HMA
Heads of Medicines Agencies

EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

FAIR Catalogue sustainability

Technical workshop on real-world metadata for regulatory purposes
Virtual meeting, April 12, 2021

Presented by Prof. Miriam Sturkenboom
University Medical Center Utrecht

Background & requirements

2016: ‘FAIR Guiding Principles for scientific data management and stewardship’ were published in *Scientific Data*.

Guidelines to **F**indability, **A**ccessibility, **I**nteroperability, and **R**euse of digital assets.

Principles emphasize machine-actionability (i.e., the capacity of computational systems to find, access, interoperate, and reuse data with no or minimal human intervention)

Background & requirements

Findable: The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers.

- Data should have an identifier, rich metadata, and be indexed in a searchable resource

Accessible: Once the user finds the required data, she or he needs to know how can they be accessed.

- Metadata should be retrievable by their identifier

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.

- (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

Reusable: 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.

Source: https://www.go-fair.org/fair-principles/

What is sustained in the ecosystem of RWE?

EU-funded initiatives for real world evidence: descriptive analysis of their characteristics and relevance for regulatory decision-making

Kelly Plueschke [1], Patricia McGettigan [1, 2], Alexandra Pacurariu [1, 3], Xavier Kurz [1], Alison Cave [1]
Correspondence to Kelly Plueschke; kelly.plueschke@ema.europa.eu

Abstract

Introduction A review of European Union (EU)-funded initiatives linked to 'Real World Evidence' (RWE) was performed to determine whether their outputs could be used for the generation of real-world data able to support the European Medicines Agency (EMA)'s regulatory decision-making on medicines.

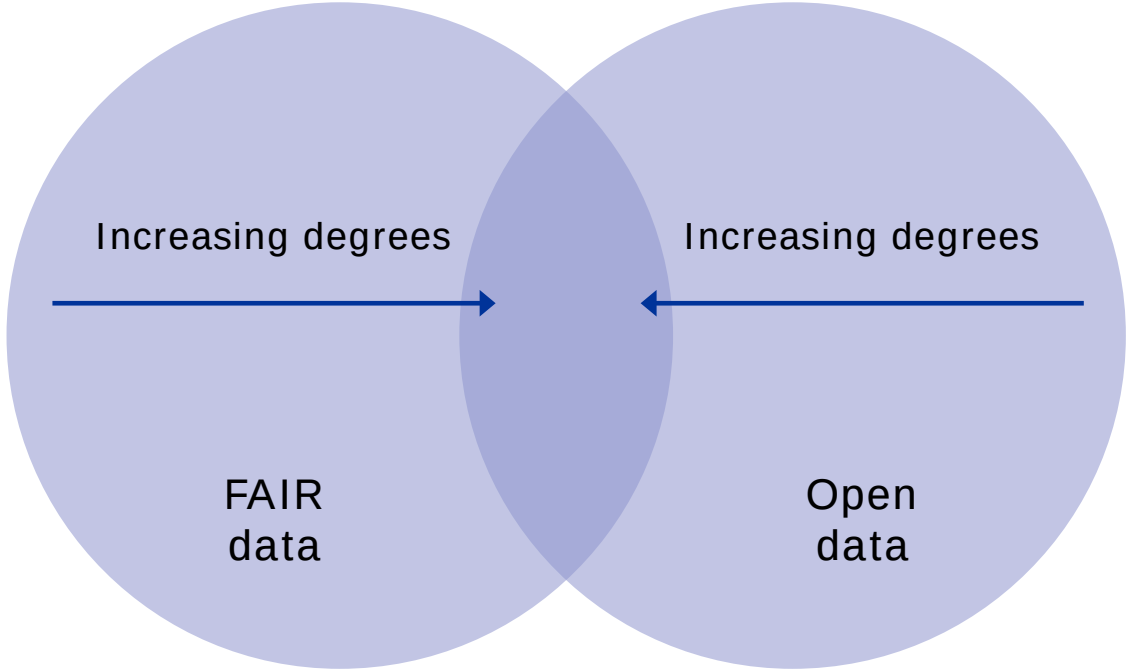
Conclusions This landscape of EU-funded initiatives linked to RWE which started before 31 December 2016 highlighted that the immediate utilisation of their outputs to support regulatory decision-making is limited, often due to insufficient available information and to discrepancies between outputs and objectives. Furthermore, the restricted sustainability of the initiatives impacts on their downstream utility. Multiple projects focussing on the same therapeutic areas increase the likelihood of duplication of both efforts and resources. These issues contribute to gaps in generating RWE for medicines and diminish returns on the public funds invested.

