D6.2. Report on FAIR data certification roadmap

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

Document Information

GRANT AGREEMENT NUMBER 824666				ACRONYM: FAIR4Health
Full title	Improving Health Research in EU through FAIR Data			
Horizon 2020 Call	SwafS-04-2018: Encouraging the re-use of research data generated by publically funded research projects			
Type of action	Research and Innovation Action			
Start Date	1st December 2018	**Duration**	36 months	
Website	www.fair4health.eu			
Project Officer	Katharina BUSE			
Project Coordinator	Carlos Luis Parra Calderón, Andalusian Health Service (P1–SAS)			
Deliverable	D6.2. Report on FAIR data certification roadmap			
Work Package	WP6. Sustainability, Economic Impact and Digital Single Market			
Date of delivery	**Contractual**	M36	**Actual**	November 2021
Nature	Report	**Dissemination Level**	Public	
Lead beneficiary	P6-HL7			
Responsible Author	Catherine Chronaki			
Contributors	Alicia Martínez-García (P1-SAS), Matthias Loebe (P7-IMISE), Celia Alvarez (P1-SAS), Giorgio Cangioli (P6-HL7 Europe)			
e-mail	chronaki@gmail.com	**Phone number**	+306944594990	
Reviewers	A. Anil Sinaci (P15 SRDC), Nagarajan Ganapathy (P10 PLRI)			
Keywords	FAIR data certification roadmap, FAIR principles, health data, HL7 FHIR, health research performing organisations, European Open Science Cloud (EOSC)			

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

Document history

Version	Issue Date	Stage	Changes	Contributor(s)
0.1	21.10.2021	Draft	Table of Content	Alicia Martínez-García (P1-SAS)
0.2	29.10.2021	Draft	Contributions for the sections related to the RDA WG	Celia Alvarez (P1-SAS)
0.3	04.11.2021	Draft	Contributions related to metadata and FHIR4FAIR sections	Alicia Martínez-García (P1-SAS)
0.4	07.11.2021	Draft	Contributions related work and introduction	Matthias Löbe (P7-IMISE)
0.5	08.11.2021	Draft	Executive Summary, Introduction, Methodology, Conclusions	Catherine Chronaki (P6-HL7 Europe)
0.6	08.11.2021	Draft	Contributions related work	Giorgio Cangioli (P6-HL7 Europe)
0.7	10.11.2021	Draft	Update of Related Work	Matthias Löbe (P7-IMISE)
0.8	10.11.2021	Draft	Integrated changes from Alicia and Celia	Catherine Chronaki (P6-HL7 Europe)
0.9	15.11.2021	Draft	Incorporation of comments following review by A. Anil Sinaci (P15 SRDC), Nagarajan Ganapathy (P10 PLRI)	Catherine Chronaki (P6-HL7 Europe)
0.9a	16.11.2021	Draft	Updates to metadata section	Alicia Martínez-García (P1-SAS)
0.10	16.11.2021	Draft	Update of Related Work	Matthias Löbe (P7-IMISE)
1.0	18.11.2021	Prefinal	Updated Milestones, Formatting	Alicia Martínez-García (P1-SAS) Catherine Chronaki (P6-HL7 Europe)
1.1	18.11.2021	Final	Updated related work section	Matthias Löbe (P7-IMISE)
1.2	29.11.2021	Final	Updated of milestones expected in November 2021	Catherine Chronaki (P6-HL7 Europe) Alicia Martínez-García (P1-SAS)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

Table of Contents

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

List of figures

List of tables

List of acronyms

BoF	Birds of a Feather session
CEN	European Committee for Standardization
CTS	Core Trust Seal
EOSC	European Open Science Cloud
FAIR	Findable Accessible Interoperable Reusable
FHIR	Fast Healthcare Interoperability Resources
HL7	Health Level 7
HRPOs	Health Research Performing Organisations
IG	Implementation Guide
RDA	Research Data Alliance
TAB	Technical Advisory Board
TEHDAS	Joint Action Towards the European Health Data Space
WG	Working Group

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

1 Executive summary

The report on FAIR4Health data certification roadmap for health research institutions in the European Union envisions a time in the next 3-5 years (2022-2027) when health research institutions will be able to publish or give access through algorithms to certified FAIR health data sets.

Health research institutions offering certified FAIR health data sets need to:

1) Develop a FAIR data policy of the institution and supporting internal organizational processes following the relevant RDA guidelines
2) Establish an infrastructure similar to the FAIR4Health solution for the FAIRification of data sets and the publication of relevant FAIR metadata adhering to quality metrics
3) Adopt the HL7 FHIR 'FAIR for FHIR' Implementation Guide (IG) once available to enhance interoperability and reusability of data
4) Engage with European Open Science Cloud and cooperate with other FAIR initiatives
5) Train the administrative and research personnel on how to deliver and use FAIR data sets
6) Certify the processes adopted according to the ISO 9000 family of quality standards.

The development of this deliverable has been a long process that has followed the functional design of the FAIR4Health platform – FAIRification workflow (D2.2), the guidelines for implementing open FAIR data policy in health research (D2.3), as well as the FAIR4Health platform and agent implementation guideline (D3.1). The preparatory work of this deliverable has concentrated on the HL7 FHIR standard, while contacts with the European Committee for Standardization Technical Committee 251 (CEN TC/251) aim to inform developments towards standards for data quality.

With these activities as background, constant horizon scanning of relevant FAIR activities, and partnership with RDA, FAIR4Health worked relentlessly to advance the conditions for the creation of certifiable FAIR data sets. Since these activities are still in progress, we report intermediate milestones that will lead us to wide availability of certified FAIR health data in 3-5 years, a time when the European health data spaces[1] will be mature.

[1] https://ec.europa.eu/health/ehealth/dataspace_en

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

2 Introduction

The FAIR (Findable, Accessible, Interoperable, Reusable) principles have been established as best practice in the refinement of health datasets for open science, research, and innovation. Indeed, the introduction of the FAIR principles by Wilkinson et al. [1] is considered as the tipping point for open health science. However, the FAIR principles are in many ways general and abstract making implementation, let alone development of certification criteria a challenging task. Moreover, there are limited tools to help accelerate the FAIRification process.

FAIR4Health aims to encourage FAIRification and reuse of health data generated by publicly funded health projects and advance FAIR practices in health care accelerating and catalysing processes of institutional change. FAIR4Health has defined a FAIRification workflow to FAIRify health datasets within healthcare institutions, and has created, and now validates, a multicentre platform for the FAIRification and reuse of health data sets by Privacy Preserving Data Mining (PPDM) algorithms [2].

The Research Data Alliance (RDA, www.rda-alliance.org) is a global research community organization established in 2013 by the European Commission, the American National Science Foundation and National Institute of Standards and Technology, and the Australian Department of Innovation. The goal of the RDA is to build the social and technical infrastructure to enable open sharing and re-use of data. The RDA FAIR data maturity model WG developed the FAIR data maturity model including a set of qualified FAIR assessment indicators in 2020 [3].

FAIR4Health and the RDA FAIR Data Maturity WG initiated the HL7 FAIRness for FHIR IG project [4] to develop an implementation guide (IG) to explore how to deliver FAIR health data sets using HL7 FHIR[2]. Specifically, the IG aims to provide guidance on how to assess the FAIRness of health data sets using the RDA FAIR Data Maturity Model and to help researchers deliver FAIR health datasets using the HL7 FHIR standard. In the process of creating the IG, the FHIR for FAIR IG team has collected and analysed a large and diverse set of examples and use cases to populate the associated FHIR resources. These efforts illustrate how FAIRness can be implemented, assessed, and potentially improved using the HL7 FHIR standard.

At the same time, FAIR4Health initiated a working group in RDA to develop a guideline for health institutions that generate FAIR data sets or give access to algorithms that may use FAIR data sets for research purposes. This new RDA WG is under the umbrella of the RDA Health Data Interest Group.

Certification procedures and more importantly criteria for FAIR health data sets are essential for health researchers that use them. Frequently certification comes to be synonymous to trust. Thus, acquiring a certificate is generally not a small expense.

[2] http://hl7.org/fhir/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

Nevertheless, from the user's point of view, a number of advantages can result from certification procedures:

- Increasing the quality of research data
- Increasing the attractiveness for subsequent use by third parties
- Increasing the visibility and showing the legitimacy of one's own work through certificates or badges
- Fulfilment of requirements imposed by funders or scientific journals
- Improvement of internal workflows and increase of acceptance for the additional workload caused by FAIR among the staff members.
- Use of the auditor's expertise in the case of formal certifications

Most importantly, establishing such processes streamlines the participation of the organization in the digital economy. Despite the advantages outlined above, certifications in the field of FAIR data are currently not widespread. According to the experts from FAIR4Health, reasons for this may include the following:

- Lack of clarity about which certifications exist, how the certifications differ, and how to find the appropriate certification
- Uncertainty about the added value of a certification
- Lack of clarity about the personnel and other financial costs required
- Difficulty in quantifying the added value and communicating the value to management or third parties
- Uncertainty about the success of the certification process in principle
- Giving access to sensitive internal data and IP by third parties
- Existing local certifications that already cover certain parts of the envisaged certification.

