





The FAIR Cookbook and FAIRsharing: *guidance and resources to make data FAIR*

Susanna-Assunta Sansone, PhD

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

 [@SusannaASansone](https://twitter.com/SusannaASansone)

 0000-0001-5306-5690

*ELIXIR Interoperability Platform,
Co-Lead*



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



*EOSC FAIR Metrics and Data Quality
Task Force, Member*



FAIRcookbook

(Life Science focused)



Recommended by
European funders,
embedded and used
by research
infrastructures

FAIRsharing.org

(All disciplines)



Annex 4

Horizon Europe

Work Programme 2023-2024



FAIRCOOKBOOK

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

FAIRcookbook

The FAIR Cookbook for FAIR doers

Authored by almost 100 **data professionals** from **industry** and **academia**, led by ELIXIR Nodes, with participation of USA NIH

Hands-on, technical step-by-step examples

Use
it

Join
it

Adopt
it

Write recipes, share your expertise, showcase your tools

Recommend in policies, use in educational material

Sustained in ELIXIR and adopted internationally!



IMI2 project guidelines for open access to publications and research data

IMI reference: IMI2/OUT/2018-07040


Annex 4

Horizon Europe

Work Programme 2023-2024


faircookbook.elixir-europe.org

FAIR as driver of the digital transformation



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 Wise¹ ✉, Alexandra Grebe de Barron², Andrea Splendiani³, Beeta Balali-Mood¹, Drashtti Vasant², Eric Little⁴, Gaspare Mellino⁵, Ian Harrow¹, Ian Smith⁶, Jan Taubert⁷, Kees van Bochove⁸, Martin Romacker⁵, Peter Walgemoed⁹, Rafael C. Jimenez¹⁰, Rainer Winnenburger¹¹, Tom Plasterer¹², Vibhor Gupta¹³, Victoria Hedley¹⁴

[Show more](#)

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

Under a Creative Commons [license](#)

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

- 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

Fair Data Principles in Pharma Industry to Derive Meaningful Insights

Published 4 months ago on October 7, 2019
by [Pharma R&D](#)

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

Posted on January 6th, 2020 by [Ani Marrs](#) in [Pharma R&D](#)

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

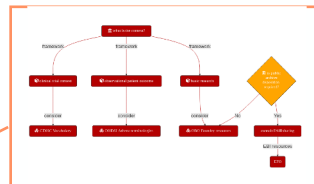
Build capacity for high quality data management in the private and public sectors

Anatomy of a recipe

components

Contents

- 7.1. Main Objectives
- 7.2. Graphical Overview
- 7.3. Capability & Maturity Table
- 7.4. FAIRification Objectives, Inputs and Outputs
- 7.5. Table of Data Standards
- 7.6. Ingredients
- 7.7. Step by step process
- 7.8. Conclusions
- 7.9. References
- 7.10. Supplementary material
- 7.11. Authors
- 7.12. License



Ingredients

An idea of tools/skills needed

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

Practical elements, code snippets

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

Step by step process

Guidelines, process, description

Conclusions

What should I read next?

Examples

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

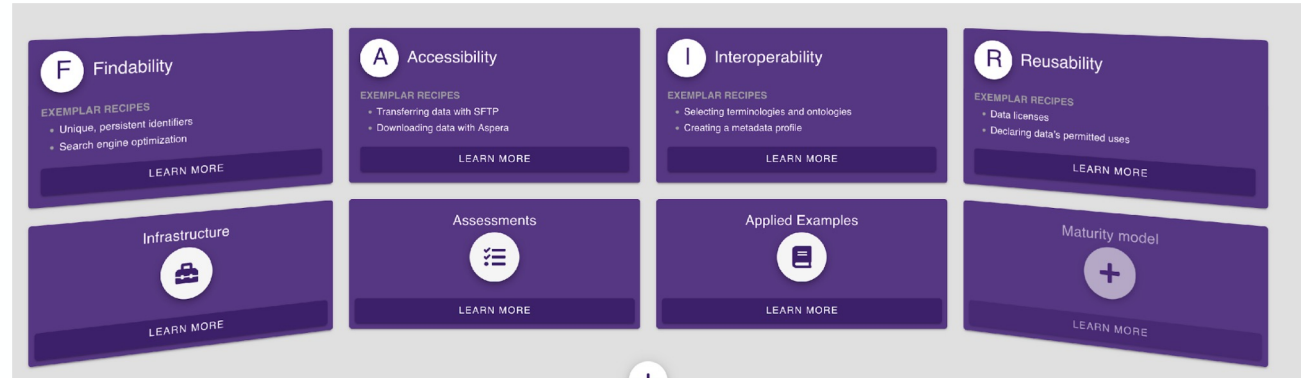
RDMkit

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



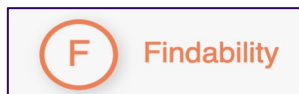
- Content covers:
 - omics
 - pre-clinical
 - clinical areas

