, well-described, and machine-readable metadata, including the identifier of the data in a clear and explicit way, offering automatic discovery of datasets and services [1].

Similar considerations are also applicable to metadata, where, depending on what the metadata is, the metadata elements can be represented within the same resource documenting the data, or preferably according to the FAIR principles by a set of linked resources.

FAIR principles state that digital object describing the meta data of a health data set has to be distinct from the actual data referenced. This arrangement helps advance the "Findability" of health datasets that are protected within an institution. The metadata may reference the FHIR profiles used to access the actual data. This helps increase not only (A)ccess, but also (I)nteroperability and (R)reuse of the health data set. Sometimes the health data set is not allowed to leave the health institution that serves as data steward. In this case, a research question can be analyzed by "sending the algorithm to the health data set". In this case the importance of metadata is even higher as it provides the detailed context for responding to the research question and respecting GDPR provisions is paramount. Data elements can be analyzed only in part by the FAIR maturity indicators. In addition, the FAIRness for FHIR IG provides the basis for a certification roadmap for health data

sets focusing in part on all dimensions of FAIR, but most importantly on the use of HL7 FHIR in different scenarios.

GoFAIR and EOSC are engaged in the development of certification guidelines [16, 17]. GoFAIR is addressing FAIR related certification of organizations, people, processes, and products. The FAIR4Health FAIR data certification roadmap focusing on the technical implementation of the FAIR principles using FHIR. The FHIR4FAIR IG informs this effort (see Figure 2). The work also addresses the gaps identified by the EOSC report for the FAIR certification of data repositories.

Figure 3 – GoFAIR certification with HL7 FHIR4FAIR IG (adapted from [15]).

Conclusions

Providing guidance on how the HL7 FHIR standard could support the FAIRification process in the case of health datasets, it's a very relevant topic which could provide relevant contributions to the scientific community.

The 'FAIRness for FHIR' project is ongoing, with the short-term aim to present the FHIR4FAIR IG in HL7 September 2021 Ballot. This work serves to provide practical underpinnings for the FAIR4Health FAIRification workflow [15], which is a domain-specific extension of the GoFAIR process.

The team working in this project stimulates discussion on how HL7 FHIR standards and associated tools can simplify the FAIRification process, over different kinds of health data (healthcare datasets, publications, clinical trials and other kinds of research studies, etc.).

Acknowledgements

This work has been co-supported by the European Union through FAIR4Health project. This project has received funding from the European Union's Horizon 2020 Research and Innovation programme under grant agreement No 824666.

Also, this research has been co-supported by the Carlos III National Institute of Health, through the IMPaCT Data project (code IMP/00019), and through the Platform for Dynamization and Innovation of the Spanish National Health System industrial capacities and their effective transfer to the productive sector (code PT20/00088), both co-funded by European Regional Development Fund (FEDER) 'A way of making Europe'.

References

[1] M. D. Wilkinson, M. Dumontier, J. J. Aalbersberg, et al, The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data 3* (2016), article number: 160018.

[2] GoFAIR website, https://www.go-fair.org/.

[3] HL7 International, 'HL7 FHIR v4.0.1'. HL7 International, 30-Oct-2019.

[4] FAIR4Health website, https://www.fair4health.eu/.

[5] A. H. Nordo, H. P. Levaux, L. B. Becnel, J. Galvez, P. Rao, K. Stem, et al, Use of EHRs data for clinical research: Historical progress and current applications, *Learning Health Systems* 3(1) (2019), e10076.

[6] Research Data Alliance https://www.rd-alliance.org/about-rda

[7] C. Bahim, C. Casorrán-Amilburu, M. Dekkers, E. Herczog, N. Loozen, K. Repanas, et al, The FAIR Data Maturity Model: An Approach to Harmonise FAIR Assessments, *Data Science Journal*, 19(1) (2020).

[8] Force11. Guiding Principles for Findable, Accessible, Interoperable and Re-usable Data Publishing version b1. 0. (2014), available at: https://www.force11.org/fairprinciples.

[9] FAIR4Health Project. Report on the analysis of FAIR Facets, Principles and Dataports Vs HL7® FHIR®. (2019).

[10] G. Jiang, R. C. Kiefer, D. K. Sharma, E. Prud'hommeaux, H. R. Solbrig, A consensus-based approach for harmonizing the OHDSI common data model with HL7 FHIR, Studies in health technology and informatics 245 (2017), 887.

[11] 'FAIRness for FHIR' project website: https://confluence.hl7.org/pages/viewpage.action?pageId=91991234.

[12] FHIR4FAIR HL7 FHIR Implementation Guide website, under development: http://build.fhir.org/ig/HL7/fhir-forfair/.

[13] C. Alvarez-Romero, A. Martinez-García, C. L. Parra-Calderón, Tools for the support of workflow with data in Health Research to make them Findable, Accessible, Interoperable and Reusable, 2nd EU-China Health Summit 2020.

[14] C. Alvarez-Romero, A. Martinez-Garcia, M. Gencturk, A. A. Sinaci, C. L. Parra-Calderon, FAIR4Health FAIRification Tools, 16th International Digital Curation Conference.

[15] A. A. Sinaci, F. J. Núñez-Benjumea, M. Gencturk, M. L. Jauer, T. Deserno, C. Chronaki, G. Cangioli, C. Cavero-Barca, J. M. Rodríguez-Pérez, M. M. Pérez-Pérez, G. B. Laleci Erturkmen, T. Hernández-Pérez, E. Méndez-Rodríguez, C. L. Parra-Calderón, From Raw Data to FAIR Data: The FAIRification Workflow for

Health Research, *Methods of information in medicine* 59(S 01) (2020), e21–e32.

[16] https://www.go-fair.org/certification/

[17] EOSC Secreteriat. Recommendations on certifying services required to enable FAIR within EOSC, 2021 https://op.europa.eu/en/publication-detail/-/publication/70aa74b5-53bf-11eb-b59f-01aa75ed71a1/language-en

Address for correspondence

Corresponding author: Alicia Martínez-García.

Preferred method of contact: email.

Full mailing address: Virgen del Rocío University Hospital. 'Centro de Documentación Clínica Avanzada' building. Manuel Siurot av., s/n - 41013 Seville Spain.

Email: alicia.martinez.garcia@juntadeandalucia.es

Phone: +34 955 01 36 16