For this purpose, in the process of developing this report, we have analysed the environment and worked on two supporting initiatives in RDA and in HL7, since one of the most frequent complaints is the lack of tools and the weakness of RDA indicators for interoperability.
The next sections present the Methodology followed (section 3), Related Work (Section 4), Information on datasets available through the FAIR4Health platform (section 5), RDA WG on guidelines (Section 6), and HL7 FHIR for FAIR IG (Section 7). The certification work of European Open Science Cloud (EOSC) is considered in the related work. Each section introduces their own milestones, which are then integrated in FAIR4Health data certification roadmap (Section 8).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

3 Methodology

The methodology we followed in developing this report has been comprehensive and systematic bringing together the RDA community with the international HL7 community and the FAIR4Health consortium.

1. Co-creation of the FAIRness for FHIR project scope statement.
2. Launch of the RDA WG on "Raising FAIRness in health data and Health Research Performing Organisations (HRPOs)" to propose the guidelines as RDA recommendation
3. Develop the HL7 FHIR for FAIR IG.
4. Engage in constant horizon scanning and partnership building to address harmonization of metadata. Data certification initiatives are presented along with tools for certification in the next section.
5. Analyse the process of generating metadata for health data sets in the FAIR4Health platform.
6. Share findings and knowledge with the global FAIR data community to accelerate adoption of the guidelines.
7. Collect the expected milestones for the contributing activities and present a roadmap for FAIR data certification in health research organizations following the ISO 9000 family of quality standards.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

4 Related work

4.1 FAIR Data Certification

The following section contains an analysis of existing certification efforts connected to the basic ideas of FAIR research data management. By way of introduction, three things are emphasized:

- **Relation to FAIR**: A number of approaches exist in the domain of research data archiving and sharing that do not always cover explicitly the term FAIR but do so intentionally. The reason for this is that they already existed before the coining of the term FAIR. This mainly concerns library approaches to the long-term archiving of primary research data with regards to data sharing.
- **Scope:** The FAIR data principles are intentionally rather abstract to be broadly applicable. They apply to data, metadata, and refer to:
    - (meta-)data in the sense of **directly accessible datasets**: databases, static snapshots, data collections from different parts and in different formats
    - (Meta-)data in the sense of **indirectly accessible data** from repositories or via application programming interfaces, where the target resources are software systems or infrastructures
    - (Meta-)data in the sense of **data models**, schemas, standards, where the focus is not on specific instance data, but on the suitability of the existing modelling constructs of the standard to achieve FAIRness
    - **Executable algorithms** or tools that assist users or machines in generating FAIR (meta-)data or provide an assessment of the current level of FAIRness achieved
    - **Proof of competence**: enabling scientists or data curators to support aspects of FAIR during the lifecycle of digital resources, e.g., by demonstrating specific knowledge or skills (training)
- **Terminology**: Certification means "assurance given by an independent certification body that a product, service or system meets the requirements of a standard[3]." In this section, the requirement of an independent body will be waived in order to more broadly address existing prior work. Furthermore, approaches that do not support explicit metrics for measuring FAIR will also be presented.

4.2 CoreTrustSeal (CTS)

- URL: https://www.coretrustseal.org/
- Target Audience:
    - Consortia and research data infrastructures
    - National libraries and archives
    - Technical service providers
- CoreTrustSeal: Non-profit foundation under Dutch law

[3] According to the International Organization for Standardization (ISO): https://www.iso.org/glossary.html

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

- Direct ancestors: **Data Seal of Approval (DSA)** [4] Certification of Trusted Data Repositories (2008-2018) developed by Data Archiving and Networked Services (DANS)[5] and **World Data System (WDS)**[6] Certification of Regular Members (2011-2018)

Relation to FAIR Data Certification:

The relationship to FAIR data certification is not explicit only indirect through the repositories:

- CoreTrustSeal covering a lot of similar aspects (PIDs, data discovery, access, licenses, provenance, reuse) from all of the 4 axes of FAIR.
- **Focus**: *basic certification* (level 1) standard for trusted digital repositories with regards to long-term preservation[7]. It envisions to be "the first step in a global framework for repository certification."
- United to RDA Repository Audit and Certification DSA–WDS Partnership WG
- Developed a catalogue of 16 guidelines or requirements[8], focusing on Organizational Infrastructure and Digital Object Management
- suitable for repositories that are committed to the curation and long-term preservation of the usability of said repository considering requirements and knowledge base of a "designated community"

Certification process

The characteristics of the CoreTrustSeal are as follows:

- Certification process starts with self-assessment of all requirements with the help of a guidance document[9] including 16 criteria (each one is given a compliance level ranging from "not applicable" to "completely fulfilled")
- self-assessment is then reviewed by two independent experts and discussed in the CoreTrustSeal board
- 5 levels of compliance possible (0 – Not applicable, 4 – The guideline has been fully implemented in the repository)
- Short answers with references to public documentation via the CoreTrustSeal Application Management Tool[10]
- Review of the self-assessment by two independent experts with final discussion in the CoreTrustSeal Board
- Cost: 3-year-certification for (as of time of writing) 1000€

[4] https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:116038

[5] https://dans.knaw.nl/en

[6] https://www.worlddatasystem.org/

[7] According to "Memorandum of Understanding to create a European Framework for Audit and Certification of Digital Repositories" http://www.trusteddigitalrepository.eu/Memorandum%20of%20Understanding.html

[8] https://zenodo.org/record/3638211

[9] https://zenodo.org/record/3632533

[10] https://amt.coretrustseal.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

- Current certified repositories[11]
    - WDS certified: 42 (declining)
    - DSA certified: 16 (declining)
    - CTS certified: 114 (increasing)
    - but hardly any repositories with focus on biomedical data

Table 1. How CoreTrustSeal supports FAIR.

F	**R13**. The repository enables users to discover the data and refer to them in a persistent way through proper citation.
	R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.
A	**R10**. The repository assumes responsibility for long-term preservation and manages this function in a planned and documented way.
	R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.
	R16. The technical infrastructure of the repository provides for protection of the facility and its data, products, services, and users.
I	**R11**. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.
	R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.
R	**R2**. The repository maintains all applicable licenses covering data access and use and monitors compliance.
	R7. The repository guarantees the integrity and authenticity of the data.
	R11. The repository has appropriate expertise to address technical data and metadata quality and ensures that sufficient information is available for end users to make quality-related evaluations.
	R15. The repository functions on well-supported operating systems and other core infrastructural software and is using hardware and software technologies appropriate to the services it provides to its Designated Community.

[11] https://www.coretrustseal.org/why-certification/certified-repositories/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

4.3 Nestor seal for trustworthy digital long-term archives

- URL: https://www.langzeitarchivierung.de/Webs/nestor/EN/Zertifizierung/zertifizierung_node.html
- Target Audience:
  - Long-term archives, especially technical libraries
  - University libraries, service provider, governmental archives, large reesrach organisations
- Nestor is primarily a German initiative based on DIN 31644:2012-04 "Information and documentation - Criteria for trustworthy digital archives"[12]
- DIN 31644 consists of 34 criteria[13]
- Nestor also includes organizational concerns (human resources, qualification, processes, responsibilities), legal concerns (legal regulations, contractual coverage) and financial concerns (ensuring sustainability)

Relation to FAIR Data Certification:

The relationship to FAIR data certification is implicit:

- nestor covering a lot of similar aspects (PIDs, data discovery, access, licenses, provenance, reuse) from all of the 4 axes of FAIR, e.g.
  - Interpretability (C24)
  - Identification (C27)
  - Metadata management (C28-C32)

Certification

- **Focus**: *Extended certification*[7] (level 2) standard for trusted digital repositories
  - Meant to be an addition to CoreTrustSeal
  - Focuses on "trustworthiness"
- Certification is starting with a self-audit of the 34 criteria
- Review of the self-audit and audit report by two independent reviewers from nestor
  - For every criterion, 0 to 10 points are awarded (0-no concept, 3-concept available, 6-complete concept, 10-concept also implemented in practice)
  - 10 points are expected for the core criteria C1 – C12
  - An average of 7 points must be achieved in the assessment of the applicable criteria C13 - C34
  - Sample assessment form[14]

[12] https://www.langzeitarchivierung.de/Webs/nestor/EN/Services/nestor_Siegel/nestor_siegel_node.html

[13] https://d-nb.info/1189191830/34 (in German), older Version https://d-nb.info/1047613859/34 in English

[14] http://files.dnb.de/nestor/zertifizierung/Einreichungsformular_EN.docx

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

    - Sample certification contract for Data Archiving and Networked Services (DANS)[15]
- Duration of review: about 4 months
- Cost for review is approximately 500 Euro
  Current certified repositories: 4[16]

4.4 ISO 16363 Audit and certification of trustworthy digital repositories

- URL: http://www.iso16363.org/
- Target Audience: large repositories for digital assets
- ISO 16363 (originally CCSDS 652-R-1[17]) is based upon the Trusted Digital Repositories and Audit Checklist (TRAC)[18] standard
    - In addition to ISO 16363, the working group wrote and submitted ISO 16919:2014" Requirements for Bodies providing Audit and Certification"
    - Which itself relies heavily on ISO/IEC 17021, a standard for auditing and certification of general types of management systems
- Current version is from 2012 (reviewed and confirmed in 2017)
- 50 main criteria, 109 in total
- 3 main parts
    - Organizational infrastructure
    - Handling of digital objects
    - Infrastructure and risk management

Relation to FAIR Data Certification:

