The FAIR Cookbook and FAIRsharing:

guidance and resources to make data FAIR

Susanna-Assunta Sansone, PhD


susanna-assunta.sansone@oerc.ox.ac.uk


@SusannaASansone


0000-0001-5306-5690

Oxford Academic Lead for Research Practice,
Engineering Science, Professor of Data Readiness
Oxford e-Research Centre, Associate Director

ELIXIR Interoperability Platform,
Co-Lead

EOSC FAIR Metrics and Data Quality
Task Force, Member

FAIR cookbook

(Life Science focused)

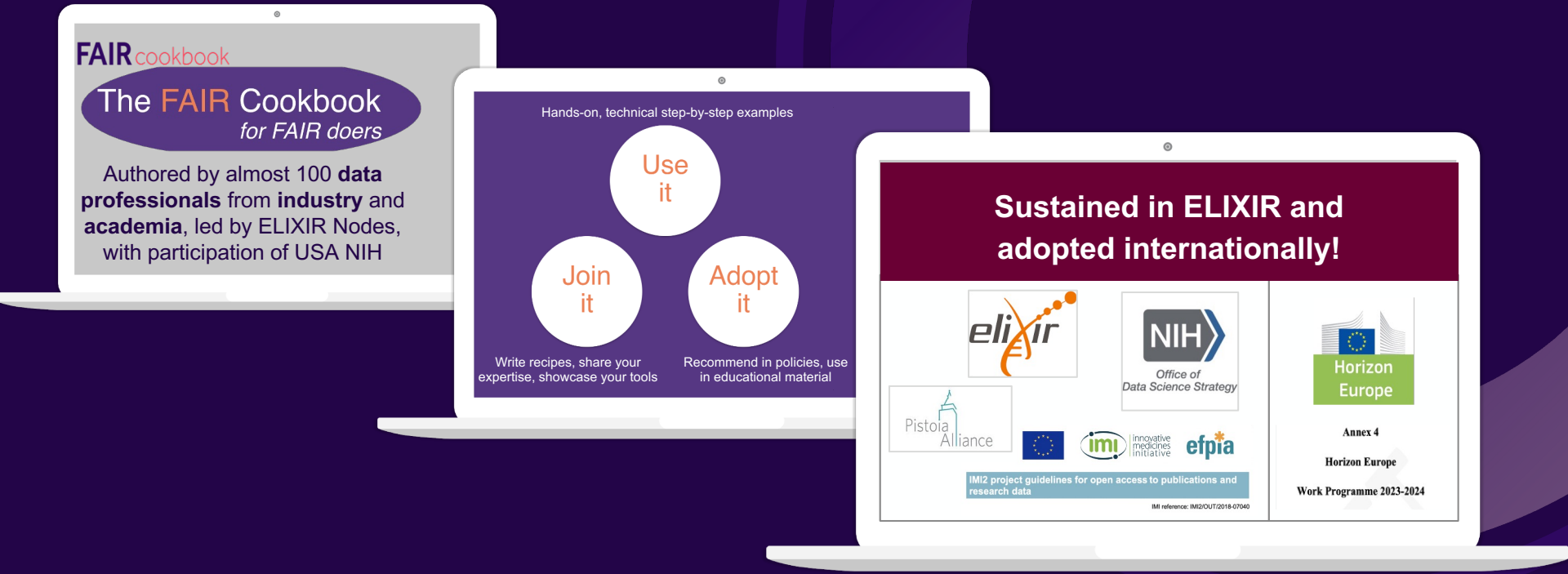
Recommended by European funders, embedded and used by research infrastructures

FAIRsharing.org

(All disciplines)

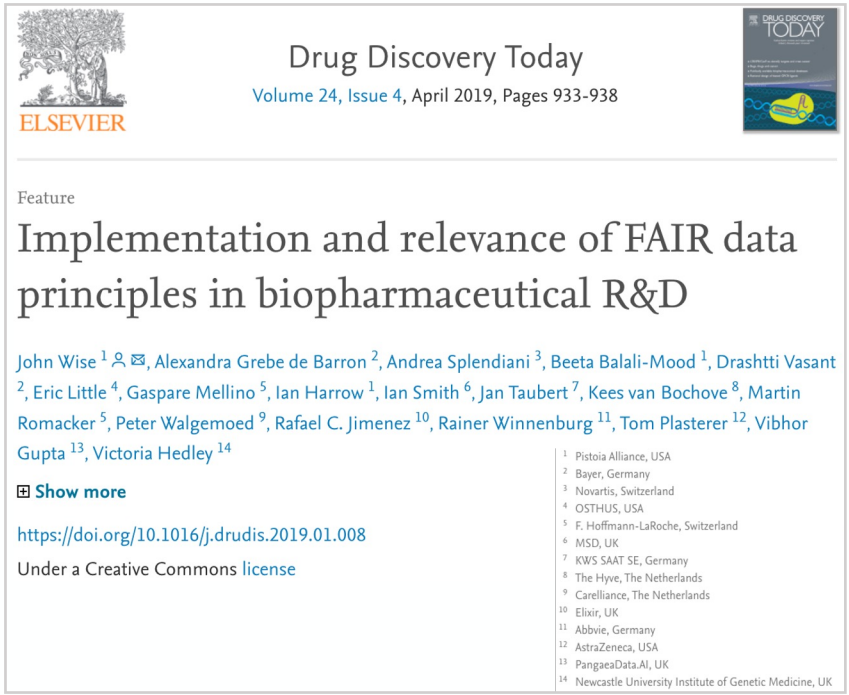
FAIRCOOKBOOK

An open, live resource for the life science with recipes that cover the operation steps of FAIR data management



faircookbook.elixir-europe.org

FAIR as driver of the digital transformation



ELSEVIER

Drug Discovery Today

Volume 24, Issue 4, April 2019, Pages 933-938

Feature

Implementation and relevance of FAIR data principles in biopharmaceutical R&D

John Wise [1], Alexandra Grebe de Barron [2], Andrea Splendiani [3], Beeta Balali-Mood [1], Drashtti Vasant [2], Eric Little [4], Gaspare Mellino [5], Ian Harrow [1], Ian Smith [6], Jan Taubert [7], Kees van Bochove [8], Martin Romacker [5], Peter Walgemoed [9], Rafael C. Jimenez [10], Rainer Winnenburg [11], Tom Plasterer [12], Vibhor Gupta [13], Victoria Hedley [14]

Show more

1 Pistoia Alliance, USA
2 Bayer, Germany
3 Novartis, Switzerland
4 OSTHUS, USA
5 F. Hoffmann-LaRoche, Switzerland
6 MSD, UK
7 KWS SAAT SE, Germany
8 The Hyve, The Netherlands
9 Carelliance, The Netherlands
10 Elixir, UK
11 Abbvie, Germany
12 AstraZeneca, USA
13 PangaeaData.AI, UK
14 Newcastle University Institute of Genetic Medicine, UK

https://doi.org/10.1016/j.drudis.2019.01.008

PHARMA

Fair Data Principles in Pharma Industry to Derive Meaningful Insights

Published 4 months ago on October 7, 2019
By Piyush Rishi

How can we get pharma R&D to embrace FAIR data?

Posted on January 6th, 2020 by Ani Marrs in Pharma R&D

- To improve biopharma R&D productivity
- To enables powerful new AI analytics to access data for ML and prediction

- **Requirements**
  - financial, technical, training
- **Challenges**
  - change the culture, show business value, achieve the 'FAIR enough' on an enterprise scale

Motivations and ambitions to build the FAIR Cookbook

beyond the hype

Motivations

Large body of generic FAIR guidance

Non-specific guidance for the life sciences

Lack of practical examples of ‘how-to’ with different data types and scenarios

Ambitions

Target specific situations to deliver a guide with applied examples

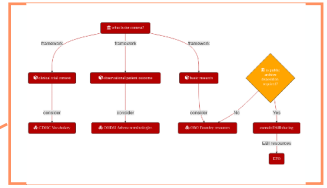
Join academia and industry forces to make the case for FAIR data management

Build capacity for high quality data management in the private and public sectors

Anatomy of a recipe

components

Contents



Ingredients
An idea of tools/skills needed

Tool Name	Tool Location	Tool function
ROBOT	http://robot.obolibrary.org/	ontology management cli

Step by step process
Guidelines, process, description

Conclusions
What should I read next?