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.:



Recipe Overview

- Reading Time: 30 minutes
- Executable Code: No
- Difficulty: 3/5

Recipe Type: Background information

Audience: Principal Investigator, Data Manager, Data Scientist

Maturity Level & Indicator: DSM-1-C0

Cite me with FCB006

<https://w3id.org/faircookbook/FCB006>

Recipe Overview

- Reading Time: 15 minutes
- Executable Code: No
- Difficulty: 3/5

Recipe Type: Hands-on

Audience: Cheminformatician, Data Curator, Data Manager, Data Scientist

Maturity Level & Indicator: DSM-4-C1

Cite me with FCB007

<https://w3id.org/faircookbook/FCB007>

Recipe Overview

- Reading Time: 10 minutes
- Executable Code: No
- Difficulty: 3/5

Recipe Type: Guidance

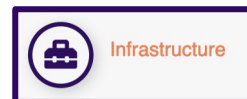
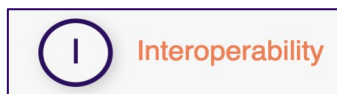
Audience: Software Developer, Data Scientist

Maturity Level & Indicator: not applicable

Cite me with FCB010

<https://w3id.org/faircookbook/FCB010>

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



Recipe Overview

- Reading Time: 15 minutes
- Executable Code: No
- Difficulty: 3/5

Recipe Type: Guidance

Audience: Principal Investigator, Data Manager, Terminology Manager, Data Scientist, Ontologist

Maturity Level & Indicator: DSM-3-C4

Cite me with FCB020

<https://w3id.org/faircookbook/FCB020>

Recipe Overview

- Reading Time: 20 minutes
- Executable Code: No
- Difficulty: 3/5

Recipe Type: Survey / Review

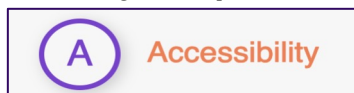
Audience: Data Manager, Data Scientist, Terminology Manager, System Administrator, Procurement Officer