- The commitment to the FAIR aspects is comparable to nestor, because ISO 16363 doesn't add specific constraints on digital assets

Certification process

- **Focus**: *Formal certification*[7] (level 3) standard for trusted digital repositories
    - Meant to be an addition to CoreTrustSeal and/or nestor-like certifications, not the one to start with
- Self-assessment template in Excel[19]
- Full external audit: Primary Trustworthy Digital Repository Authorisation Body (ISO-PTAB) is reviewing an application
    - For every criterion, 0 to 4 points are awarded (0-non complaint or missing, 4-fully complaint)

[15] https://www.langzeitarchivierung.de/Webs/nestor/SharedDocs/Downloads/DE/Zertifizierung/pruefberichtDANS.html?nn=182282

[16] https://www.langzeitarchivierung.de/Webs/nestor/DE/Zertifizierung/nestor_Siegel/nestor_siegel_node.html

[17] https://public.ccsds.org/Pubs/652x0m1.pdf#search=652

[18] https://www.crl.edu/sites/default/files/d6/attachments/pages/trac_0.pdf

[19] http://www.iso16363.org/?smd_process_download=1&download_id=30

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

    - Two-stages audit, see PTAB ISO 16363 Audit of the GPO Repository System (Public Report) page 5 for a detailed description of the workflow[20]
- Certification only allowed by organizations/persons certified according to ISO 16919 (Requirements for Bodies Providing Audit and Certification of Candidate Trustworthy Digital Repositories)
- Cost approximately 10,000 – 30,000 EUR[21] (unofficial estimation)
- Valid for 3 years
- As of July 2020, the United States Government Publishing Office is currently the only organization in the world to hold ISO 16363:2012 certification[22]

4.5 German Data Forum (RatSWD)

- URL**:** https://www.konsortswd.de/datenzentren/akkreditierung/
- Target Audience:
    - Research data centers (mostly permanently publicly financed)
    - Domain-specific for repositories covering social sciences, economics, education and health
- RatSWD: The "Rat für Sozial- und Wirtschaftsdaten" (literally: Council for Social and Economic Data) is an independent advisory board of the German Federal Government
- Existing since 2004
- Main tasks are strategic development of the research data infrastructure and representing the interests of data producers and data users
- RatSWD accreditation is the most popular certification in Germany

Relation to FAIR Data Certification:

The relationship to FAIR data certification is not explicit mentioned but similar goals are addressed:

- The main overlap is that data sharing and reuse are the primary drivers for the existing of the accreditation
- **Focus**: certification standard for digital repositories with social, behavioural, and economic data
- RatSWD accreditation adds technical constraints like providing at least one data access path and ensuring the long-term availability of the (meta-)data

Certification process

- Three mandatory criteria: access procedures, documentation of datasets, long-term availability

[20] https://www.fdlp.gov/file-repository/preservation/3910-the-ptab-iso-16363-audit-of-the-gpo-govinfo-repository-system-public-report
[21] https://datascience.nih.gov/trusted_digital_repository
[22] https://www.fdlp.gov/preservation/trusted-digital-repository-iso-16363-2012-audit-and-certification and https://www.govinfo.gov/features/only-tdr-certification

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

- Repositories must be fully operational and proof of a minimum of three external data users
- Additional criteria[23] relate to operational aspects such as proof of an operational concept, response times to data requests, proof of staff and infrastructure, and also a data quality assurance concept
- Certification process starts with self-assessment of all requirements
- The Committee for Data Access (FDI Committee examines the application and submits a recommendation
- Review is performed annually

  Current certified repositories: 40[24] (38 as of May 2020[25])

4.6 FAIRsFAIR – RDA FAIR Metrics

FAIRsFAIR has developed a framework for measuring FAIRness based on the work of the RDA FAIR Data Maturity Model WG, FAIR data maturity model specifications and guidelines[26]. It developed 17 minimum viable metrics to systematically measure the extent to which research data objects are FAIR and a tool to validate it (F-UJI[27], FAIR-Aware). A research data object may comprise data, metadata, and documentation (such as policies and procedures). These components influence the implementation of the FAIR assessment. For instance, they can either be resources to be evaluated or evidence of enabling FAIR.
https://www.fairsfair.eu/fairsfair-data-object-assessment-metrics-request-comments

4.7 FAIR-TLC: Metrics to Assess Value of Biomedical Digital Repositories

https://zenodo.org/record/203295#.YXKb1J5ByUI

4.8 FAIRsFAIR certification for repositories

In line with the "Turning FAIR into Reality" report from the European Commission expert group on FAIR data, FAIRsFAIR will augment existing certification mechanisms for digital data repositories, such as CoreTrustSeal. These established procedures and standards of the CoreTrustSeal requirements emerged from research data community work to identify key practices for data repositories which support long term access to reusable data.

Objectives:

- to develop extensions to existing approaches to "core" data repository certification with a view to identifying the levels of capability maturity necessary to support data which is assessed as FAIR.

[23] https://www.konsortswd.de/en/datacentres/accreditation/
[24] https://www.konsortswd.de/en/datacentres/all-datacentres/
[25] https://www.konsortswd.de/aktuelles/publikation/fdi-factsheet-oct20/
[26] https://www.rd-alliance.org/sites/default/files/FAIR%20Data%20Maturity%20Model.pdf
[27] https://www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

- to identify and develop a European network of trustworthy repositories enabling FAIR data.
- to offer support and guidance for data repositories aspiring repository certification, and for reviewers of certification applications.
- to provide stakeholders with a trustworthy tool to find relevant certified repositories.
- to define the scope of "core" level repository certification, and through pilots define more rigorous requirements to support the assessment of data in trustworthy repositories as FAIR

https://www.fairsfair.eu/fair-certification

4.9 GO FAIR

The three-point-FAIRification-Framework aims to guide FAIRification by maximizes reuse of existing resources, maximizes interoperability, and accelerates convergence on standards and technologies supporting FAIR data and services. The three points are as follows:

a) Metadata for Machines (M4M) Workshops: a community of practice that considers its domain-relevant metadata requirements and other policy considerations and formulates machine-actionable metadata components.
b) FAIR Implementation Profile (FIP) comprises re-usable metadata schemata produced in the M4M workshop.
c) FAIR Data Points (FDP) or FAIR Digital Objects (FDO) are configured based on the FAIR implementation guide and contribute to a global Internet of FAIR data and services

https://www.go-fair.org/how-to-go-fair/

Certification Process

GO FAIR has also developed a policy for certification for FAIR resources which is available here: https://www.gofairfoundation.org/certification/. This is part of a larger approach towards certification that aims to address people, organizations, and data sets in a holistic way.

The main elements of GO FAIR Foundation Pioneer program on certification using first generation criteria are co-created in an open, transparent process to produce certification criteria and evaluation mechanisms (certification schema) that has five stages:

a) Schema Development: define objective requirements and compliance tests that compose the GO FAIR Foundation certification schema.
b) Schema Review: independent third-parties review of the GO FAIR Foundation schema.
c) Schema Publication: GO FAIR Foundation publishes the schema [as FAIR machine-actionable resources].
d) Certification: Professional certification organizations can then adopt the GO FAIR Foundation schema to build and execute compliance tests (automated where possible) and grant certifications [as FAIR, machine-actionable resources].

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

e) Schema Maintenance: Regular, iterative review and update of the schema track technological advances

4.10 Research Data Alliance

Repository Audit and Certification Catalogues

4.11 ECRIN

- URL: https://ecrin.org/activities/data-centre-certification
- Target Audience: academic clinical study centers (non-commercial CROs)
- ECRIN: European Clinical Research Infrastructures Network
    - Main idea: Support for multinational clinical trials
    - Overcoming language barriers, finding local experts, assisting during administrative and regulatory processes
- 106 requirements in 16 domains (version 4.0[28])

Relation to FAIR Data Certification:

- ECRIN has a special role within the initiative mentioned here: it doesn't aims at arbitrary research repositories but at clinical trials centers, that means specialized organizations not only maintaining data, but also being part of the data collection process
- To relation to FAIR is weak in the sense that no technical formats or standards are described or assumed here
- on the other hand, organizational processes and quality assurance procedures for clinical research data are assumed here that also apply to other types of health data

Certification process

- Focus: data management and information technology as well as clinical trial management
- The ECRIN Datacenter Certification Program enables non-commercial clinical trial sites to have their processes for creating and maintaining GCP-compliant trial databases monitored and evaluated
- An application form can be requested
- Certification starts with self-assessment
- Review is done through an on-site audit of the facility's data management activities and the IT infrastructure used to support these activities
- Duration: 2-3 days, involve a team of three auditors
- If certification is successful, it will be granted for 4 years
- Currently 17 datacenters are certified[29]

[28] https://zenodo.org/record/1240941#.YZbJCmDMK3D
[29] https://ecrin.org/who-we-are/partners

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

4.12re3data

- URL: https://www.re3data.org/
- Target Audience: research data repositories – medium to small scale
    - Must be run by a legal entity
    - Must focus on research data
- re3data.org (Registry of Research Data Repositories) is an internationally recognized meta-repository for research data repositories
    - service was developed as part of two projects funded by the German Research Foundation (2012-2016)
    - main goal of the adjacent re3data COREF (Community Driven Open Reference for Research Data Repositories) project is to further professionalize re3data and to provide reliable and customizable descriptions of research data repositories
- Used by researchers and services worldwide for finding datasets
- Machine-readable XML schema representation[30]
- Based on DataCite metadata vocabulary[31]