Practical elements, code snippets

```
#Python3
#zooma-annotator-script.py file
def get_annotations(propertyType, propertyValues, filters = ""): "
```

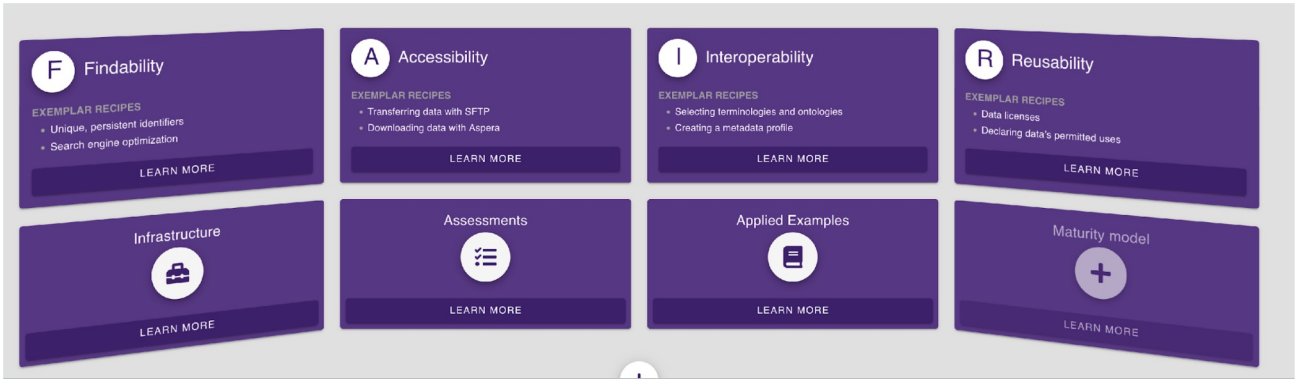
Examples

7.12.1. Competency questions for the Ontology ROBOT use case
7.12.2. Application ontology for metabolomics

RDMkit
More about Identifiers from the RDM Toolkit

Pistoia Alliance
The Pistoia Alliance FAIRtoolkit use cases: Adoption and Impact of an identifier policy at Astra-Zeneca

Coverage and learning objectives



- Content covers:
  - omics
  - pre-clinical
  - clinical areas

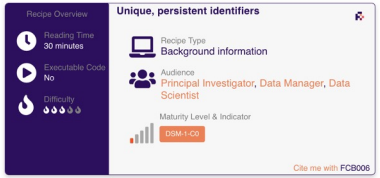
But not limited to it!

1. Learn how to improve the FAIRness with **exemplar datasets**
2. Understand the levels and **indicators** of FAIRness
3. Discover **open source** technologies, tools and services
4. Find out the required **skills**
5. Acknowledge the **challenges**

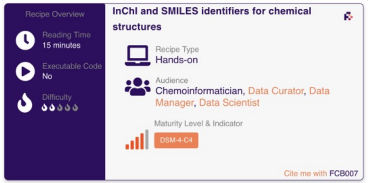
Focuses on addressing needs

Goal: **improving visibility of content**, e.g.:

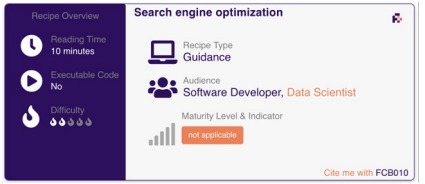
F Findability



https://w3id.org/faircookbook/FCB006



https://w3id.org/faircookbook/FCB007

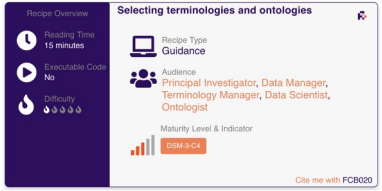


https://w3id.org/faircookbook/FCB010

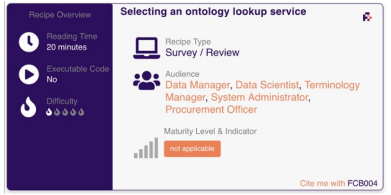
Goal: **semantic integration of datasets from multiple sources**, e.g.:

I Interoperability

Infrastructure



https://w3id.org/faircookbook/FCB020

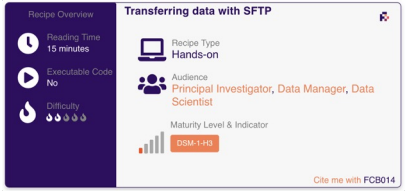


https://w3id.org/faircookbook/FCB004

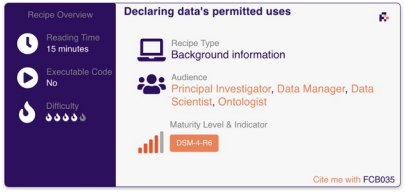
Goal: **security compliance and with regulators**, e.g.:

A Accessibility

R Reusability

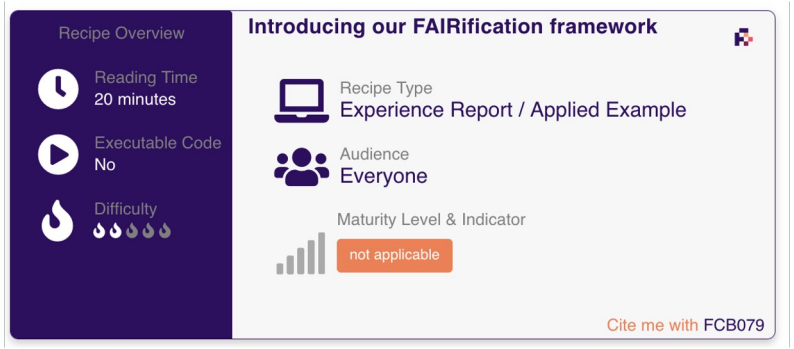


https://w3id.org/faircookbook/FCB014

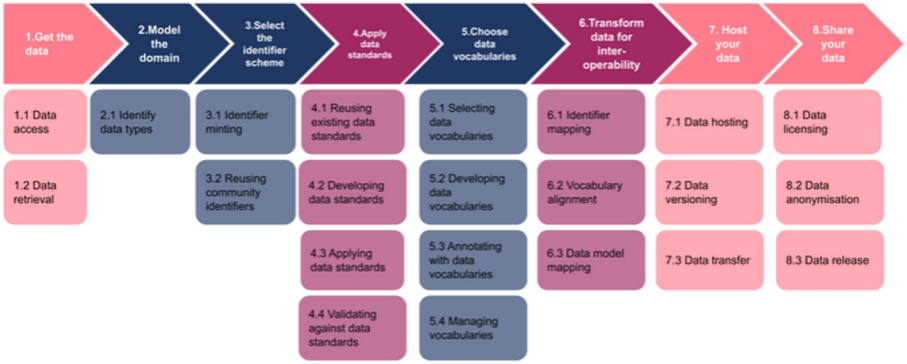


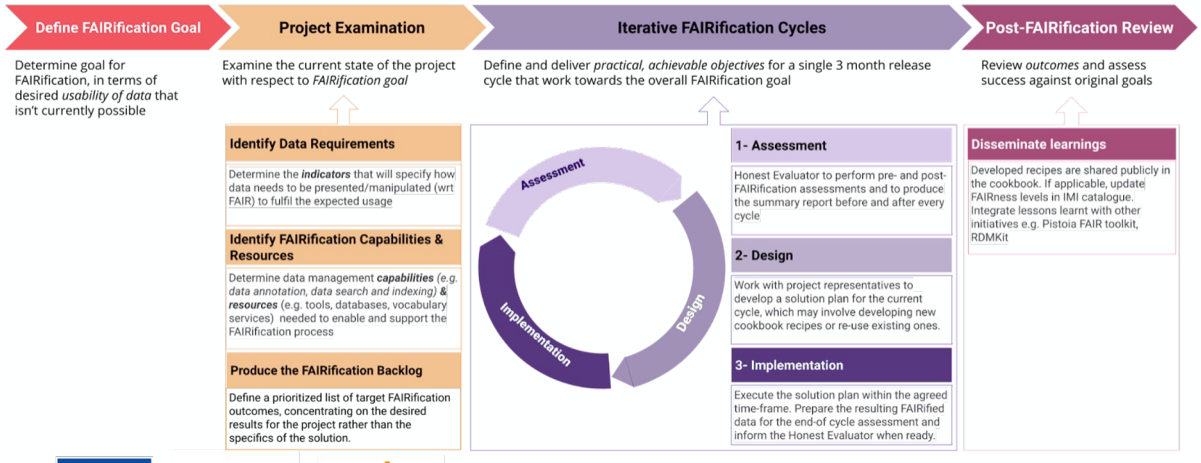
https://w3id.org/faircookbook/FCB035

The FAIRification framework in a recipe

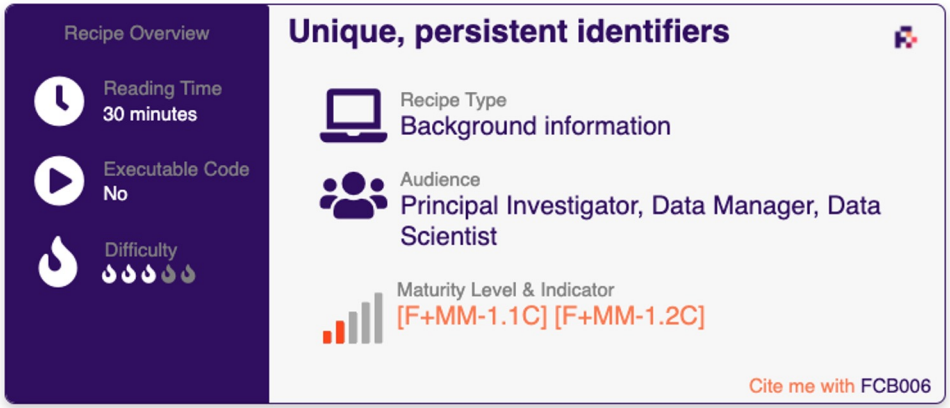


https://w3id.org/faircookbook/FCB079





Credit and citability: because all contributions matters!