Maturity Level & Indicator: not applicable

Cite me with FCB004

<https://w3id.org/faircookbook/FCB004>

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



Recipe Overview

- Reading Time: 15 minutes
- Executable Code: No
- Difficulty: 3/5

Recipe Type: Hands-on

Audience: Principal Investigator, Data Manager, Data Scientist

Maturity Level & Indicator: DSM-1-R0

Cite me with FCB014

<https://w3id.org/faircookbook/FCB014>

Recipe Overview

- Reading Time: 15 minutes
- Executable Code: No
- Difficulty: 3/5

Recipe Type: Background information

Audience: Principal Investigator, Data Manager, Data Scientist, Ontologist

Maturity Level & Indicator: DSM-4-P0

Cite me with FCB035

<https://w3id.org/faircookbook/FCB035>

The FAIRification framework in a recipe

Recipe Overview

Reading Time
20 minutes

Executable Code
No

Difficulty

Introducing our FAIRification framework

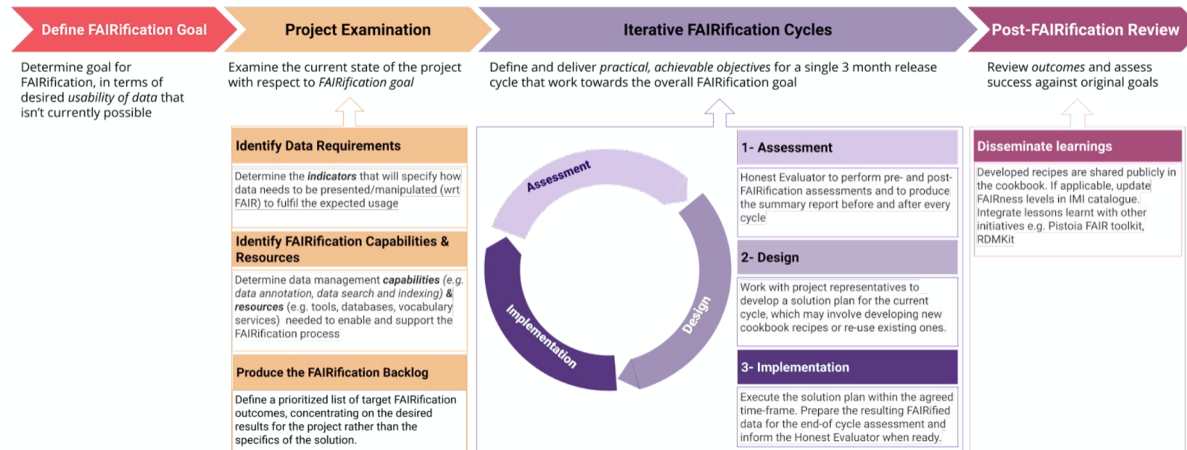
Recipe Type
Experience Report / Applied Example

Audience
Everyone

Maturity Level & Indicator
 not applicable




Cite me with FCB079

<https://w3id.org/faircookbook/FCB079>






Credit and citability: because all contributions matters!

Recipe Overview

-  Reading Time
30 minutes
-  Executable Code
No
-  Difficulty
🔥🔥🔥

Unique, persistent identifiers

-  Recipe Type
Background information
-  Audience
Principal Investigator, Data Manager, Data Scientist
-  Maturity Level & Indicator
[F+MM-1.1C] [F+MM-1.2C]




















Cite me with FCB006

<https://w3id.org/faircookbook/FCB006>



CreDiT
attribution ontology



Name	Orcid	Affiliation	Type	Elixir Node	Credit Role
 Andrea Splendiani		Novartis AG			Conceptualization
 Alasdair J G Gray		Heriot Watt University			Writing - Original Draft
 Chris Evelo		University of Maastricht			Writing - Original Draft
 Egon Willighagen		University of Maastricht			Writing - Original Draft
 Philippe Rocca-Serra		University of Oxford			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