Relation to FAIR Data Certification:

- Although FAIR is not explicitly mentioned, there is an extremely close proximity in terms of content, which is underlined not least by the use of machine-readable, community-consented vocabularies and a REST API

Certification process

- Assessment form for self completion32
    - 42 (complex) criteria as of schema version 3.1[33], some multiple
- Application is reviewed by two members
- No cost for the submitter
- Provides no certificate, but a number of badges that characterize important features and add a citeable identifier (DOI)
- Focus: trustworthy repositories
- Currently lists more than 2600 repositories for research data[34]

4.13 Tools to assess FAIRness

In our efforts to define a roadmap for FAIR data certification for Health research organization, the existence of tools that may accelerated the availability of FAIR datasets is

[30] https://www.re3data.org/schema
[31] https://datacite.org/
[32] https://www.re3data.org/suggest
[33] https://gfzpublic.gfz-potsdam.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_5007395
[34] https://coref.project.re3data.org/blog/using_the_re3data_api

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

important. Below we present an exemplary list of tools that have been developed for this purpose and may be deployed in a health organization as part of the FAIR data policy:

1. 5 Star Data Rating Tool
2. Data Stewardship Wizard
3. FAIR Data Self-Assessment Tool
4. FAIR Evaluation Services
5. FAIR enough?
6. FAIR metrics
7. FAIRdat
8. FAIRness self-assessment grids
9. FAIRshake
10. GARDIAN FAIR Metrics
11. FAIR maturity EVALUATOR

4.14 Other Relevant Initiatives

There are several other initiatives that contribute to the FAIR data policy landscape and have influenced our work:

- EOSC recommendations on certifying services required to enable FAIR https://op.europa.eu/en/publication-detail/-/publication/70aa74b5-53bf-11eb-b59f-01aa75ed71a1
- UNESCO Open Science recommendations https://codata.org/initiatives/data-policy/unesco-open-science-recommendation/
- FAIRPlus: The FAIR Cookbook: an open-source collection of instructions (or "recipes") for making life science data FAIR
- European Commission Horizon2020: Commission expert group on Turning FAIR data into reality https://ec.europa.eu/transparency/expert-groups-register/screen/index.cfm?do=groupDetail.groupDetail&groupID=3464
- European Framework for Audit and Certification DANS https://dans.knaw.nl/en/about/organisation-and-policy/certification?set_language=en

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

5 Metadata associated with the FAIR4Health solution

5.1 FAIR4Health Metadata

On April 2021, FAIR4Health consortium started a set of working meetings to define the metadata related to the whole FAIR4Health solution and use cases, the metadata related to the datasets (generated in the FAIRification process), and to decide how the consortium is going to publish these metadata for the scientific community's interest. Clearly, while the data itself is not public, the metadata about a data set used to answer a research question should be public.

The following subsections summarize these definitions about the specific metadata that FAIR4Health is generating, and the decisions regarding publication.

5.1.1 Metadata related to the FAIR4Health use cases

The welcome page of the FAIR4Health platform (Figure 1) shows a summary of the metadata that is part of the FAIR4Health solution, more specifically:

- Information related to the research studies currently covered, including:
  - A specific description for each research study, detailing the kind of study.
  - The number of sites (data providers) involved in each research study.
  - The total number of patients whose data have been used to generate models in each research study.
  - The different algorithms covered by each research study.
- The Common Data Model used by FAIR4Health solution.
- The open and free publication of the metadata related to datasets, generated carrying out the FAIRification process.

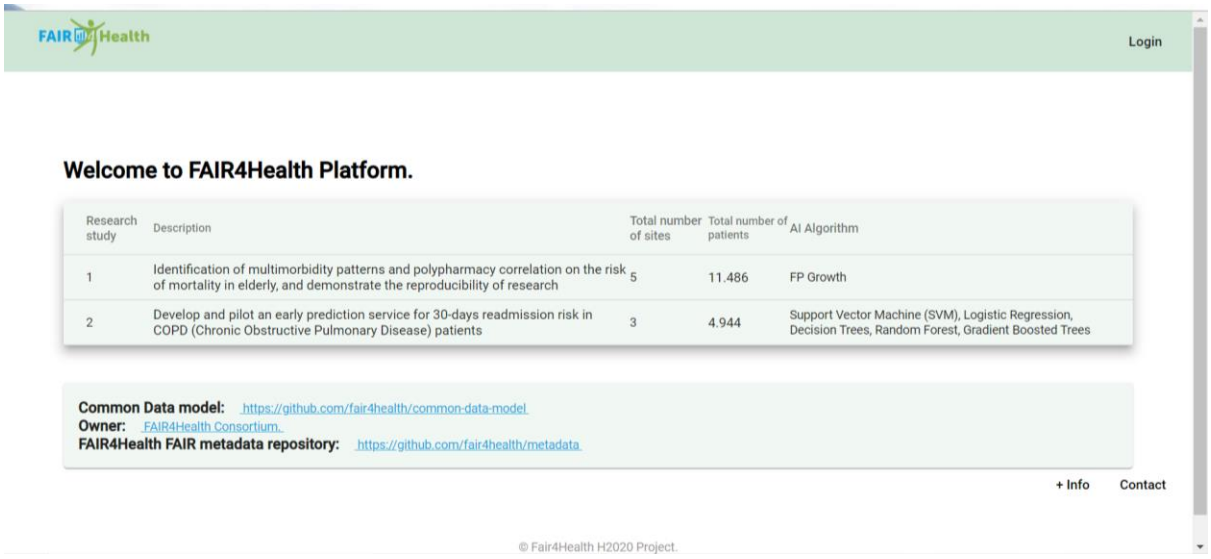

Figure 1. Welcome page of the FAIR4Health platform.

With this information, the scientific community can know the kind of research studies covered by the complete FAIR4Health solution.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

5.1.2 Metadata related to datasets

The FAIR4Health FAIRification process generate some metadata related to the FAIRified datasets, more specifically:

- Provenance information about the transformed data such as the author, ownership, timestamp and version of the FAIR data (following the format defined by the HL7 FHIR Resource Provenance).
- Minimal documentation about the transformed dataset, including the licence of use (following the format defined by the HL7 FHIR Resource DocumentManifest).
- Description of the capabilities offered by each site (following the format defined by the HL7 FHIR Resource CapabilityStatement).

The FAIR metadata related to the FAIRified datasets generated in the FAIRification process, was published in the FAIR4Health GitHub[35] in August 2021, so it is available to the scientific community.

One folder by each clinical partner participant in the FAIRification process has been created:

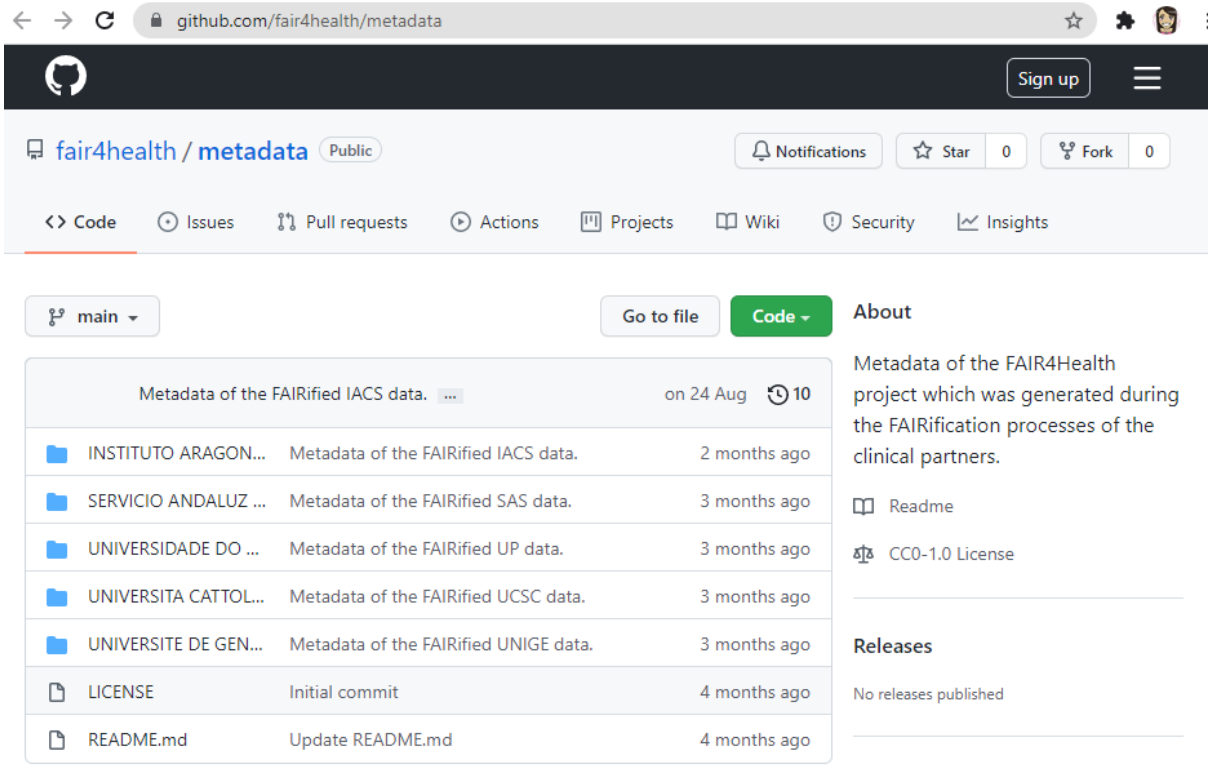

Figure 2. Metadata related to datasets, published in GitHub.