 Source: https://pubmed.ncbi.nlm.nih.gov/29903798/

FAIR is not always Open

FAIR should be applied broadly to all objects (including metadata, identifiers, software, and data management plans) that are essential to the practice of research and should inform metrics relating directly to these objects.

Figure 4. The relationship between FAIR and Open



5 Source: https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_0.pdf

Current sustainability RWE catalogues: Bridge to Data (private)

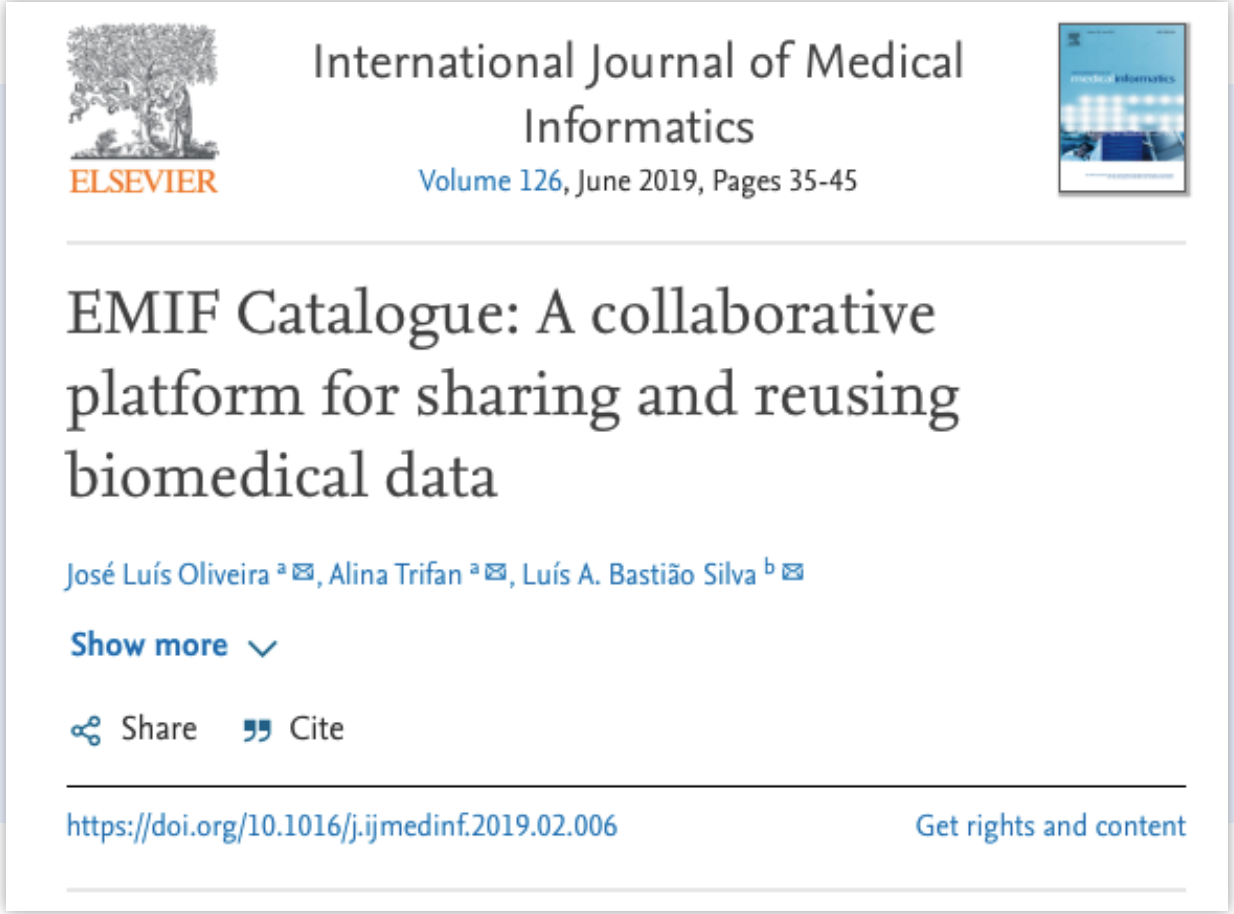
Key database features

- Types of databases included:
  - Longitudinal EMRs and claims
  - Drug- or disease-specific cohorts
  - Registries
  - National surveys
  - National surveillance systems
  - Spontaneous reporting systems
  - Tissue/blood
  - Genomic/pharmacogenetic
- > 330 standardized database profiles
- 135 defined data fields
- Profiles from 42 countries
- Regularly updated