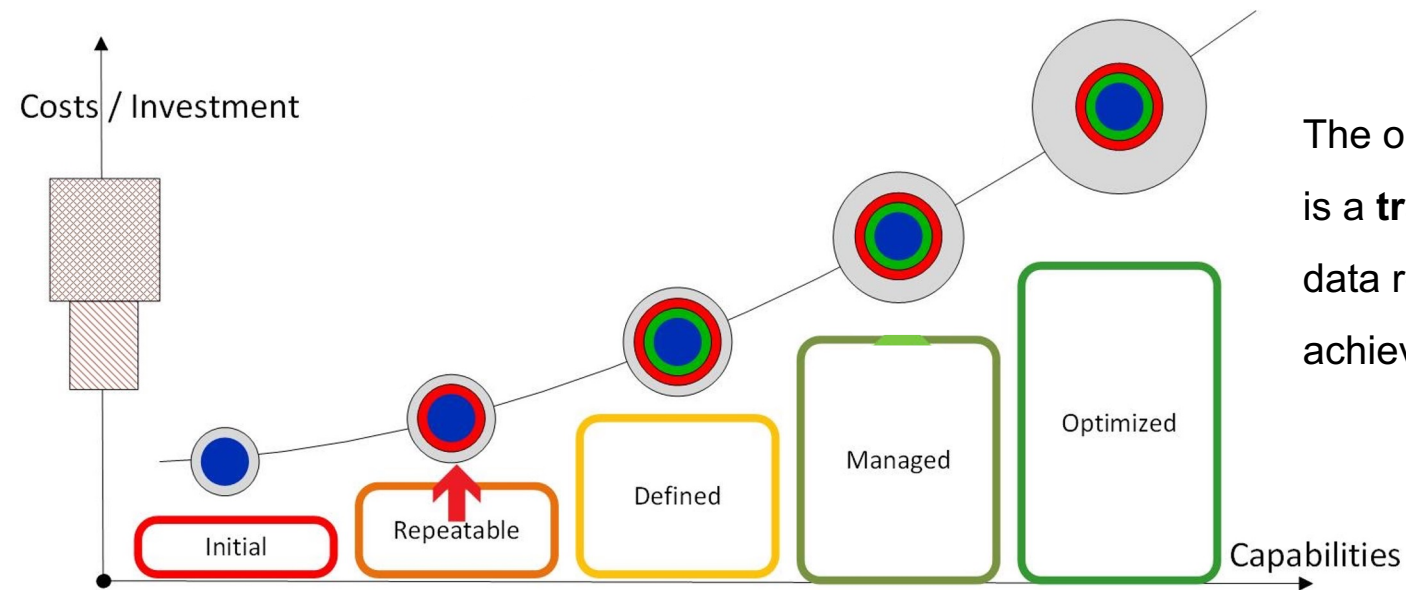


RESET

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	● ● ○ ○ ○

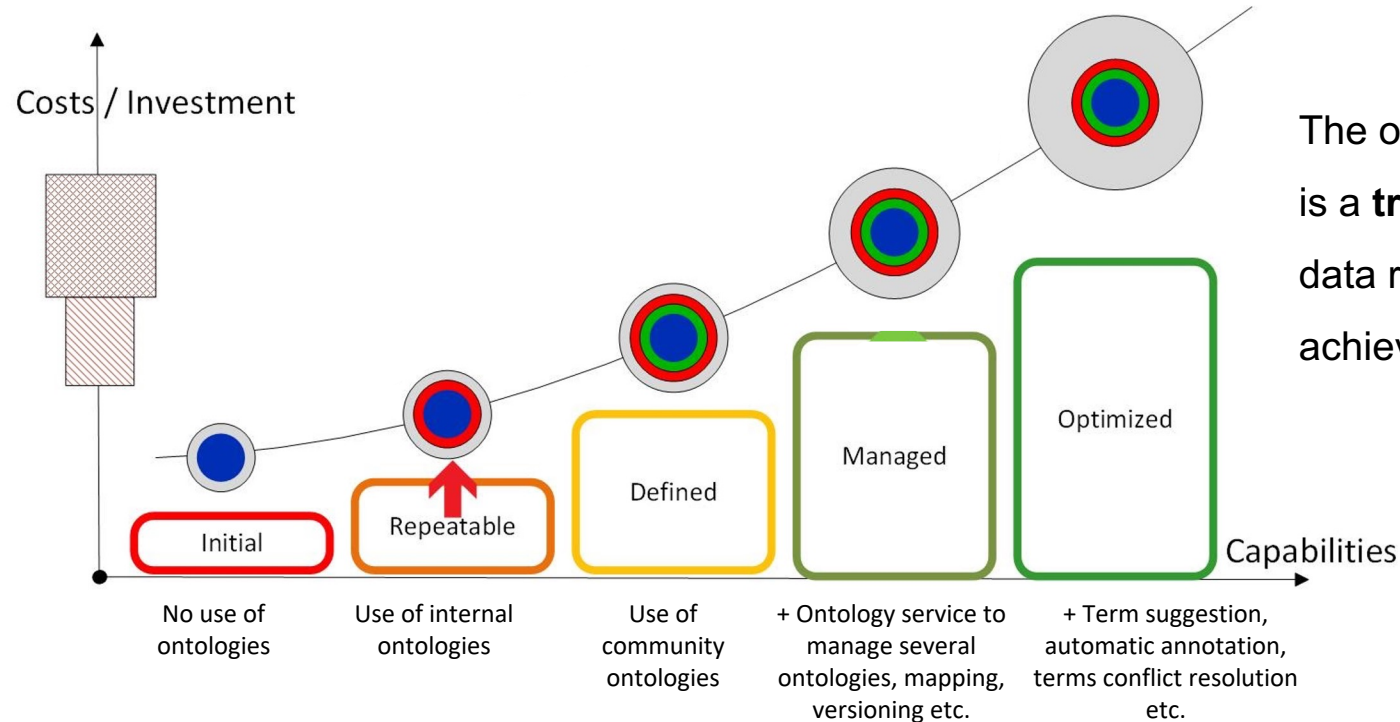
<https://faircookbook.elixir-europe.org/content/search-wizard.html>