Each folder contains 3 JSON files: Provenance.json, DocumentManifest.json, and CapabilityStatement.json.

[35] FAIR4Health metadata related to datasets published in GitHub. https://github.com/fair4health/metadata

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

5.2 Next steps

The FAIR4Health consortium continues assessing the possibilities to open and free publish these metadata in the following public repositories:

- Digital CSIC[36]. On 9th November 2021, SAS team has uploaded SAS metadata in Digital CSIC, the publication is in progress under review by Digital CSIC reviewers. When SAS metadata will be finally published in this repository, SAS will ask the rest of clinical partners participants in the FAIRification process of retrospective health data, to have their agreement to upload their metadata too like in GitHub.
- FAIRsharing[37]. On 9th November 2021, SAS team has uploaded a request to upload metadata in FAIRsharing, the publication is in progress under review by FAIRsharing reviewers (Figure 3).
- Zenodo[38]. On 9th November 2021, SAS team has uploaded SAS metadata in Zenodo, the publication is in progress under review by Zenodo reviewers (Figure 4). When SAS metadata will be finally published in this repository, SAS will ask the rest of clinical partners participants in the FAIRification process of retrospective health data, to have their agreement to upload their metadata too like in GitHub.
- Translational Data Catalog[39]. The consortium has explored without successful the possibility to upload metadata in this portal. So, for having help in the process, SAS team on 10th November 2021 has contacted with the Translational Data Catalog's team using the contact information available in the website. On 12th November 2021, the Translational Data Catalogue's team answer us, they are helping us with the process.
- Leipzig Health Atlas[40]. The FAIR4Health team is currently investigating the possibility of depositing metadata in this repository including a semantic description. On 15th November 2021, this repository is under maintenance, so this work is going to be restarted as soon as the repository is running again.
- RDA Metadata Standards Catalog WG[41]. SAS team explored the possibility of uploaded the FAIR4Health metadata in any resource offered by this WG, but we concluded this WG doesn't offer a repository for metadata records.
- re3data[42]. The FAIR4Health team explored the possibility of uploaded the FAIR4Health metadata using the services offered by this initiative, but we concluded this initiative offers a database of research repositories, but doesn't offer a repository to upload the metadata records.

[36] https://digital.csic.es/
[37] https://beta.fairsharing.org/create
[38] https://zenodo.org/deposit/new
[39] https://datacatalog.elixir-luxembourg.org/
[40] https://www.health-atlas.de/
[41] https://www.rd-alliance.org/groups/metadata-standards-catalog-working-group.html
[42] https://www.re3data.org/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

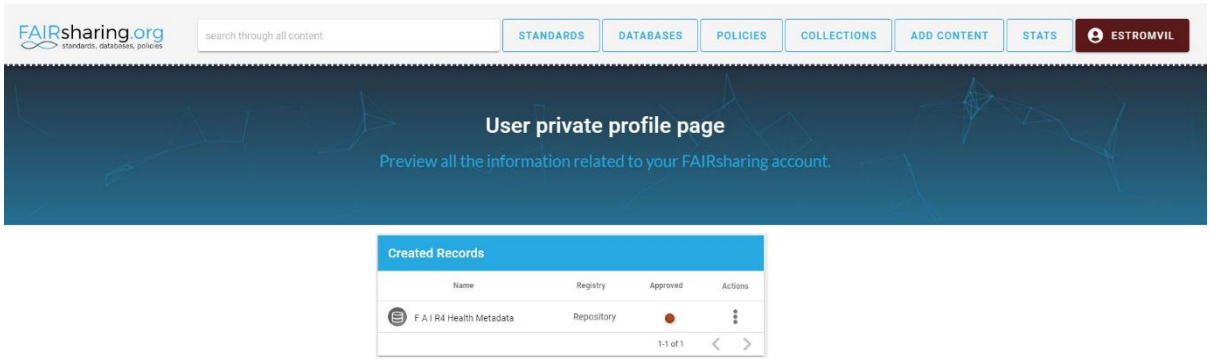

Figure 3. Metadata related to datasets, in review to be published in FAIRsharing.

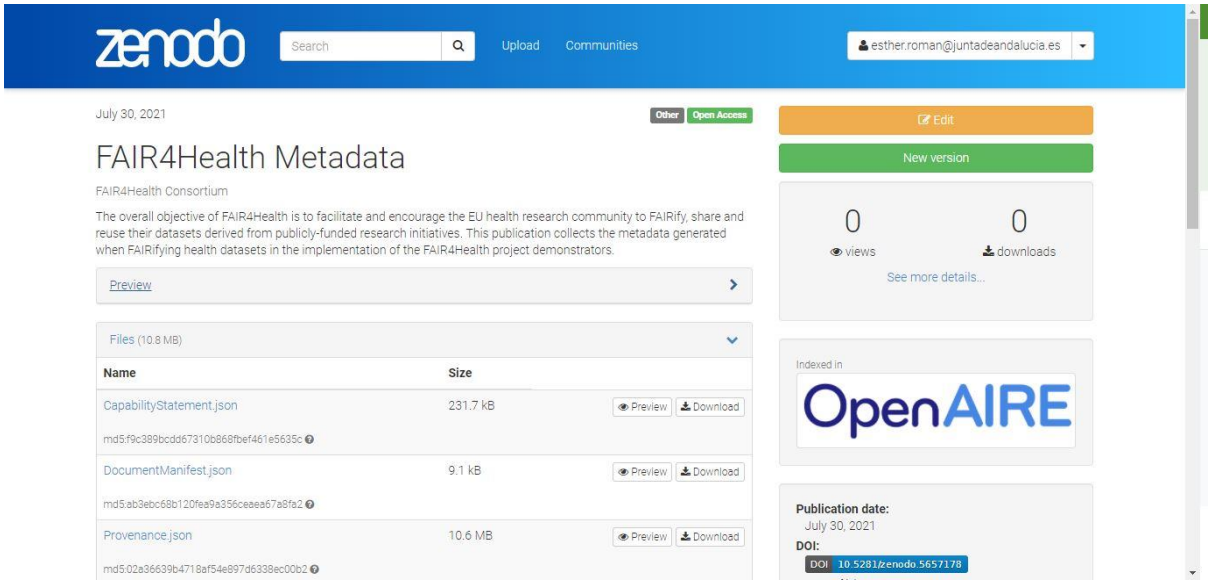

Figure 4. Metadata related to datasets, in review to be published in Zenodo.

The FAIR4Health project has contributed in the large picture of metadata initiatives, sharing the metadata related to the FAIR4Health solution itself in the welcome page of the FAIR4Health platform, and sharing the metadata related to datasets in GitHub.

The case study of FAIR4Health has been discussed in Joint Action Towards the European Health Data Space (TEHDAS) the joint action for the European Health Data Space focusing on secondary use of data. Clearly there is an issue related to what metadata should be published about a data set that cannot be accessed or downloaded but may be used for research.

We expect that in the next three years work in this area will continue and we shall see more work on the analysis of metadata published through a platform such as FAIR4Health.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

Table 2. Milestones for sharing metadata of data sets on the FAIR4Health platform.

Id	Milestone	Date
#1.1	FAIR4Health metadata published in Digital CSIC	December 2021*
#1.2	FAIR4Health metadata published in FAIRsharing	December 2021*
#1.3	FAIR4Health metadata published in Zenodo	December 2021*
#1.4	FAIR4Health metadata published in Translational Data Catalog	December 2021*
#1.5	FAIR4Health metadata published in Leipzig Health Atlas	December 2021*
#1.6	Metadata frameworks established for Health Care	Q3 2022
#1.7	Tools developed for automatic publishing of metadata	Q4 2022

** Depends on each metadata repository availability and time needed to cover the process*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

6 RDA WG on Raising FAIRness in HRPOs

6.1 Introduction and background

Adoption and application of the FAIR principles to research data has grown significantly in the few years since their first publication. The FAIR principles have been embraced in many quarters and have paved the way to creating a level playing field for data reuse. Nevertheless, they also pose challenges in some areas of research where reuse may not be an automatic right due to issues of confidentiality, privacy, commercial interests and sensitivities in general.

One sector of research where this is particularly evident is health research, and FAIR4Health proposed a new RDA working group aligned with the Health Data Interest Group[43] aiming to address the global disparities in uptake of the FAIR principles in health research and within health research performing organisations (HRPOs).

Starting from FAIR4Health deliverable D2.3. "Guidelines for implementing FAIR open data policy in health research" which identified differences in policies across selected European regions, some FAIR4Health partners initiated on the creation of an RDA WG to address the need to accelerate the adoption of FAIR practices in health research organizations. The aim was to expand the work done in FAIR4Health to create a global analysis of policies. Then, drawing on commonalities and best practices, a set of guidelines (RDA outputs or recommendations) are to be developed. HRPOs can employ these guidelines in their local contexts to address FAIRification of research data.