https://w3id.org/faircookbook/FCB006

w3id

ORCID

CreDiT attribution ontology

CRT

Name	Orcid	Affiliation	Type	Elixir Node	Credit Role
Andrea Splendiani		Novartis AG			Conceptualization
Alasdair J G Gray		Heriot Watt University		elixir UNITED KINGDOM	Writing - Original Draft
Chris Evelo		University of Maastricht		elixir NETHERLANDS	Writing - Original Draft
Egon Willighagen		University of Maastricht		elixir NETHERLANDS	Writing - Original Draft
Philippe Rocca-Serra		University of Oxford		elixir UNITED KINGDOM	Writing – Review & Editing, Conceptualization

Over 80 recipes and counting

Search the cookbook recipes

Filters:

Search recipes names

Search recipes by type

Search recipes by audience

Reading Time

From 0 min To 60 min

Executable code

Yes Both No

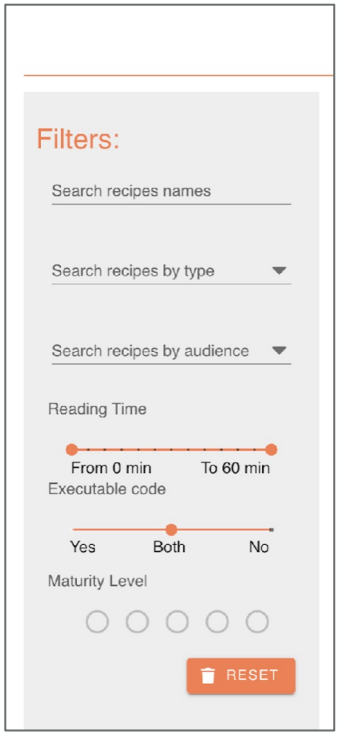
Maturity Level

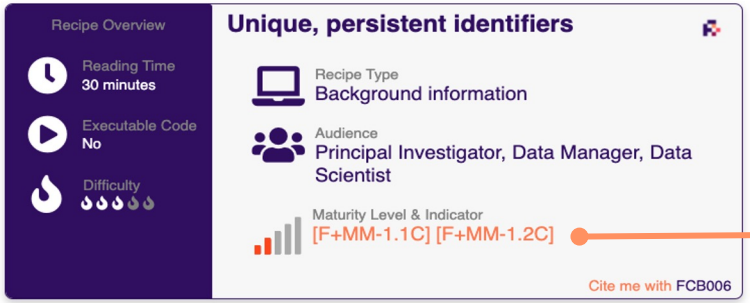
RESET

Identifier	Recipe Name	Recipe Type	Reading Time	Executable Code	Audience	Maturity Level
FCB015	Downloading data with Aspera	Hands-on	15 min	✓	Principal Investigator, Data Manager, Data Scientist	● ○ ○ ○ ○
FCB014	Transferring data with SFTP	Hands-on	15 min	✓	Principal Investigator, Data Manager, Data Scientist	● ○ ○ ○ ○
FCB054	EHDEN OHDSI discovery with Schema.org	Experience Report / Applied Example	15 min	✗	Terminology Manager, Data Manager, Data Scientist, Ontologist	● ● ○ ○ ○
FCB042	eTox - omics datasets	Experience Report / Applied Example	20 min	✗	Data Manager, Data Curator, Ontologist	● ● ○ ○ ○
FCB067	FAIR High-Content Screening data deposition	Experience Report / Applied Example	15 min	✓	Data Manager	● ● ○ ○ ○
FCB037	Making omics data matrix FAIR	Hands-on	30 min	✓	Data Manager, Data Scientist	● ● ○ ○ ○
FCB038	FAIR Data Matrices	Hands-on	30 min	✓	Principal Investigator, Data Manager, Data Scientist	● ● ○ ○ ○
FCB039	Structuring data matrices	Hands-on	30 min	✓	Principal Investigator, Data Manager, Data Scientist	● ● ○ ○ ○
FCB040	Exploring data with SPARQL	Hands-on	30 min	✓	Principal Investigator, Data Manager, Data Scientist	● ● ○ ○ ○
FCB041	Integrating data	Hands-on	30 min	✓	Principal Investigator, Data Manager, Data Scientist	● ● ○ ○ ○
FCB062	FAIR Computational Workflows	Hands-on	15 min	✓	Principal Investigator, Data Manager, Data Scientist	● ● ○ ○ ○
FCB043	ND4BB - chemical activities datasets	Experience Report / Applied Example	20 min	✗	Data Manager, Data Curator	● ● ○ ○ ○

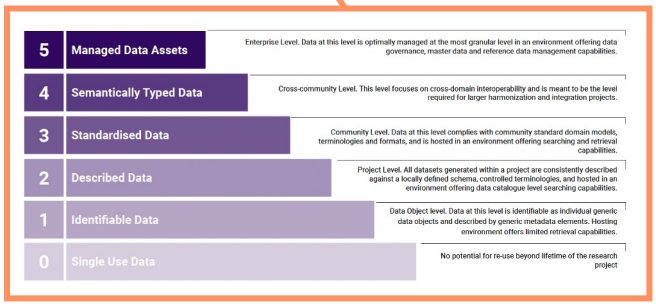
https://faircookbook.elixir-europe.org/content/search-wizard.html

Maturity level and indicators

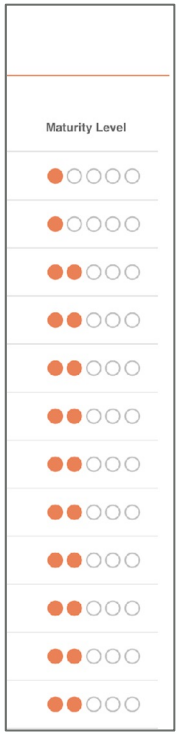




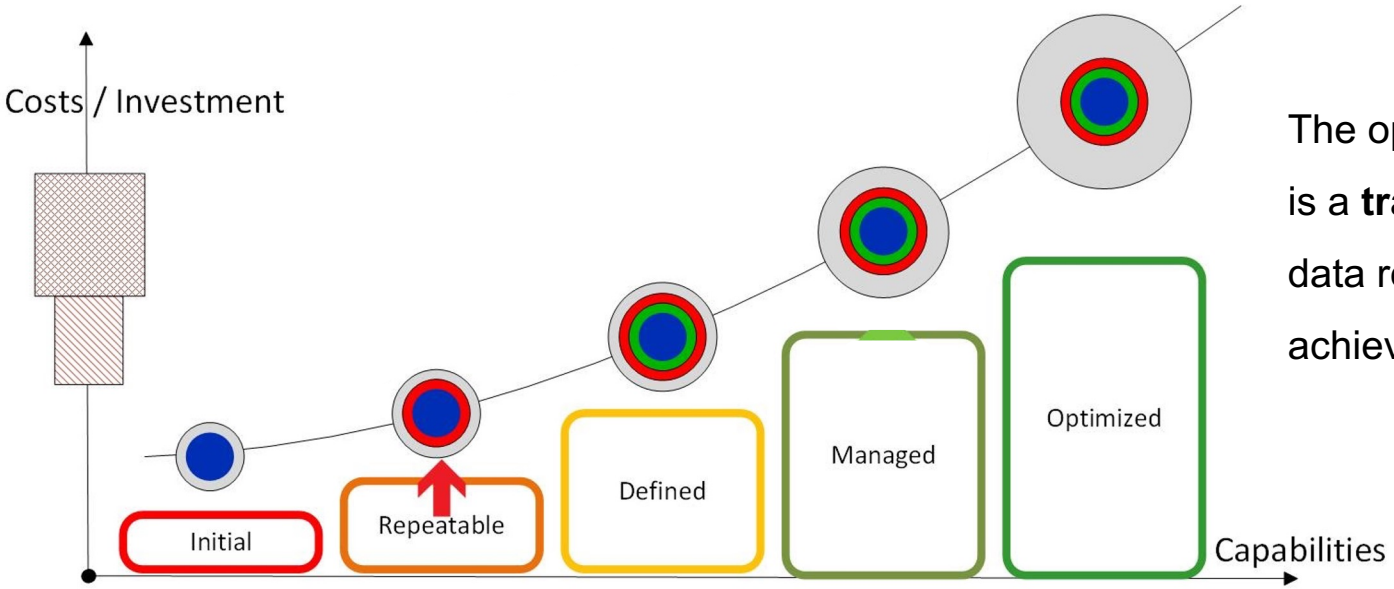
Provide insights into FAIR Maturity reached by applying a specific recipe to improve a dataset