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

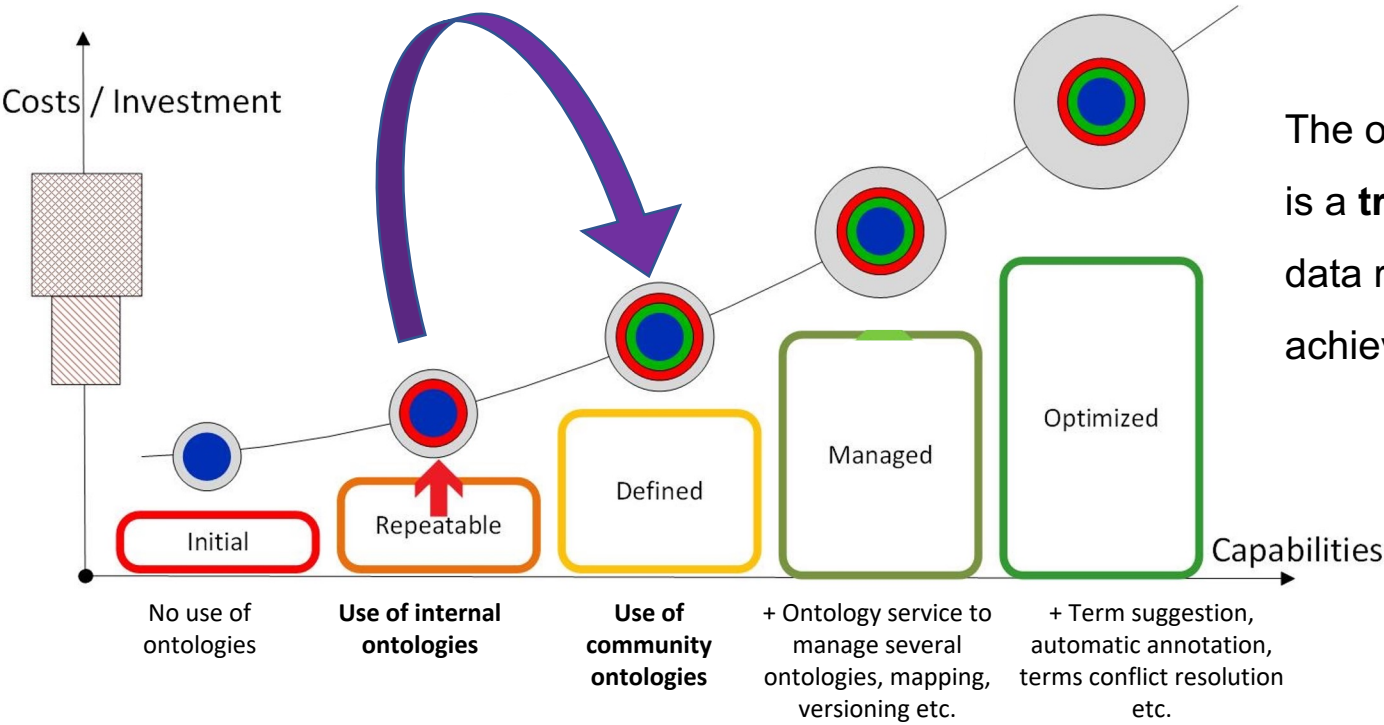
Capability needed to improve data reuse: the ontology example



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





FAIRcookbook

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**

SPRINGER NATURE

© 2023 Springer Nature Limited

<https://faircookbook.elixir-europe.org>

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

FAIRcookbook

(Life Science focused)



Recommended by
European funders,
embedded and used
by research
infrastructures

FAIRsharing.org

(All disciplines)