Concretely, the FAIR4Health consortium initiated the RDA WG on Raising FAIRness in health data and HRPOs[44] building upon previous birds of Feather (BoFs) in RDA Plenaries 13 (April 2019, Philadelphia), 14 (October 2019, Helsinki) and 15 (April 2020, Virtual). The necessity for creating useful and straightforward guidelines for HRPOs by establishing the initial approach of the principles and steps to be included in a global HRPO policy for FAIR data were discussed.

In the beginning of June 2020, the case statement prepared for this WG was submitted for review by the RDA Technical Advisory Board (TAB) and finally this WG was endorsed by the RDA Council in August 2020. In RDA Plenary 16 (November 2020, Virtual), the launch of this WG was formalized, and the attendees performed a review in depth of the principles and steps to guide HRPOs to create a policy (internationally valid) for FAIR data.

Finally, another formal working group session at RDA Plenary 17 (April 2021, Virtual) was held to discuss the global landscape analysis of FAIR adoption which was initiated by the Chairs of this WG. In addition, at RDA Plenary 17, Alicia Martínez-García who is member of

[43] https://www.rd-alliance.org/groups/health-data.html

[44] https://www.rd-alliance.org/groups/raising-fairness-health-data-and-health-research-performing-organisations-hrpos

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

the 'FAIRness for FHIR' HL7 project participated to introduce the work done related to the 'FHIR4FAIR' HL7 FHIR IG, aiming to establish synergies between both working groups.

On the other hand, the Chairs of this WG (S. Venkataraman, Celia Alvarez-Romero, Kristan Kang and Anupama Gururaj) meet monthly to follow up and include external input at separate monthly teleconferences.

6.2 Current status

WG sessions were previously held in which the WG Chairs solicited feedback on the WG activities from the broader community. Additionally, this WG has regularly scheduled monthly meetings (both with the WG members and the Chairs' meetings) which enabled the Global Landscape Analysis that is the focus of this WG to be conducted and completed that is a goal of this WG, collecting legal and ethical considerations and national and international guidelines and policies.

In addition, a survey has been designed to collect information on the national and international policies and requirements for health research data sharing (including FAIR data) mandated by various countries and research funders and it is planned for dissemination during RDA Plenary 18 (November 2021, Virtual).

Likewise, work is also underway on a document to compile the information gathered through the Landscape Analysis to assess the impact of such constraints or considerations, which will also be presented at RDA Plenary 18 as well.

Furthermore, the objective of the sessions planned for RDA Plenary 18 is mainly to demonstrate the ongoing tasks that were outlined in the WG case statement. Specifically, the impact assessment of FAIR adoption which were initiated by the Chairs and required input and review by the WG members and the RDA community were discussed in these sessions.

6.3 Next planned steps

A significant progress has been achieved by the RDA WG on Raising FAIRness in health data and HRPOs (see details in section 6.1 and 6.2 of this document). This WG expects to present the draft guidelines at RDA Plenary 19 (June 2022) that is currently planned to be in person meeting, allowing for better participation that we have had in the last virtual plenaries due to the pandemic. However, the WG Chairs have requested an extension of this WG end date and postponement of the deadline of submission of the results of this WG, because the WG still requires some time to complete the remaining tasks identified. The request for an extension of the WG timeframe must be discussed and approved by the RDA TAB. By that date, it is planned to deliver the final output as a written set of principles to be adopted by the HRPO community, taking care that they are not considered as rules to be followed depending on the regional and local contexts.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

We expect that once available, this guideline will be used by the health research organizations to cultivate trust and carry out the reorganization necessary to support FAIR and create FAIR data sets.

Table 3. Milestones for generating the guideline of the RDA WG on Raising FAIRness.

Id	Milestone	Date
#2.1	RDA draft recommendations delivered by the RDA WG on Raising FAIRness in health data and HRPOs	June 2022

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

7 FAIRness for FHIR HL7 project

7.1 Groundwork for the FHIR for FAIR FHIR implementation guide.

Taking the FAIR4Health project as a starting point, on August 2019 in MedInfo conference in Lyon, on July 2020 in MIE conference, on October 2020 in EU-China conference, and on May 2021 in MIE 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 to leverage not only the reuse of health data for research and innovation, but also the repeatability of research results.

So, the 'FAIRness for FHIR' HL7 project was created[45], with the short-term aim of developing a new HL7 FHIR Implementation Guide (IG) called FHIR for FAIR. Since January 2021, the 'FAIRness for FHIR' project has advanced its agenda with weekly calls.

Other related events with 'FAIRness for FHIR' participation:

- January 2021. FAIR track in the 26th HL7 FHIR connect-a-thon[46]
- April 2021. Poster titled 'FAIRness for FHIR project: Making Health Datasets FAIR using HL7 FHIR' at RDA Plenary 17[47]
- September 2021. Paper titled 'FAIRness for HL7 FHIR: supporting interoperability of health datasets' at the Computing in Cardiology 2021 conference[48]
- October 2021. Paper titled 'FAIRness for FHIR: Towards Making Health Datasets FAIR using HL7 FHIR' at the MedInfo 2021 conference[49]

7.2 Current status

The goals of HL7 FAIRness for FHIR project under the HL7 Service Oriented Architecture (SOA) WG is threefold: (a) facilitate the collaboration between the FAIR and the FHIR communities (b) enable a cooperative usage of the FHIR and FAIR paradigms. (c) support the health data FAIRness assessment by using HL7 FHIR conformance resources.

[45] https://confluence.hl7.org/display/SOA/HL7+FHIR+FHIR4FAIR+IG+PSS
46 FAIR track in the 26th HL7 FHIR connect-a-thon. https://confluence.hl7.org/display/FHIR/2021-01+FAIR
47 RDA Plenary 17 website.https://www.rd-alliance.org/rdas-17th-plenary-meeting-programme
48 FAIRness for HL7 FHIR: supporting interoperability of health datasets' at the Computing in Cardiology 2021 conference.https://www.cinc.org/2021/Program/accepted/336_Preprint.pdf
49 FAIRness for FHIR: Towards Making Health Datasets FAIR using HL7 FHIR' at the MedInfo 2021 conference. https://imia-medinfo.org/medinfo21/information-and-knowledge-management-track-information-standards-theme/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

The objective is to Develop a HL7 FHIR Implementation guide - called FHIR4FAIR FHIR IG - providing guidance on how HL7 FHIR can be used for supporting FAIR health data implementation and assessment.

The project timeline is available in

Figure 5.

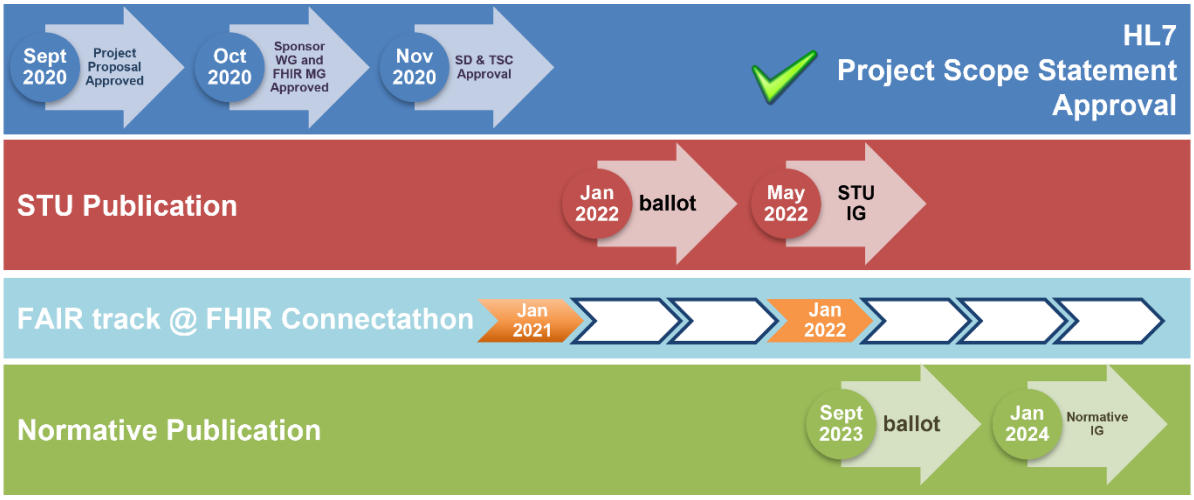

Figure 5. Timeline of the FAIRness for FHIR Project.

Work develops in weekly calls where everyone may participate. In this way FAIR4Health has initiated a level playing field so that HL7 FHIR is used in a consistent way to advance FAIRness of health data sets, particularly with respect to interoperability and privacy.