Source: https://www.bridgetodata.org

Current sustainability RWE catalogues: IMI-EMIF Catalogue

Not updated anymore after project ended



ELSEVIER

International Journal of Medical Informatics

Volume 126, June 2019, Pages 35-45

EMIF Catalogue: A collaborative platform for sharing and reusing biomedical data

José Luís Oliveira [a] ✉, Alina Trifan [a] ✉, Luís A. Bastião Silva [b] ✉

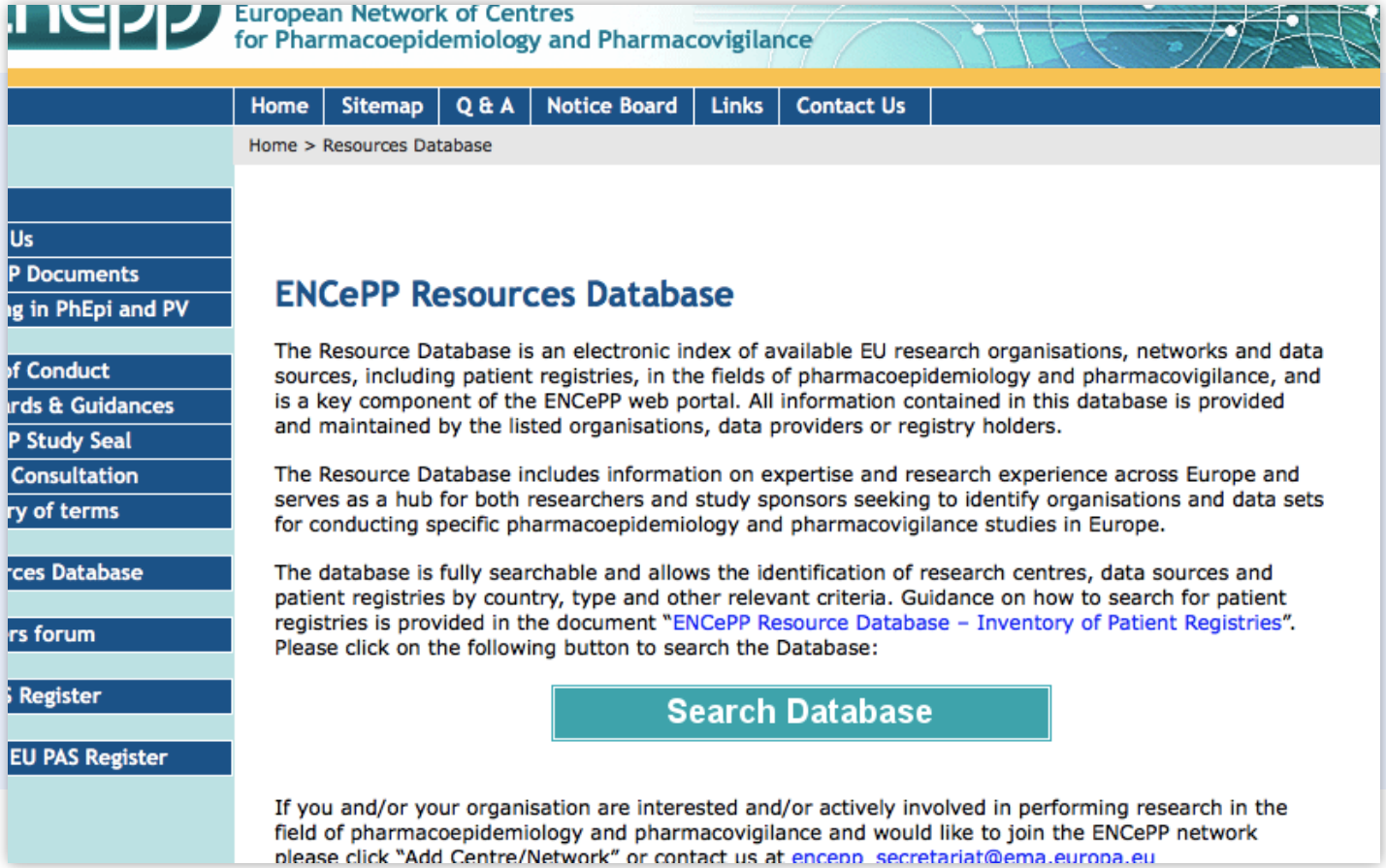
Show more

Share Cite

https://doi.org/10.1016/j.ijmedinf.2019.02.006 Get rights and content

Current sustainability RWE catalogues: ENCePP resource

Sustainable but updated irregularly