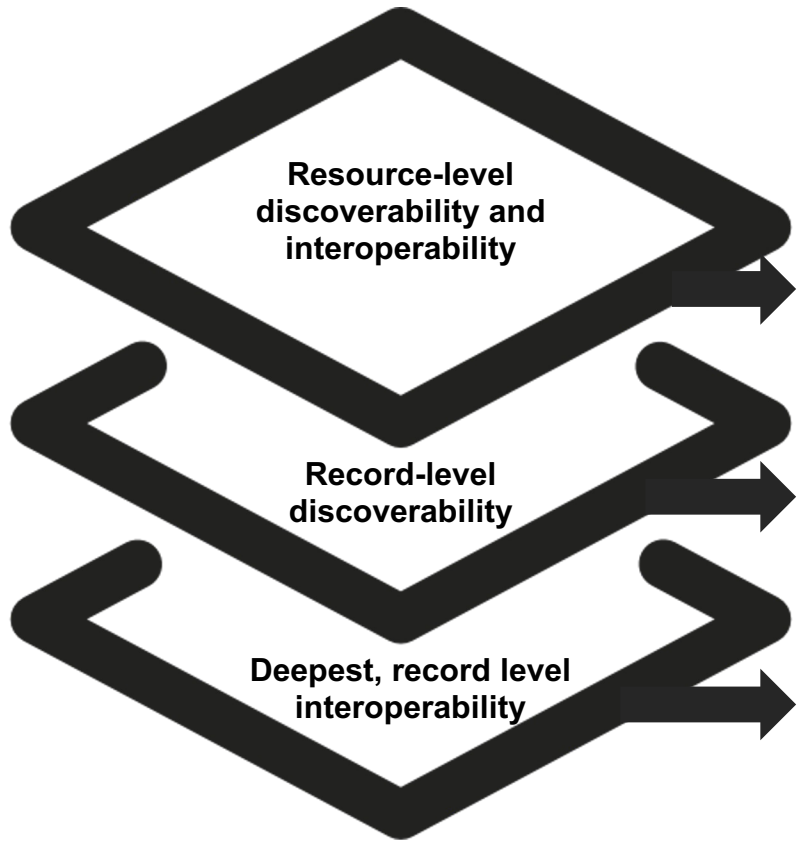
Annex 4

Horizon Europe

Work Programme 2023-2024



Standards for different purposes



A database or among databases

W3C
Data Catalog Vocabulary (DCAT)

Datasets in a database

Schema.org
DublinCore

Datasets and data reuse

SRAXml VO AAO MIAME EC number DOI PURL
 OBI CHEBI MIRIAM MIQAS URL LSID IVC
 FASTA PATO ENVO MIX MIQAS REMARK ORCID Handle IVC
 DICOM MOD TEDDY MIGEN REMARK
 SBRML XAO BTO MIAPE CONSORT RRID InChI ...
 3 OMOP SEDML DO PRO IDO ARRIVE MIASE MISFISHIE ...

