Findable Accessible Interoperable Reusable (FAIR) Principles for FHIR

Catherine Chronaki
HL7 Europe Foundation
chronaki@HL7europe.org

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824666
What do data scientists do?

and where standards can help..

https://www.youtube.com/watch?v=X3paOmcTjQ
History of FAIR data

- In 2016, the ‘FAIR Guiding Principles for scientific data management and stewardship’ were published in *Scientific Data*.
- The authors intended to provide guidelines to improve the Findability, Accessibility, Interoperability, and Reuse of digital assets.
- The principles reflect the capacity of computational systems to find, access, interoperate, and reuse data with none or minimal human intervention.
- Big data and data science accent this imperative because of the increase in volume, complexity, and creation speed of data.
A framework guiding FAIRification

FAIRification process begins when a community of practice considers its domain-relevant metadata requirements and policy considerations, and formulates machine-actionable metadata components.

- **Metadata for Machines** (M4M) Workshops.
- **FAIR Implementation Profile** (FIP).
- Configuration of FAIR infrastructure with **FAIR Data Points** (FDP) or FAIR Digital Objects (FDO) in the global Internet of FAIR Data and Services.
- FAIR principles refer to three types of entities: data (or any digital object), metadata (information about that digital object), and infrastructure.

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

The first step in (re)using data is to find them. Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the **FAIRification process**.

- **F1.** (Meta)data are assigned a globally unique and persistent identifier
- **F2.** Data are described with rich metadata (defined by R1 below)
- **F3.** Metadata clearly and explicitly include the identifier of the data they describe
- **F4.** (Meta)data are registered or indexed in a searchable resource
Findable: **F1. (Meta)data are assigned a globally unique and persistent identifier**

- The most important principle of FAIR, key to the vision of Open Science
- Enables data integration
- Identifiers consist of an internet link and remove ambiguity in the meaning of your published data
  - Assigns identifiers to metadata elements
  - To concepts/measurements of the dataset
- Some data repositories automatically generate globally unique and persistent ids for the deposited datasets
  - can be obtained by registries services
  - Example: OIDs in HL7

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No. 824666
Findable: **F2. Data are described with rich metadata**

- Compliance with F2 allows
  - Generous and extensive metadata including information on context, quality and condition or characteristics
- F2 allows to locate data and increase re-use and citations
- Do not presume that you know who will use your data and for what purpose
- Types of metadata: intrinsic and contextual
- Intrinsic: captured automatically by machines
- User-controlled or Contextual: details about the data, e.g. devices used, units of captured data, protocol document, concepts linked to data, etc.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824666
Findable: **F3. Metadata clearly and explicitly include the identifier of the data they describe**

- This principle instructs to separate metadata from the dataset.
- This can be achieved by mentioning the gid of the data set in the metadata.
Findable: **F4. (Meta)data are registered or indexed in a searchable resource**

- Identifiers and rich metadata descriptions alone cannot ensure “findability”
- Availability of a dataset, service, repository should be made discoverable through indexing
- Example:
  - Metadata of FAIR datasets can be published on FAIR Data points
  - Registries of FAIR datasets
Once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation.

- **A1. (Meta)data are retrievable by their identifier using a standardised communications protocol**
- **A1.1 The protocol is open, free, and universally implementable**
- **A1.2 The protocol allows for an authentication and authorisation procedure, where necessary**
- **A2. Metadata are accessible, even when the data are no longer available**
Accessible: **A1. (Meta)data are retrievable by their identifier using a standardised communications protocol**

- This principle states that FAIR data retrieval should be mediated without specialized or proprietary tools or communication methods.
- This principle focuses on how data and metadata can be retrieved from their globally unique and persistent identifiers.
- Examples:
  - http or ftp protocol.
Accessible: **A1.1 The protocol is open, free, and universally implementable**

- To maximize data reuse, the protocol should be free (no-cost) and open-sourced and thus globally implementable to facilitate data retrieval.
- This criterion influences choices about the repository where you share your data.
- Examples:
  - HTTP, FTP, SMTP are acceptable,
  - Skype or Microsoft exchange are not, as they are proprietary.
Accessible: A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

- Accessible is sometimes misunderstood. It does not necessarily mean ‘open’ or ‘free’.

- This principle implies that the conditions under which the data are accessible should be clear – even private data or data for sale can be FAIR.

- Accessibility is specified in such a way that a machine can automatically execute the requirements.

- This criterion will also affect your choice of the repository where you will share your data.