https://fairplus.github.io/Data-Maturity

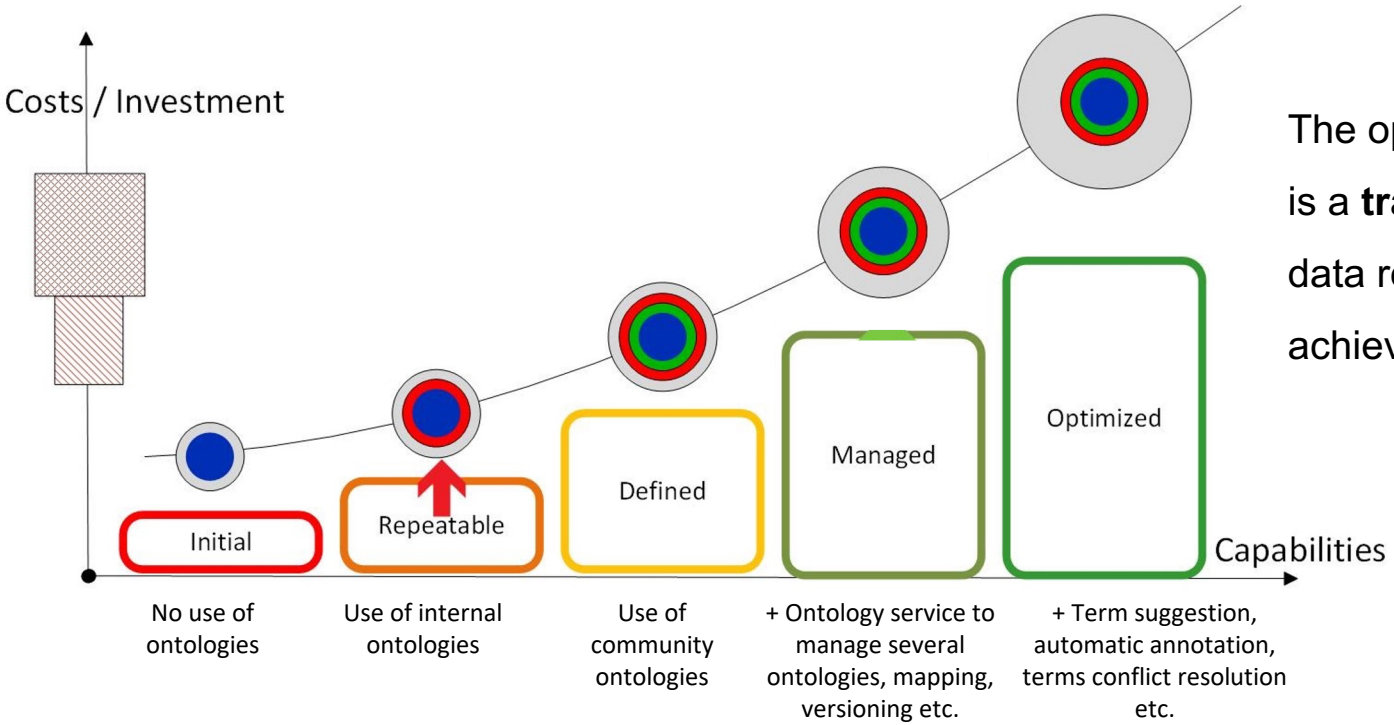


Capability needed to improve data reuse



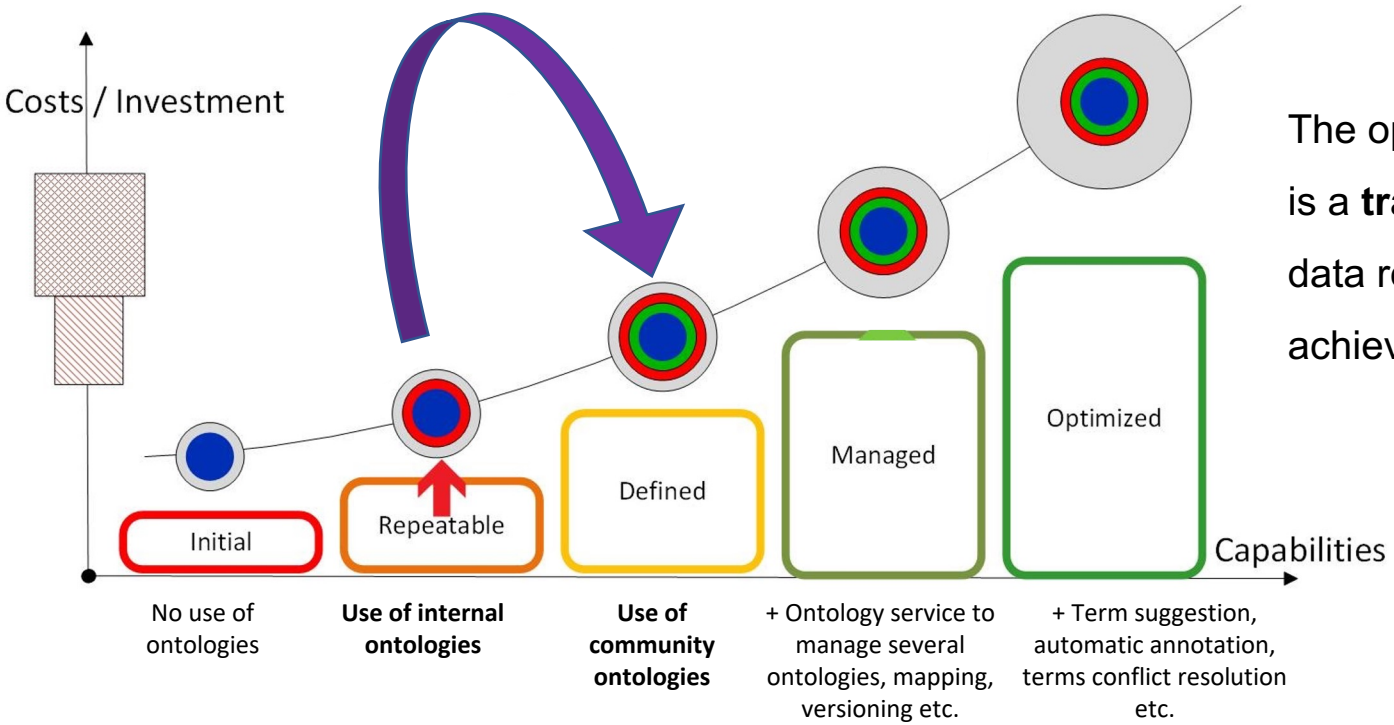
The optimum level of FAIRness is a **trade-off** between desired data reuse level and cost to achieve that level

Capability needed to improve data reuse: the ontology example



The optimum level of FAIRness is a **trade-off** between desired data reuse level and cost to achieve that level

FAIR cookbook
How to build application ontology



The optimum level of FAIRness is a **trade-off** between desired data reuse level and cost to achieve that level

Evaluating the FAIR Cookbook utility and value based on three uses

- As an **educational material** on FAIR in a training context
- As **a practical guidance** to improve day-to-day tasks for FAIRer data
- As **contributor towards changing the culture** and to **identify investment areas** in research data management
  - (behind the firewalls) outcomes were expressed in terms of satisfaction of the value of the recipes, against specific tasks, or challenges addressed
  - they reported a positive contribution towards their discussion on return on investment to operationalize FAIR

FAIRCOOKBOOK

An open, live resource for the life science with recipes that cover the operation steps of FAIR data management

Pre-print:
doi.org/10.5281/zenodo.7156792

Article in press at:

scientific data

https://faircookbook.elixir-europe.org

Editorial Board: faircookbook-ed@elixir-europe.org

FAIR cookbook

(Life Science focused)

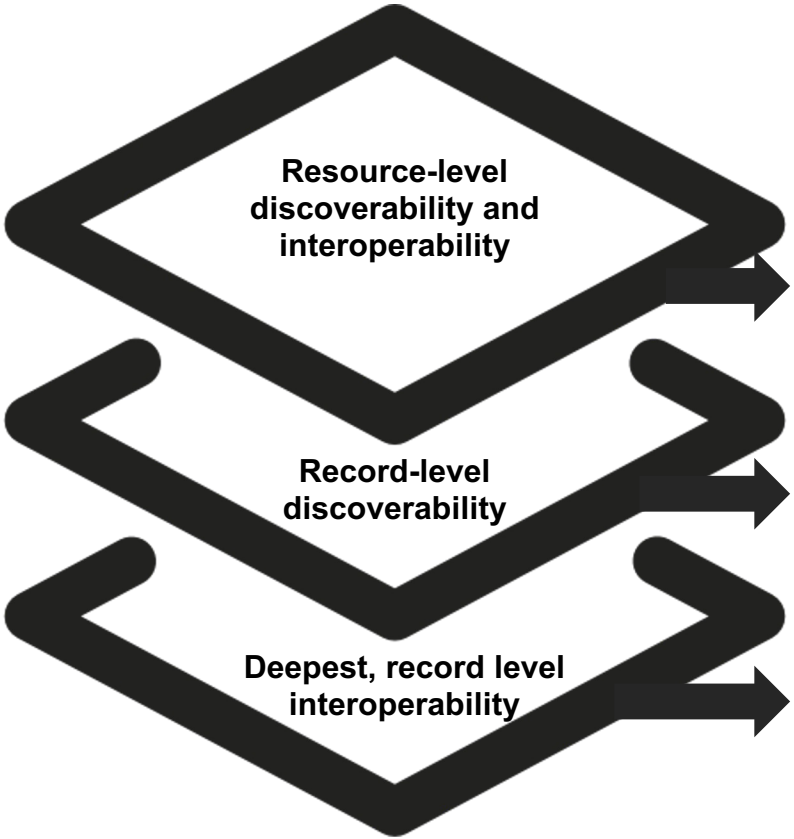
Recommended by European funders, embedded and used by research infrastructures