The draft implementation guide is available at: https://build.fhir.org/ig/HL7/fhir-for-fair/

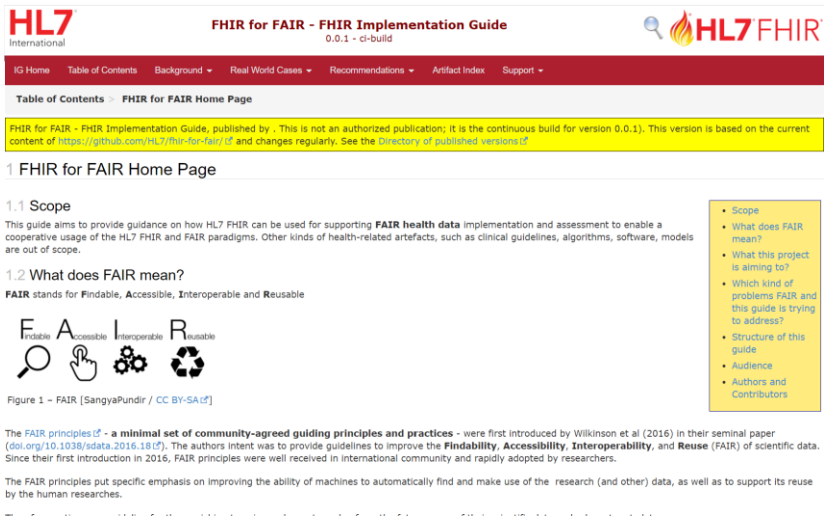

Figure 6. FHIR for FAIR FHIR Implementation Guide.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

The guide is based on FHIR R4b to take advantage of the new capabilities offered by resources as the Citation, keeping in any case the compliance with already existing implementations based on FHIR R4 (as this project). It has been designed following an incremental, iterative and meet-in-the -middle approach, without pretending to cover in this version of the guide all the aspects related to the data FAIRification.

It provides some background information about FAIR and FHIR; it describes some of the challenges that the FHIR implementation of the FAIR principles need to tackle (e.g., metadata / data mapping, use of GUPRIs (*Globally Unique, Persistent and Resolvable Identifiers*), etc).

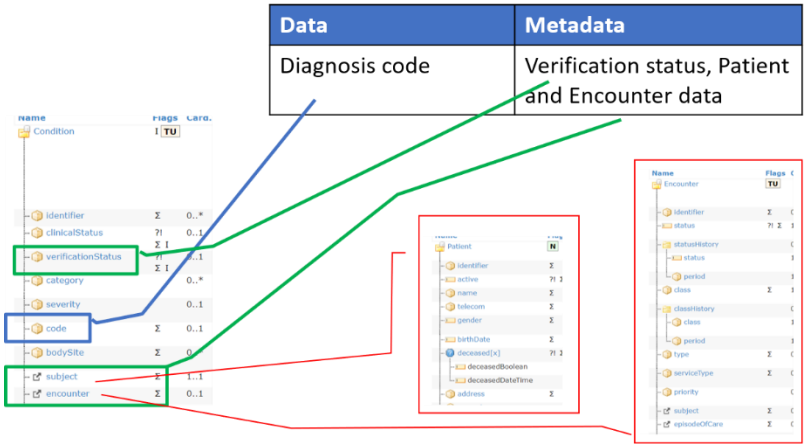

Figure 7. Mapping FAIR data and metadata in FHIR.

A set of selected real-world scenarios used as case study, including FAIR4Health, have been then summarised, highlighting challenges, lessons learned and describing how the adoption of HL7 FHIR may help in improving FAIRNess (based on RDA indicators assessment).

The core part of the guide is however the 'Recommendations', that constitutes of three main parts:

1. General recommendations (https://build.fhir.org/ig/HL7/fhir-for-fair/recommendations.html)
2. A page describing how to approach FAIR principles for HL7 FHIR based implementations (https://build.fhir.org/ig/HL7/fhir-for-fair/FHIRandFAIR.html)
3. A page describing how to interpret the RDA indicators in the HL7 FHIR space (https://build.fhir.org/ig/HL7/fhir-for-fair/FHIRandRDAMetrics.html)
4. Plus, a page about terminologies (https://build.fhir.org/ig/HL7/fhir-for-fair/terminology.html).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

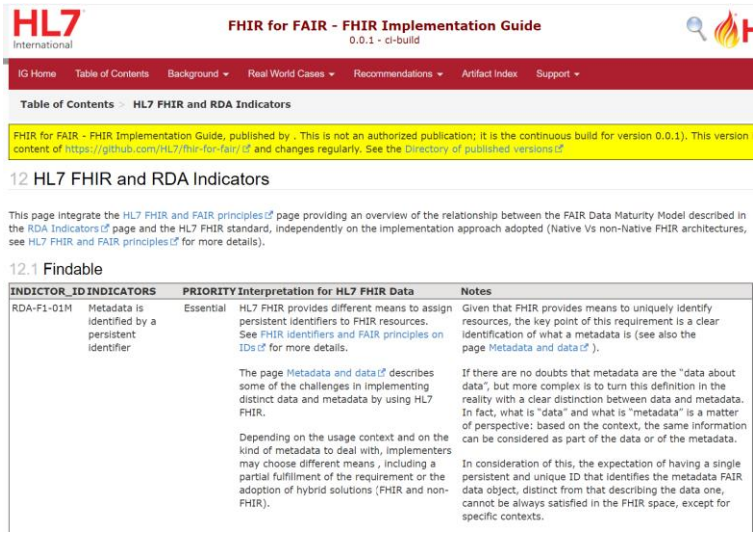

Figure 8. FHIR for FAIR FHIR Implementation Guide: HL7 FHIR and RDA Indicators.

Finally, the guide includes a series of FHIR profiles, extensions and examples representing some study level metadata information and help the implementation of the FAIR principles.

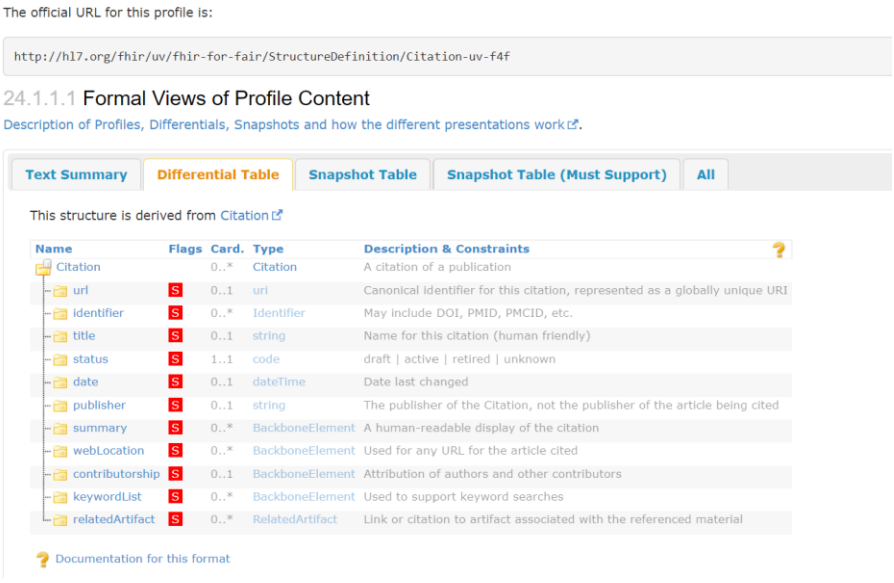

Figure 9. FHIR for FAIR FHIR Implementation Guide: profile defined on the Citation resource.

7.3 Next planned steps

The HL7 FHIR for FAIR implementation guide is up for ballot in January 2022. However, work will continue after January to further elaborate different cases, receive feedback from its use and further elaborate.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

Table 4. Milestones for the HL7 FHIR for FAIR IG.

Id	Milestone	Date
#3.1	January ballot of FHIR for FAIR IG	January 2022
#3.2	FHIR for FAIR STU Publication	March 2022[50]
#3.3	Second version of FHIR for FAIR IG (STU 2)	September 2022
#3.4	Third version of FHIR for FAIR IG (Normative Ballot)	September 2023

[50] The actual publication date depends on the final publication of FHIR R4b.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

8 Towards a FAIR data Certification roadmap

Health research institutions offering certified FAIR health data sets need to:

1) Develop the FAIR data policy of the institution and supporting internal organizational processes following the relevant RDA guidelines
2) Establish infrastructure similar to the FAIR4Health platform for the FAIRification of data sets and the publication of relevant FAIR metadata
3) Adopt the FAIR4FHIR IG once available
4) Engage with European open science cloud and cooperate with other FAIR initiatives
5) Train the administrative and research personnel on how to deliver FAIR
6) Certify the processes adopted according to the ISO9000 family of quality standards.

Still the FAIR data ecosystem is immature to have a clear view of the FAIR data certification prospects ahead. The efforts of the FAIR4Health consortium to accelerate this maturation process will continue after the end of the project and have generated some milestones for the year ahead. We expect more initiatives to blossom so that by the end of 2024, the first certified Health research organizations appear.

In addition, we would like to highlight that throughout the FAIR4Health project and in line with the analyses carried out in the RDA WG on Raising FAIRness in health data and HRPOs, we have identified the importance of implementing a FAIR strategy in health research, including training and skills development to facilitate researchers' compliance with FAIR principles.

Table 5. FAIR Data certification Roadmap.