- Example: HMAC, HTTPS, FTPS
**Accessible:** **A2. Metadata are accessible, even when the data are no longer available**

- Datasets tend to degrade or disappear over time for different reasons
  - Because of the cost of maintaining online data sets or resources
- Links become invalid and users search data that are not there.
- So, this principle states that metadata should persist even after data can no longer be sustained.
- This is related to registration and indexing issues (F4).
Interoperable

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

- **I1.** (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- **I2.** (Meta)data use vocabularies that follow FAIR principles
- **I3.** (Meta)data include qualified references to other (meta)data
Interoperable: 1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

- Humans should be able to exchange and interpret each other’s data
- This principle applies also to machines without specialized or ad hoc algorithms, translators, or mappings
- For this to happen and ensure automatic findability and interoperability:
  - Use commonly used controlled vocabularies, ontologies, and thesauri
  - Good data model and well defined framework to describe and structure metadata
- Example: RDF, OWL, JSON LD, etc.
Interoperable: **I2. (Meta)data use vocabularies that follow FAIR principles**

- The controlled vocabulary used to describe datasets needs to be documented and resolvable using gids.
- This documentation needs to be findable and accessible by those using the dataset.
Interoperable: **I3. (Meta)data include qualified references to other (meta)data**

- A qualified reference is a cross-reference that explains its intent
- The objective is to create meaningful links among metadata
- Example: one dataset builds on another, additional data sets are needed, complementary information is a different dataset
  - FAIR data point
The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

- **R1. (Meta)data are richly described with a plurality of accurate and relevant attributes**
- **R1.1. (Meta)data are released with a clear and accessible data usage license**
- **R1.2. (Meta)data are associated with detailed provenance**
- **R1.3. (Meta)data meet domain-relevant community standards**
Reusable: R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

- It is easier to find and reuse data if there are many labels are attached to the data (related to F2 principles)
- This principle focuses on the ability of a human or machine user to decide if the data is useful in a particular context
- So, the publisher should provide metadata that allows discovery and describes the context under which the data was generated.
  - The publisher should not attempt to predict the data consumer’s identity and needs
- Example: scope of the data; particularities and limitations; dates; raw/processed data; variable names self-explanatory

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824666
Reusable: **R1.1. (Meta)data are released with a clear and accessible data usage license**

- This principle is about legal interoperability
- It implies:
  - Provide clarity on license status for reuse of data
  - Clearly articulate of license restrictions
- The conditions under which the data can be reused should be clear to human and machine users.
- Example: creative commons license linked to data set
Reusable: **R1.2. (Meta)data are associated with detailed provenance**

- Users that wish to reuse a data set should know where the data comes from
  - How to cite and acknowledge the origin
- It is recommended that users include description of the workflow that led to the data set
  - Who generated or collected the dataset?
  - How it was processed if at all?
  - Has it been published before?
  - Does it contain data from another source?
Reusable: R1.3. (Meta)data meet domain-relevant community standards

- Data available in a widely accepted standard are easier to reuse
- For example, include a well-established file format, template or common vocabulary.
- Best practices can help promote common standards
Uptake of the FAIR Data Principles?

- After improved data stewardship was shown as a result of the implementation of the FAIR data principles, adherence to the FAIR data principles is now frequently expected by researchers, publishers, funding agencies and policy makers such as the Health Research Board in Ireland.

FAIR as a Journey: Maturity Models

FAIR data requires a change: technical and cultural

- understand the level of FAIRness: “as-is”
- a roadmap for improvement: “target”
- benchmarking against the standards

Assessment tools provides:

- a sequence of maturity levels
- a set of indicators for measuring the maturation levels
Core Criteria to Normalise Assessment:

- list of indicators for assessing adherence to the FAIR principles
- various evaluation methods can be developed
- common understanding of the indicators, the maturity levels and the prioritisation
RDA Indicators - Overview

The full list of 42 indicators and report can be found on the group page.