What needs to be sustained to make RWE FAIR?

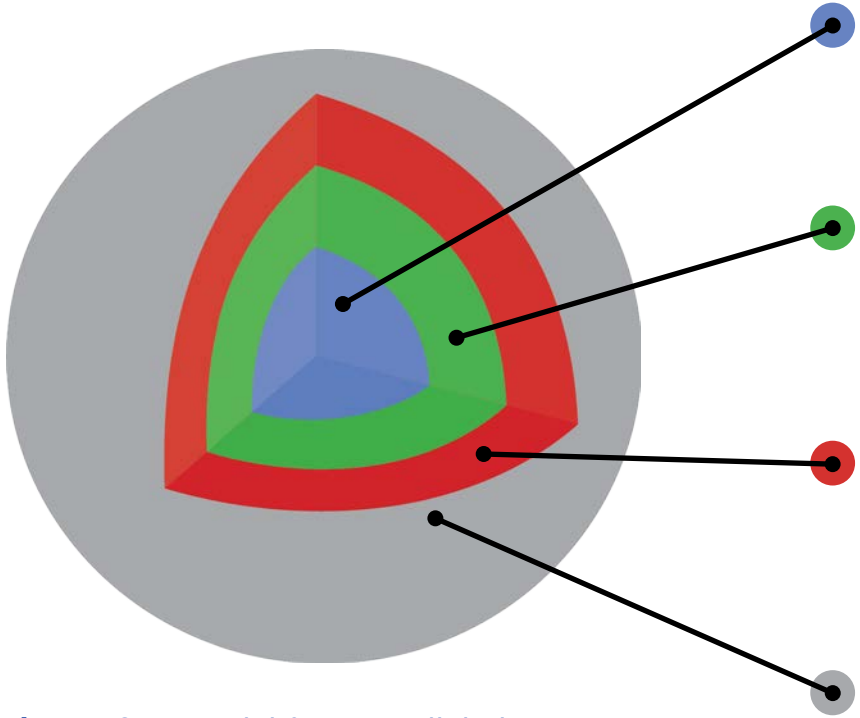


Figure 8. A model for FAIR digital objects, noting the elements that need to be in place for data to be Findable, Accessible, Interoperable, and Reusable

DIGITAL OBJECT

Data, code, and other research outputs

At its most basic level, data or code is a bitstream or binary sequence. For this to have meaning and to be FAIR, it needs to be represented in standard formats and be accompanied by persistent identifiers (PIDs), metadata, and documentation. These layers of meaning enrich the object and enable reuse.