FAIRsharing.org

(All disciplines)

Standards for different purposes

FAIRsharing.org



A database or among databases

Datasets in a database

Datasets and data reuse

W3C Data Catalog Vocabulary (DCAT)

Schema.org

DublinCore

Standards for data and metadata - a dynamic ecosystem

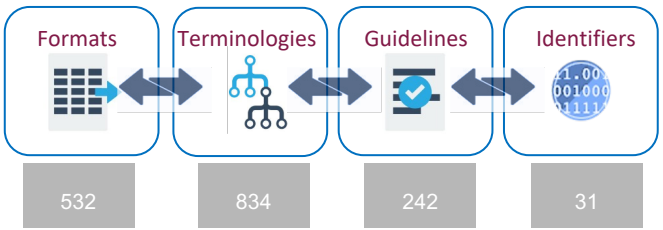
FAIRsharing.org

Conceptual model, conceptual schema, exchange formats to represent, contain and move information

Controlled vocabularies, thesauri, ontologies to disambiguate terms and enable semantic relationships

Minimum information reporting requirements, or checklists to report the same core, essential information

Unambiguous, persistent and context-independent schema to identify data and metadata elements



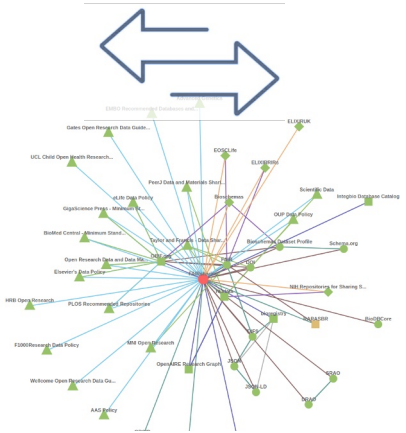
As of 4 April 2023

An informative and educational resource, and a service

FAIRsharing.org

POLICIES

by funders, journals
and other organizations



DATABASES

including repositories
and knowledgebases

COMMUNITY STANDARDS

Formats | Terminologies | Guidelines | Identifiers

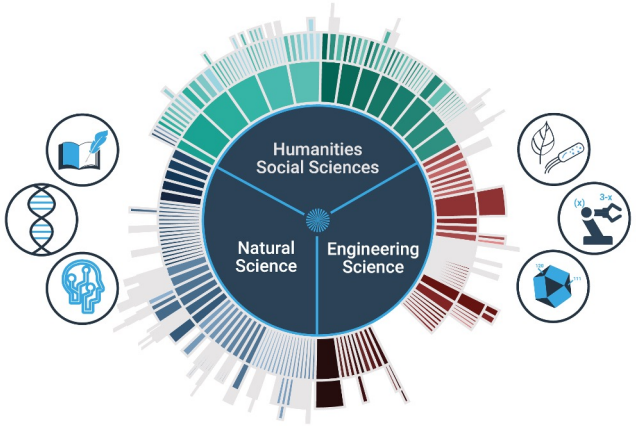
FAIRsharing provides **curated descriptions** and **relationship graphs** of **standards**, **databases** and **policies**

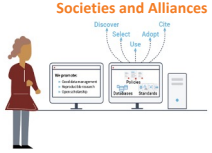
FAIRsharing mission and how we deliver it

FAIRsharing.org

Promote the *value* and *use* of **standards**, **databases** and **policies** in all disciplines,

by engaging **stakeholders** across all sectors, through all stages of the research life cycle





Guide **consumers** to *discover*, *select* and *use* these **resources** with confidence

Help **producers** to make their **resources** more *visible*, more widely *adopted* and *cited*

https://fairsharing.org/communities

FAIRsharing.org

FAIRsharing in numbers: content, contributions, coverage

1251

Contributors

Our community of record maintainers and curators

3856

Records

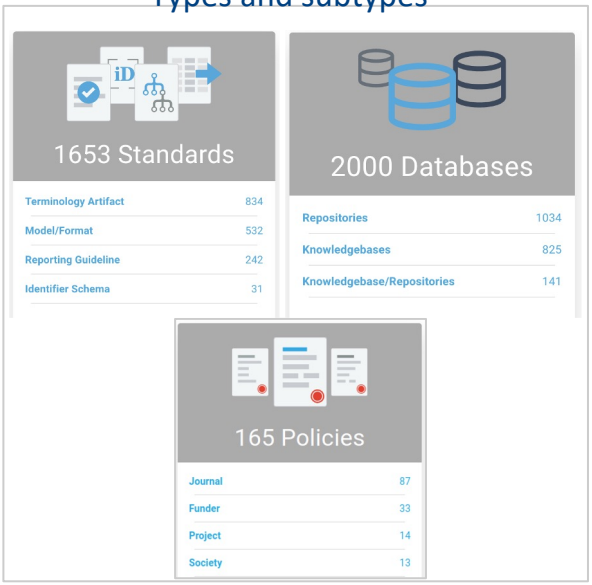
Descriptions of standards, databases and policies