Id	Milestone	Date
#1.1	FAIR4Health metadata published in Digital CSIC	December 2021*
#1.2	FAIR4Health metadata published in FAIRsharing	December 2021*
#1.3	FAIR4Health metadata published in Zenodo	December 2021*
#1.4	FAIR4Health metadata published in Translational Data Catalog	December 2021*
#1.5	FAIR4Health metadata published in Leipzig Health Atlas	December 2021*
#1.6	Metadata frameworks established for Health Care	September 2022
#1.7	Tools developed for automatic publishing of metadata	December 2022
#2.1	RDA draft recommendations delivered by the RDA WG on Raising FAIRness in health data and HRPOs	June 2022

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

#3.1	January ballot of FAIRness for FHIR IG	January 2022
#3.2	January HL7 Connectathon	January 2022
#3.3	Second version of FAIRness for FHIR IG	September 2022
#3.4	Third version of FAIRness for FHIR IG	September 2023
#4.1	First Health research organizations get certified	June 2024

** Depends on each metadata repository availability and time needed to cover the process*

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

9 Conclusions

This report on a roadmap for FAIR data certification has drawn its finding from 4 main sources: (1) landscape analysis on certification and FAIR data (2) analysis of the metadata for use cases on the FAIR4Health platform. (3) Experienced advisors (4) key findings of the FAIRness of HL7 FHIR platform.

The FAIRness for FHIR activity will continue after the end of the FAIR4Health project and the same is true for the relevant RDA activities.

Our expectation is that by mid 2024 the first health research organizations will be certified according to the ISO9000 family of standards with robust procedures for their operation related to the generation or availability of FAIR data sets.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

10 References

[1] Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

[2] Sinaci A, Núñez-Benjumea Gencturk F, et al., From Raw Data to FAIR Data: The FAIRification Workflow for Health Research, Meth. Inf. in Medicine 59(S01) (2020), e21–e32.

[3] Research Data Alliance. FAIR Data Maturity Model: specification and guidelines. DOI: 10.15497/RDA0050

[4] FAIRness for FHIR IG URL: confluence.hl7.org/ pages/viewpage.action?pageId=91991234.

[5] FAIR4Health deliverable D2.3. Guidelines for implementing FAIR open data policy in health research. https://osf.io/3u7dt/

[6] RDA WG on Raising FAIRness in health data and health research performing organisations (HRPOs). https://www.rd-alliance.org/groups/raising-fairness-health-data-and-health-research-performing-organisations-hrpos

[7] 'FHIR for FAIR' HL7 FHIR Implementation Guide. https://build.fhir.org/ig/HL7/fhir-for-fair

[8] RDA Plenary 19. https://www.rd-alliance.org/plenaries/rda-19th-plenary-meeting-part-international-data-week-20%E2%80%9323-june-2022-seoul-south-korea

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

Appendix A: HL7 FAIRness for FHIR project scope statement (PSS)

The HL7 FAIRness for FHIR PSS is available here: https://confluence.hl7.org/display/SOA/HL7+FHIR+FHIR4FAIR+IG+PSS

1a. Project Name	FAIRness for FHIR
1b. Project ID	1651
1c. Is Your Project an Investigative Project (aka PSS-Lite)?	No
1d. Is your Project Artifact being Reaffirmed or proceeding to Normative directly after being either Informative or STU?	No
1e. Today's Date	
1f. Name of standard being reaffirmed	
1g. Project Artifact Information	
1h. ISO/IEC Standard to Adopt	
1i. Does the standard include excerpted text from one or more ISO, IEC or ISO/IEC standards, but is not an identical or modified adoption?	
1j. Unit of Measure	
2a. Primary/Sponsor WG	Service Oriented Architecture
2d. Project Facilitator	Giorgio Cangioli
2e. Other Interested Parties (and roles)	BR&R would like to be kept informed of progress, particularly to ensure the relevance of FAIR principles in

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

	Clinical Research are taken account of. CDS wishes to be an Interested Party, given that we are interested in applying the same FAIR principles to the creation, management, and access of knowledge artifacts. FHIR-I would like to be kept informed of progresses. ArB would like to be kept informed of progresses.
2f. Modeling Facilitator	
2g. Publishing Facilitator	Giorgio Cangioli
2h. Vocabulary Facilitator	
2i. Domain Expert Representative	Oya Beyan (Fraunhofer, RDA DE; EOSC FAIR WG); Carlos Luis Parra Calderón (EOSC; FAIR4Health coordinator; Andalusian Health); Catherine Chronaki (HL7 Europe); ALICIA MARTINEZ GARCIA (Andalusian Health Service); Anthony Juehne (RDA Reproducible Health Services WG); Matthias Löbe (IMISE Leipzig)
2j. Business Requirements Analyst	
2k. Conformance Facilitator	Giorgio Cangioli
2l. Other Facilitators	
2m. Implementers	1) FAIR4Health Project 2) SRDC 3) ATOS
3a. Project Scope	Develop a FHIR4FAIR FHIR Implementation Guide aiming to (FAIR Principles https://www.go-fair.org/fair-principles): 1) Identify how HL7 FHIR standard fulfills data FAIRness maturity indicators a) analyse relationship between FAIR data object conceptual components (e.g. data, metadata, provenance, identifiers) and HL7 FHIR resources b) analyze how RDA FAIR Maturity indicators are supported by specific HL7 FHIR resources c) analyze how RDA Reproducible Health Data Services recommendations can be supported by HL7 FHIR d) identify a minimum set of metadata to be fulfilled for

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

	specific sets of RDA FAIRness maturity indicators extended of health-related research data sets. e) provide examples of best practices from EOSC (European Open Science Cloud https://www.eosc-portal.eu/) or NIH (https://www.nih.gov/) 2) Suggest an assessment methodology/checklist, exploring machine readable and manual assessment methods. This guide should contain a large informative part addressing the above mentioned points and a set of FHIR conformance resources and examples that provide, for selected case(s), a practical example of how FAIRness can be realized and assessed by using HL7 FHIR. This project is intended to be the result of an active collaboration between RDA and HL7, the plan is to consult the Health care communities to comment and endorse the resulting HL7 FHIR IG taking into account relevant RDA group input.
Attachments	
3b. Project Need	Provide guidance on supporting FAIRness by using HL7 and specifically: - the FAIR (Findability, Accessibility, Interoperability and Reusability) principles (https://www.go-fair.org/fair-principles/) - the proposed RDA (Research Data Alliance) - https://www.rd-alliance.org/ -recommendations on FAIR data maturity model, providing indicators for assessing adherence to the FAIR principles. - the recommendations of RDA Reproducible Health Services Working Group for describing, documenting, and sharing metadata for health data curation workflows
3c. Security Risk	Unknown
3d. External Drivers	
3e. Objectives/Deliverables and Target Dates	Connect-a-thon [Jan 2021] Connect-a-thon [May 2021] STU ballot [May 2021] Publication as STU IG (2 years) [Oct 2021]

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

	Normative ballot [May 2023] Publication as Normative IG [Oct 2023]
3f. Common Names / Keywords / Aliases:	FHIR4FAIR; FAIR;
3g. Lineage	
3h. Project Dependencies	
3i. HL7-Managed Project Document Repository URL:	.
3j. Backwards Compatibility	No
3k. Additional Backwards Compatibility Information (if applicable)	
3l. Using Current V3 Data Types?	N/A
3l. Reason for not using current V3 data types?	
3m. External Vocabularies	Unknown
3n. List of Vocabularies	
3o. Earliest prior release and/or version to which the compatibility applies	
4a. Products	FHIR Implementation Guide
4b. For FHIR IGs and FHIR Profiles, what product version(s) will the profiles apply to?	4.0.1
4c. FHIR Profiles Version	
4d. Please define your New Product Definition	
4d. Please define your New Product Family	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

5a. Project Intent	Implementation Guide (IG) will be created/modified
5a. White Paper Type	
5a. Is the project adopting/endorsing an externally developed IG?	No
5a. Externally developed IG is to be (select one)	
5a. Specify external organization	RDA
5a. Revising Current Standard Info	
5b. Project Ballot Type	STU to Normative
5c. Additional Ballot Info	
5d. Joint Copyright	No
5e. I understand I must submit a Joint Copyright Letter of Agreement to the TSC in order for the PSS to receive TSC approval.	no
6a. External Project Collaboration	RDA
6b. Content Already Developed	0%
6c. Content externally developed?	No
6d. List Developers of Externally Developed Content	
6e. Is this a hosted (externally funded) project?	No
6f. Stakeholders	Clinical and Public Health Laboratories, Quality Reporting Agencies, Regulatory Agency, Standards Development Organizations (SDOs)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

6f. Other Stakeholders	
6g. Vendors	EHR, PHR, Health Care IT
6g. Other Vendors	
6h. Providers	Clinical and Public Health Laboratories, Healthcare Institutions (hospitals, long term care, home care, mental health)
6h. Other Providers	
6i. Realm	Universal
7d. US Realm Approval Date	
7a. Management Group(s) to Review PSS	FHIR
7b. Sponsoring WG Approval Date	Sep 22, 2020
7c. Co-Sponsor Approval Date	
7c. Co-Sponsor 2 Approval Date	
7c. Co-Sponsor 3 Approval Date	
7c. Co-Sponsor 4 Approval Date	
7c. Co-Sponsor 5 Approval Date	
7c. Co-Sponsor 6 Approval Date	
7c. Co-Sponsor 7 Approval Date	
7c. Co-Sponsor 8 Approval Date	
7c. Co-Sponsor 9 Approval Date	
7c. Co-Sponsor 10 Approval Date	

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666

7e. CDA MG Approval Date	
7f. FMG Approval Date	Sep 30, 2020
7g. V2 MG Approval Date	
7h. Architecture Review Board Approval Date	
7i. Steering Division Approval Date	Nov 08, 2020
7j. TSC Approval Date	Nov 16, 2020

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666