IDENTIFIERS

Persistent and unique

Digital objects should be assigned a unique and persistent identifier such as a DOI or URN. This enables stable links to the object and supports citation and reuse to be tracked. Identifiers should also be applied to other related concepts such as the data authors (ORCIDs), projects (RAIDs), funders, and associated research resources (RRIDs).

STANDARDS & CODE

Open, documented formats

Digital objects should be represented in common and ideally open file formats. This enables others to reuse them as the format is in widespread use and software is available to read the files. Open and well-documented formats are easier to preserve. Data also need to be accompanied by the code used to process and analyse the data.

METADATA

Contextual documentation

For digital objects to be assessable and reusable, they should be accompanied by sufficient metadata and documentation. Basic metadata will enable data discovery, but much richer information and provenance is required to understand how, why, when, and by whom the objects were created. To enable the broadest reuse, they should be accompanied by a plurality of relevant attributes and a clear and accessible usage license.

Source: https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_0.pdf

Components in the FAIR ecosystem

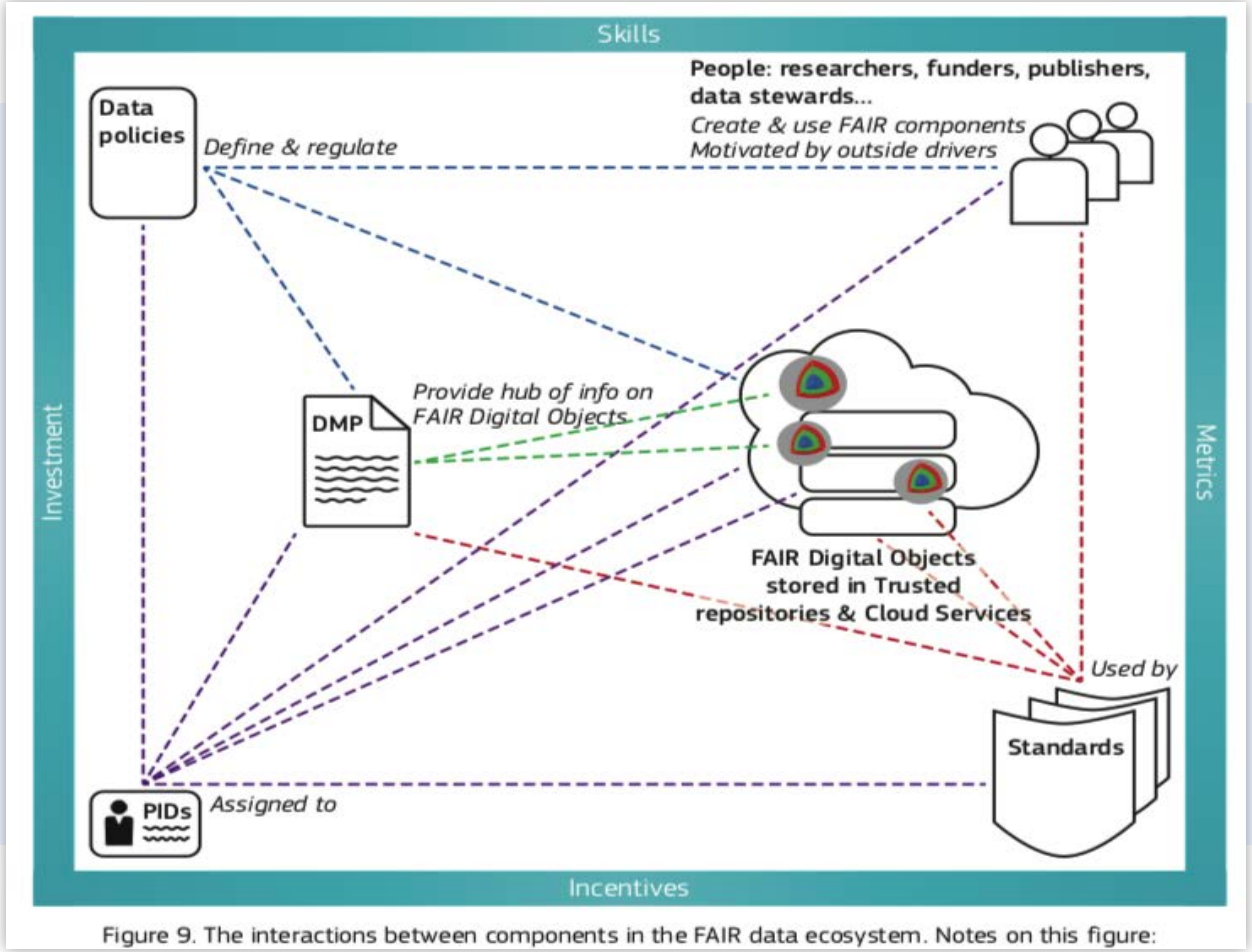


Figure 9. The interactions between components in the FAIR data ecosystem. Notes on this figure:

10 Source: https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_0.pdf

Strengthening use of real-world data in medicines development: metadata for data discoverability and study replicability

Strengthening use of real-world data in medicines development: metadata for data discoverability and study replicability

- FAIR data ecosystem & policy
- Data management plan of catalogue
- FAIR digital objects inter linkable and beyond what is currently in catalogues

Data collections

- Institution needs to assign PID
- Data source (a collection of data banks covering the same population)
- Data bank (a data collection)

Data use

- Collaborations
- Common data models
- Studies

Data definitions & standards

- No standards to describe data banks

Sources: https://www.karger.com/Article/Fulltext/502951; https://www.liebertpub.com/doi/full/10.1089/big.2013.0023

Required elements for sustainability

- **Compliance with FAIR principles: live up to the EOSC standards**
- Catalogue IT infrastructure should be a FAIR data node
- Need for PIDs for DAPS & data banks