198K

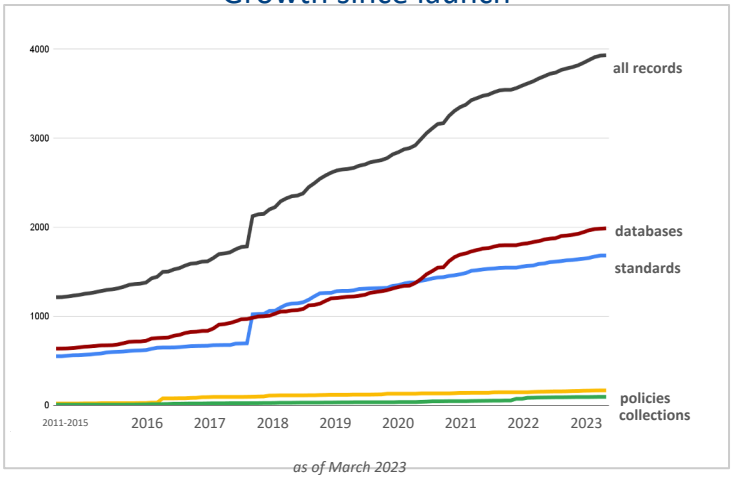
Views

Page visits since 2015

Types and subtypes

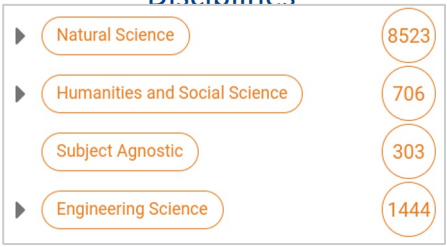


Growth since launch



as of March 2023

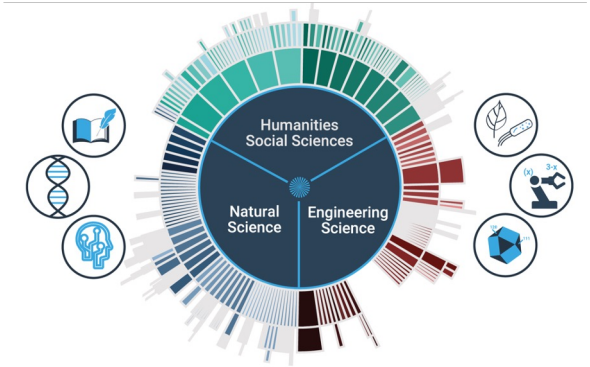
Disciplines

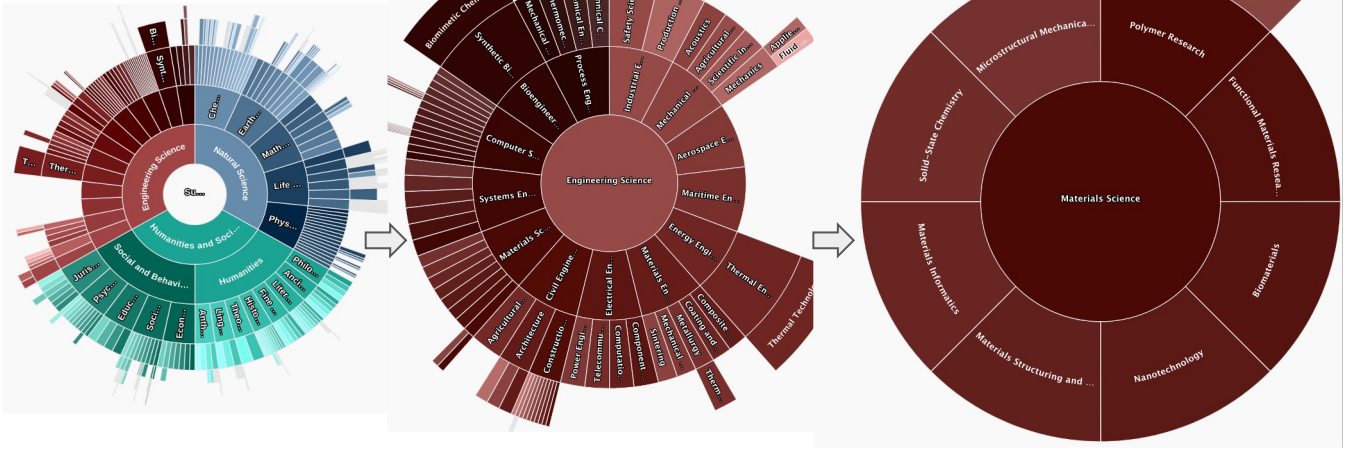


FAIRsharing.org

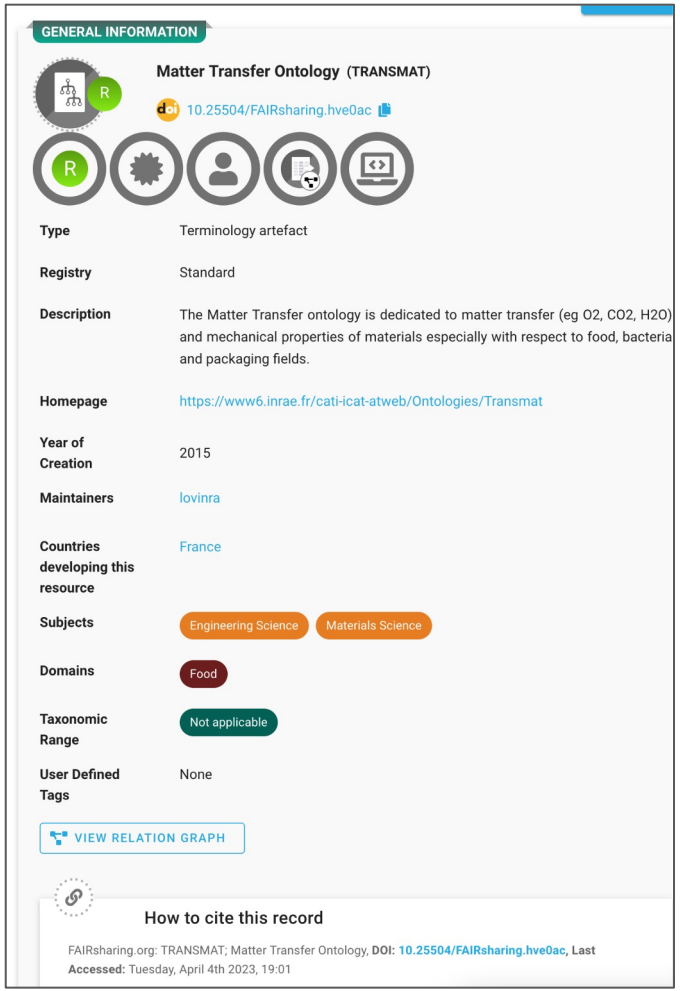
Look for resources relevant to material science

- Natural Science 8754
- Humanities and Social Science 718
- Subject Agnostic 314
- Engineering Science 1484
  - Industrial Engineering 37
  - Mechanical Engineering 31
  - Aerospace Engineering 6
  - Maritime Engineering 6
  - Energy Engineering 24
  - Materials Engineering 14
  - Electrical Engineering 32
  - Civil Engineering 153
    - Construction Engineering 114
    - Architecture 17
    - Agricultural Engineering 15
  - Materials Science 78
  - Systems Engineering 6
  - Computer Science 1144
  - Bioengineering 28
  - Process Engineering 20





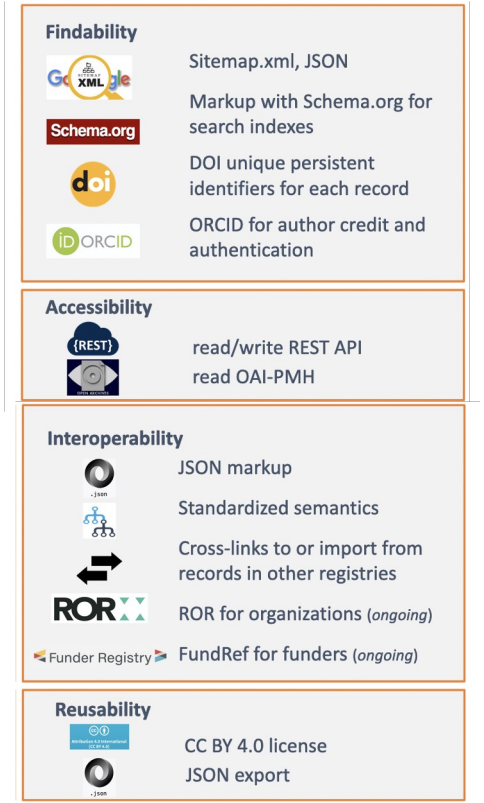
Submit any missing standards - your contribution matters!



Record DOI: 10.25504/FAIRsharing.hve0ac

Manually curated descriptions of each record, by in house team

FAIRsharing's FAIR content powers 3rd party tools



A growing number of tools and services access FAIRsharing API, and use it for **look-up**, **selection** and **content retrieval** for **standards** in particular in:

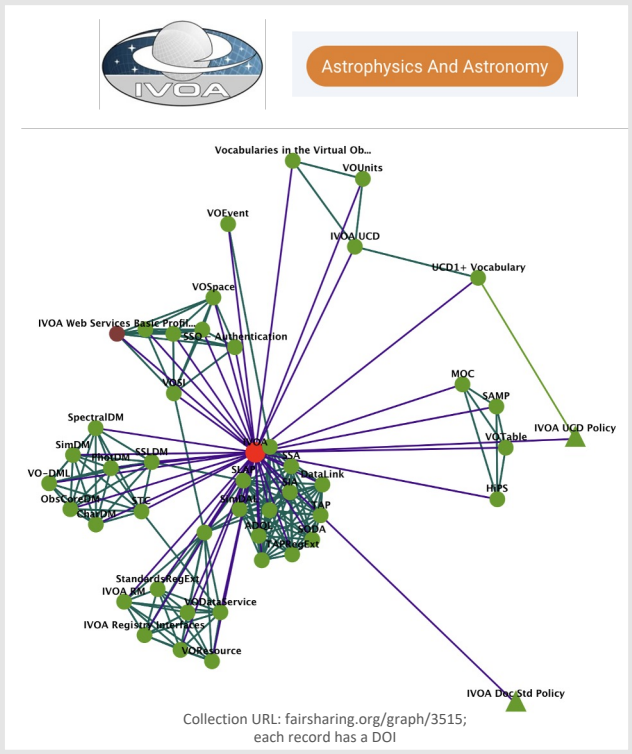
- creation of data management plans
- enrichment of guidance and training material
- assessment of FAIRness

eosc | FAIR Metrics and Data Quality Task Force

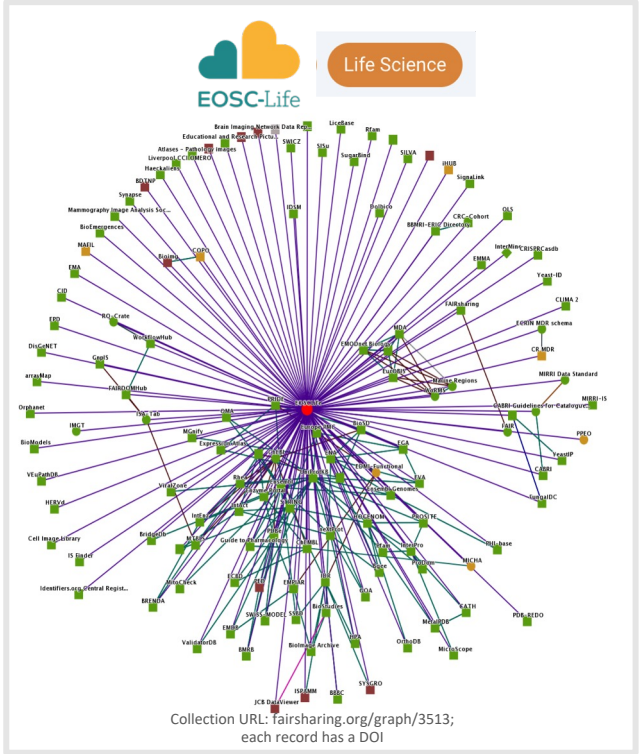
FAIR Assessment Tools: Towards an "Apples to Apples" Comparisons

doi.org/10.5281/zenodo.7463421

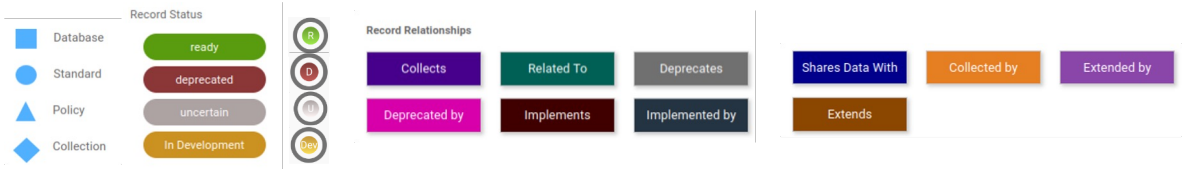
Showing organizations' "FAIR profiles" matters!