Indicators for Findability

- [F1-01M] Metadata is identified by a persistent identifier
- [F1-01D] Data is identified by a persistent identifier
- [F1-02M] Metadata is identified by a universally unique identifier
- [F1-02D] Data is identified by a universally unique identifier
- [F2-01M] Sufficient metadata is provided to allow discovery, following domain/discipline-specific metadata standard
- [F2-02M] Metadata is provided for the discovery-related elements defined by the RDA Metadata IG, as much as possible and relevant, if no domain/discipline-specific metadata standard is available
- [F3-01M] Metadata includes the identifier for the data
- [F4-01M] Metadata is offered/published/exposed in such a way that it can be harvested and indexed
Example Indicator

**RDA-F1-02M Metadata is identified by a globally unique identifier**

<table>
<thead>
<tr>
<th>Principle to which the indicator relates</th>
<th>This indicator is linked to the following principle: <em>F1 (meta)data are assigned a globally unique and eternally persistent identifier.</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Description of the indicator RDA-F1-02M</td>
<td>The indicator serves to evaluate whether the identifier of the metadata is globally unique, i.e. that there are no two identical identifiers that identify different <a href="#">metadata records</a>.</td>
</tr>
<tr>
<td>Assessment details</td>
<td>Global uniqueness of identifiers should be evaluated based on a description of how identifiers are assigned. Such a description should make it clear that the mechanism for assigning identifiers cannot possibly assign the same identifier to different resources, or assign an identifier that has already been assigned via some other mechanism/organisation. A possible way to evaluate this indicator is to verify that the identifier used for the data is listed in a registry service like the RDA-endorsed FAIRsharing.</td>
</tr>
</tbody>
</table>
Prioritization

- Mandatory: indicator MUST be satisfied for FAIRness (Essential)
- Recommended: indicator SHOULD be satisfied, if at all possible (Important)
- Optional: indicator MAY be satisfied, but not necessarily so (Useful)
FAIR4Health evaluated if and how the HL7 FHIR standard can support the data FAIRification;

highlighting the relationships between FAIR data object conceptual components and HL7 FHIR

assessing the Force11 Facets and Principles and the FAIRPorts requirements as described by Force11 community;

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824666
HL7 FHIR4FAIR IG: Goals

Identify how HL7 FHIR standard fulfils data FAIRness maturity indicators:

- Analyze relationship between FAIR data object conceptual components (e.g. data, metadata, provenance, identifiers) and HL7 FHIR resources
- Analyze how RDA FAIR Maturity indicators are supported by specific HL7 FHIR resources
- Analyze how RDA Reproducible Health Data Services recommendations can be supported by HL7 FHIR
- Identify a minimum set of metadata to be fulfilled for specific sets of RDA FAIRness maturity indicators extended for health-related research data sets.
- Provide examples of best practices from EOSC (European Open Science Cloud) or NIH (National Institutes of Health).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824666
HL7 FHIR4FAIR IG: Objectives

- Facilitate the collaboration between the FAIR and the FHIR communities
- Enable a cooperative usage of the FHIR and FAIR paradigms, by providing cross-models mapping.
- Support the health data FAIRness assessment by using FHIR conformance resources.
Informative guidance

Explaining for example how
• FAIR principles can be supported by using HL7 FHIR;
• Conceptual components of the FAIR data objects maps into FHIR;
• Data FAIRness maturity indicators may be assessed by using FHIR;

“True” FHIR IG

Including a set of FHIR conformance resources and examples that provide, for selected case(s), a practical example of how FAIRness could be assessed.
To facilitate and encourage the EU Health Research community to FAIRify, share and reuse their datasets derived from publicly funded research initiatives through the demonstration of the potential impact that such strategy will have on health outcomes and health research.

**FAIR4Health project**

**SO 1.** To design and implement an effective outreach strategy at EU level

**SO 2.** To produce a set of guidelines to set the foundations for a FAIR data certification roadmap

**SO 3.** To develop and validate an intuitive, user-centered FAIR4Health platform and FAIR4Health agents

**SO 4.** To demonstrate the potential impact in health research and health outcomes through the validation of 2 pathfinder case studies

Coordinated by **Virgen del Rocío University Hospital, Andalusian Health Service (SAS)**

17 partners from 11 EU and non-EU countries
6 health research organisations
2 universities experts in data management
4 academic partners
5 business actors

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 824666
Join the FAIR4Health Open Community!

On behalf of FAIR4Health project, we would like to invite you to register as FAIR4Health Open Community Member.

In this section into FAIR4Health website: https://www.fair4health.eu/en/membership, you will find the Membership definition, a button to register as a new member, and a button to check the current list of members.

Two standalone, desktop applications for
- Data curation and validation
  - Video tutorial
- Data de-identification and anonymization
  - Video tutorial

Can run on Windows, Linux, MacOS built on ELECTRON, requires only NPM/Node.js

Contribute and be part of the FAIR4Health development
- https://github.com/fair4health

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