- Clear standards for metadata description of health data (interaction with CDISC?)
- Updating/maintenance of the content of the catalogue data collections and data use
- Voluntary updates do not work: create incentives to keep it updated
- Funding for maintenance of the catalogue IT infrastructure

MIVERVA proof-of-concept catalogue

- A proof-of-concept process for sustainable collection of metadata in the catalogue, including quality control, will be developed for this pilot project

  1. A preliminary proposal for metadata collection and maintenance is that, **when a study is conducted** (e.g., on request by EMA or PRAC), the study investigators would use the catalogue throughout the study life cycle, and catalogue updates would be an active and funded part of the study itself. The following aspects would be considered:

     - The funding, which would be requested from the study funder
     - Each entry describing data banks or data sources in the catalogue would be associated with a specific study or institution. Users would need to be able to view and compare multiple entries of metadata for the same data banks and data sources. At this stage, resolution of differences is not proposed.
     - In many cases, the catalogue would be first populated at the point of a study, as is done in the EU PAS Registry. In the case of “emerging” data sources, that may not yet be routinely used in studies, a different route to support initial metadata entry may be required.

MIVERVA proof-of-concept catalogue

2. A second process for metadata collection and maintenance would be a **centrally coordinated and funded effort** to be implemented on a periodic basis, independent of a study. The following aspects would be considered:
   - The strategy to determine which institution(s) would receive funding to do this effort.
   - Non-native variables, e.g., the presence or absence of a disease, generally created through a study when an investigator selects and applies an algorithm to combine multiple codes (e.g., diagnostic, procedural, or treatment).
   - The governance structure.

A thorough exploration of these options is scheduled for Task 6 and will be discussed in Deliverables 8-9 of this pilot project.

Thank you!

For any question on this presentation, please contact: Malgorzata.Durka-Grabowska@ema.europa.eu
Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Address for visits and deliveries Refer to www.ema.europa.eu/how-to-find-us
Send us a question Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000