Collection URL: fairsharing.org/graph/3515;
each record has a DOI



Collection URL: fairsharing.org/graph/3513;
each record has a DOI



eosc

ESCAPE

ENVRI community

SSHOC

EOSC-Life

panosc

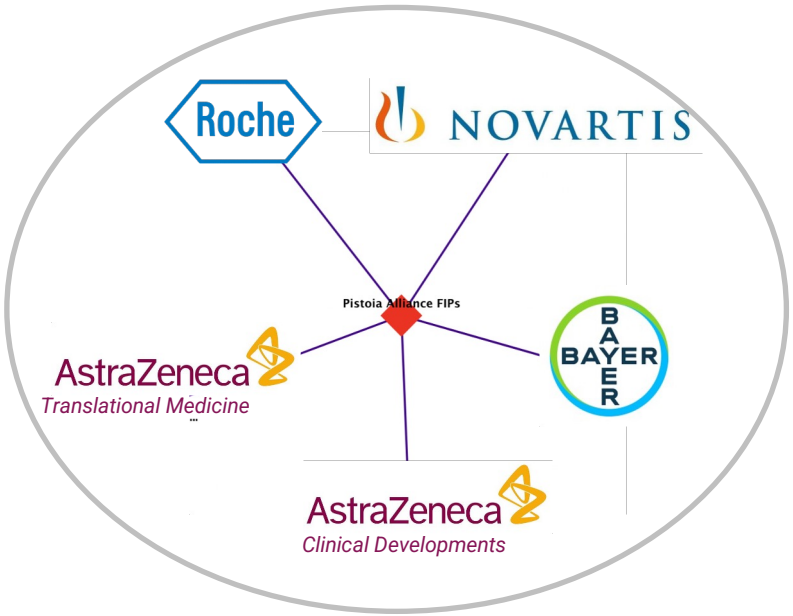
The **standards**, **repositories** and **policies** each EOSC Cluster uses or endorses

FAIRsharing.org

Building and comparing "FAIR profiles" - an example

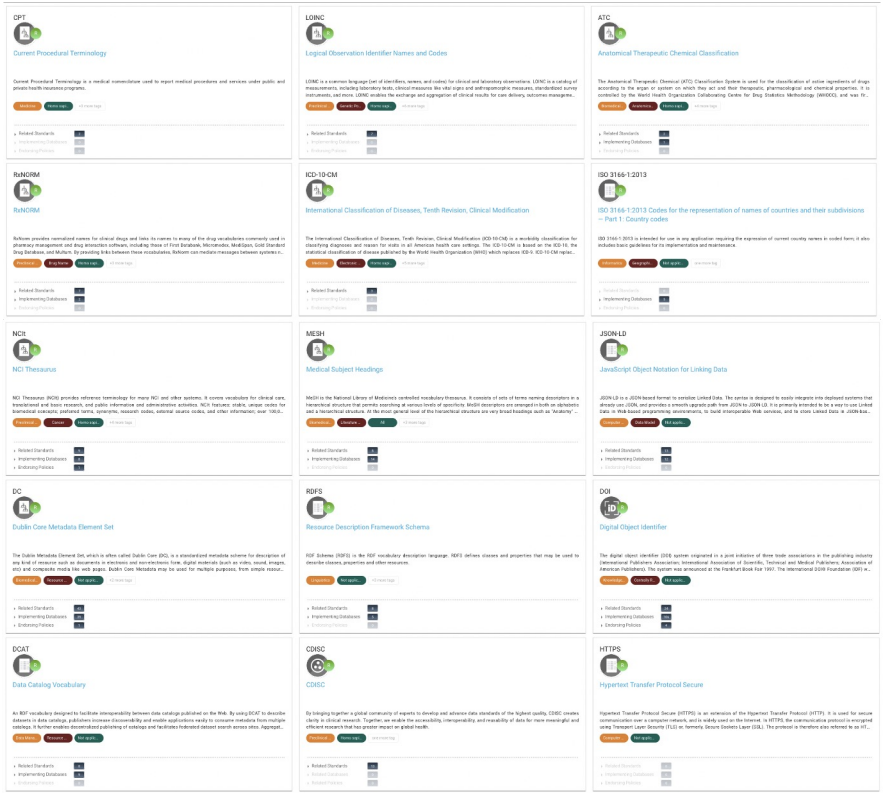
Pistoia Alliance

A collaboration with their FAIR Implementation WG



URL: beta.fairsharing.org/PistoiaAllianceFIPs

(work in progress!)



Disclaimer: These profiles speak for a limited community and do not represent any company standards

Pistoia Alliance

FAIRsharing.org

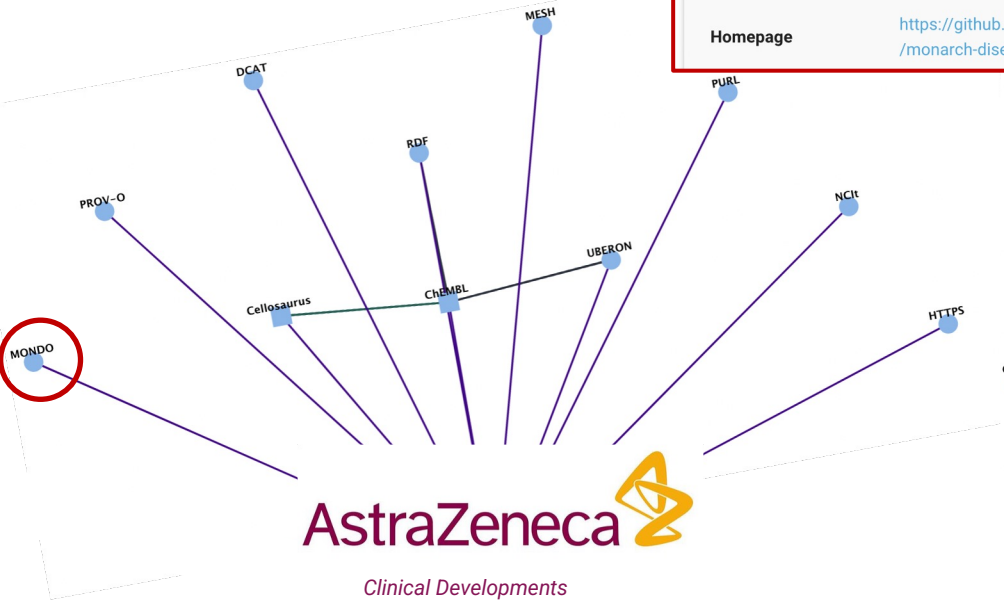
GENERAL INFORMATION

Monarch Disease Ontology (MONDO)

doi 10.25504/FAIRsharing.b2979t

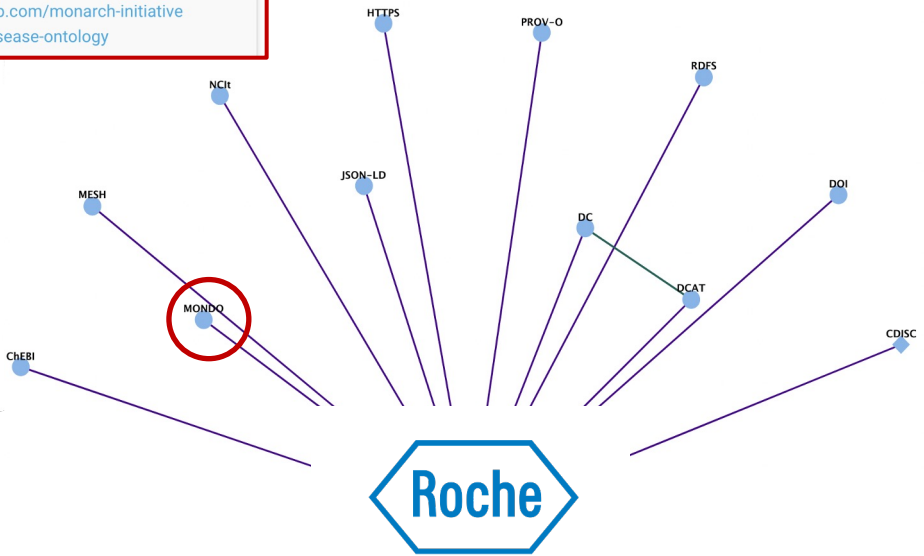
Type	Terminology artefact
Registry	Standard
Description	MonDO (Monarch Disease Ontology) is a semi-automatically constructed ontology that merges in multiple disease resources to yield a coherent merged ontology.
Homepage	https://github.com/monarch-initiative/monarch-disease-ontology

Snapshot of the semantic and syntactic standards used



AstraZeneca

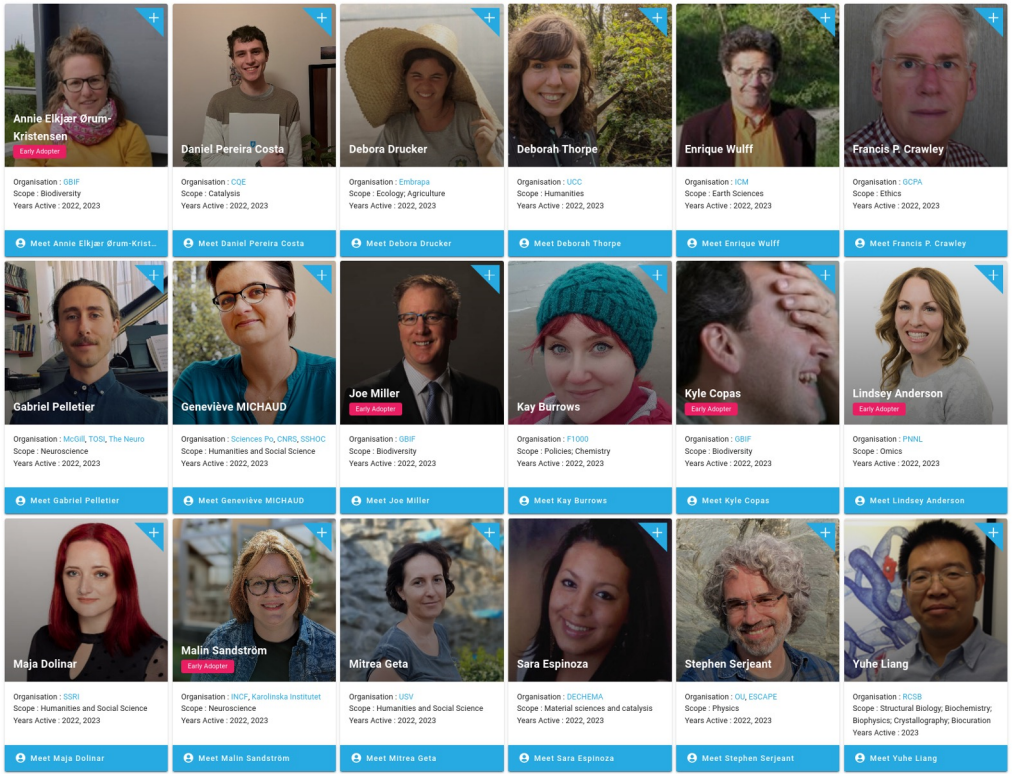
Clinical Developments



Roche

Disclaimer: These profiles speak for a limited community and do not represent any company standards

Engaging the expertise within community is essential



fairsharing.org/community_champions (2022-2023 term)

FAIRsharing Community Champion Programme

Domain experts who

- Help **curate content** content
- Contribute to **educational material**
- Gain **expertise**, **networking** and **attribution**

Funded by the:

RDA RESEARCH DATA ALLIANCE

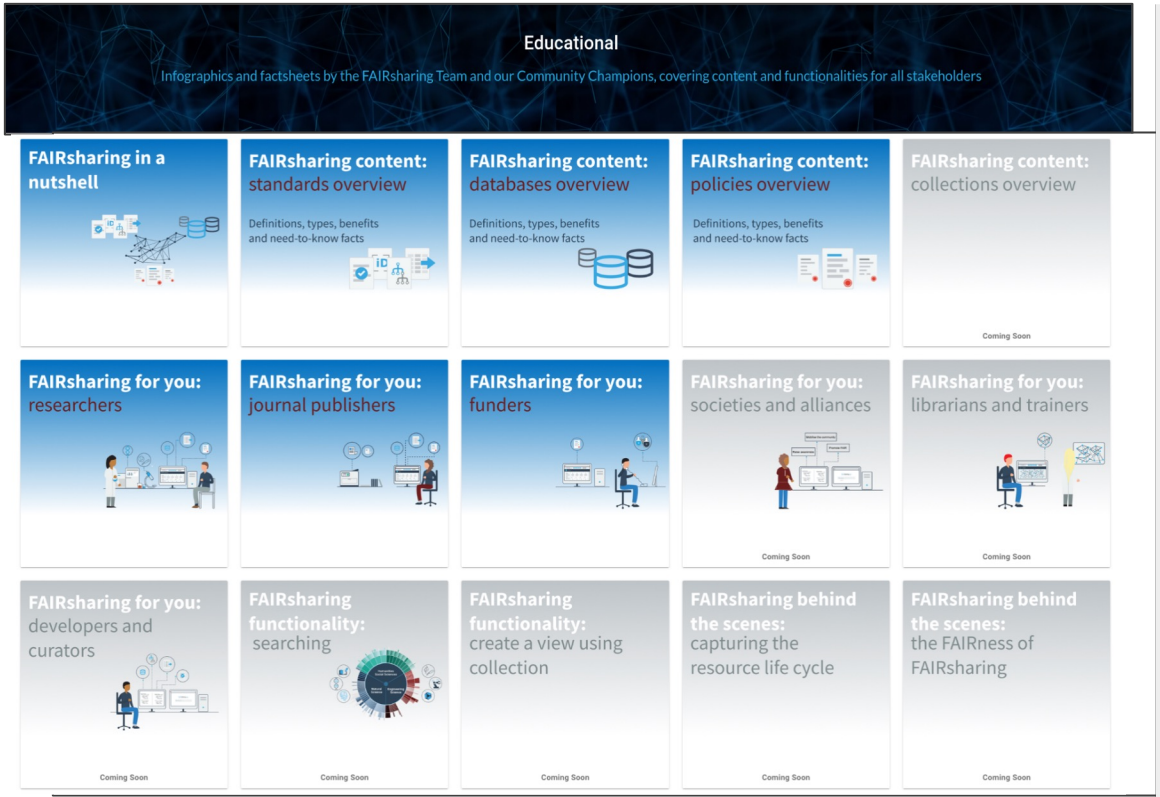
EOSC Future Funding Platform

Ambassadorship Programme

Building educational material collaboratively

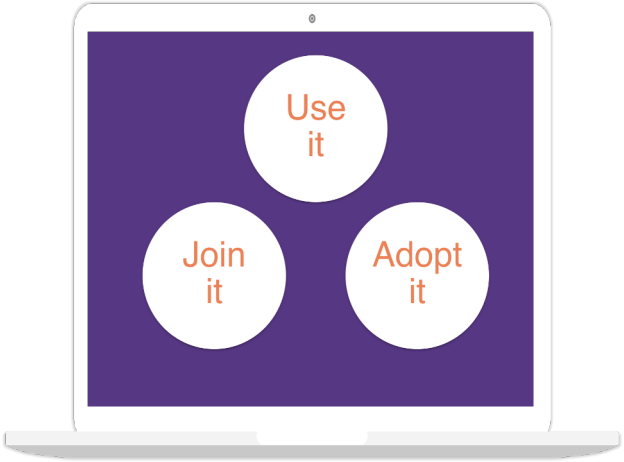
FAIRsharing.org

EXAMPLE



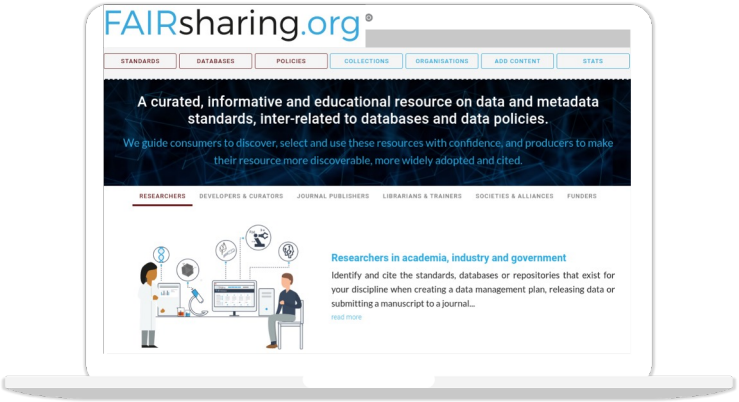
fairsharing.org/educational

(Life Science focused)

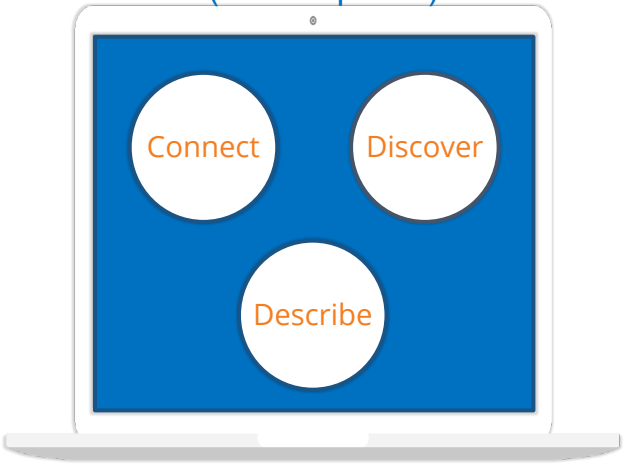


faircookbook.elixir-europe.org

faircookbook-ed@elixir-europe.org



(All disciplines)



fairsharing.org

contact@fairsharing.org