Standards for data and metadata - a dynamic ecosystem

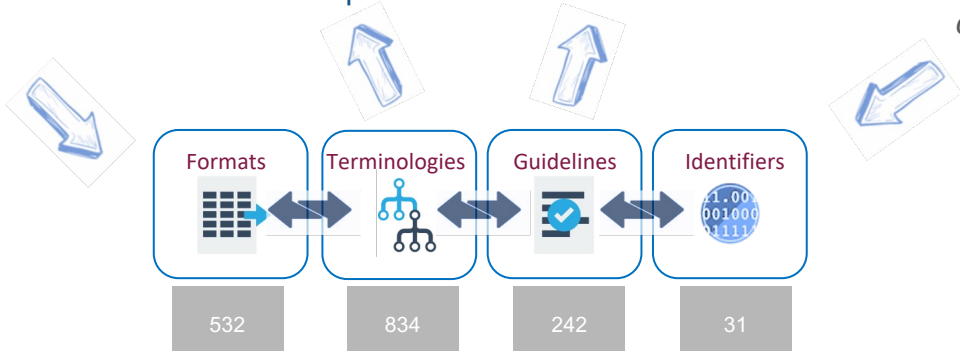
FAIRsharing.org

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

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

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

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

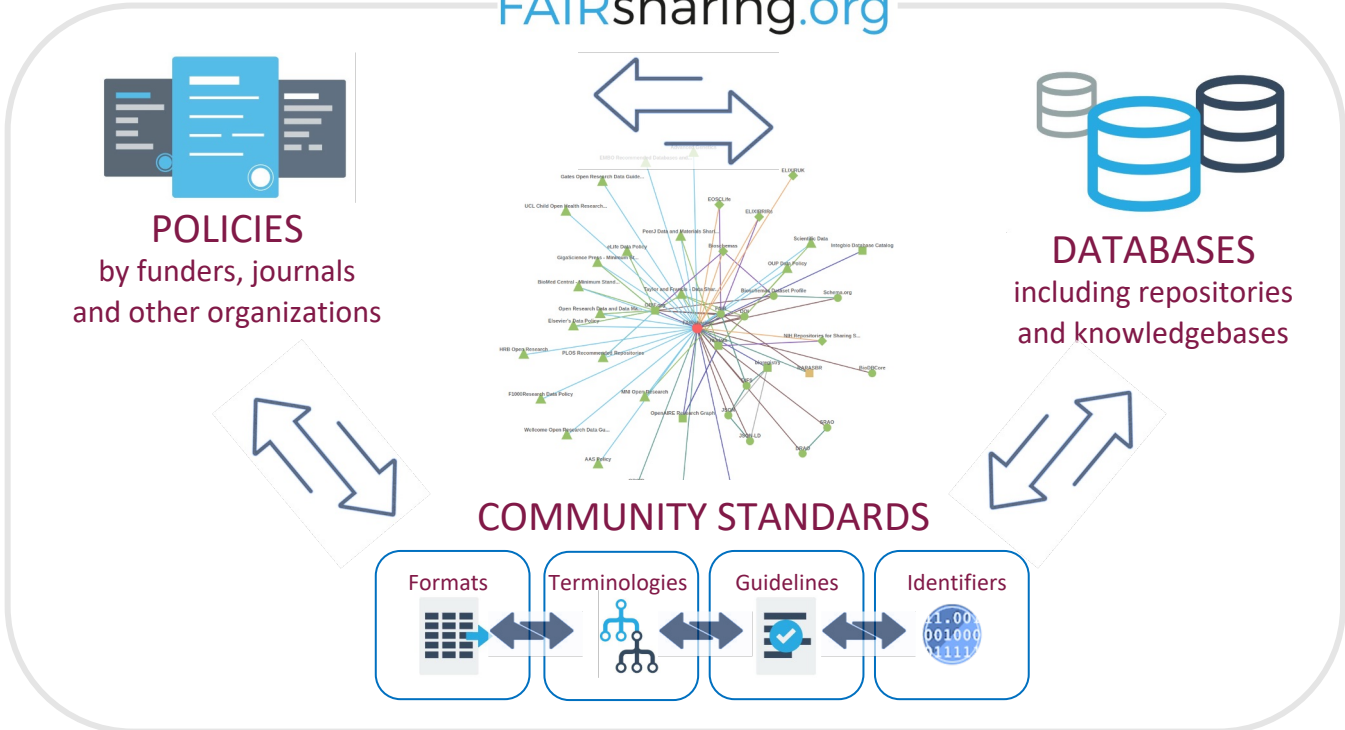


1653 Standards

As of 4 April 2023

An informative and educational resource, and a service

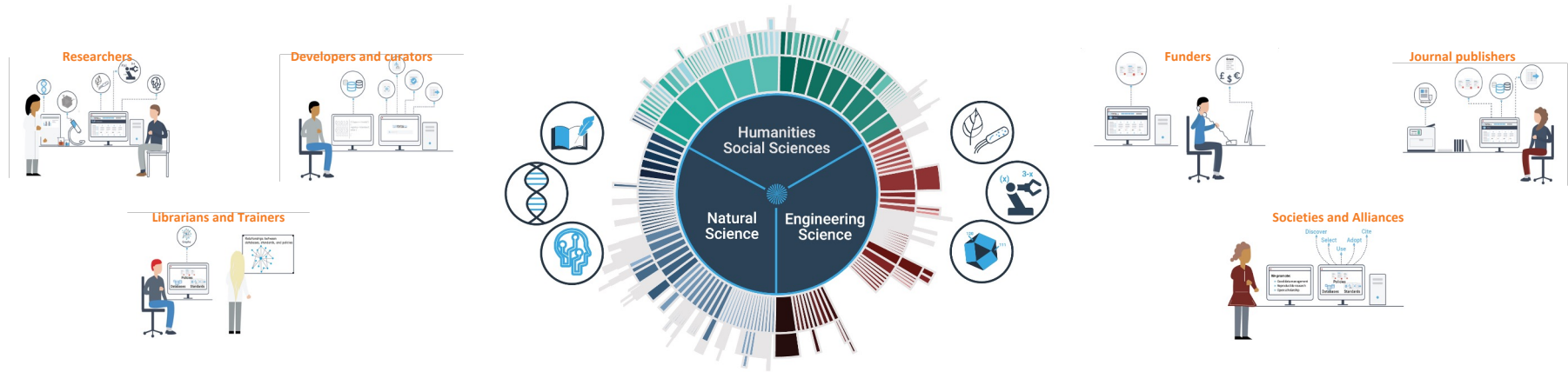
FAIRsharing.org



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

FAIRsharing mission and how we deliver it

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*

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

1653 Standards

Terminology Artifact	834
Model/Format	532
Reporting Guideline	242
Identifier Schema	31

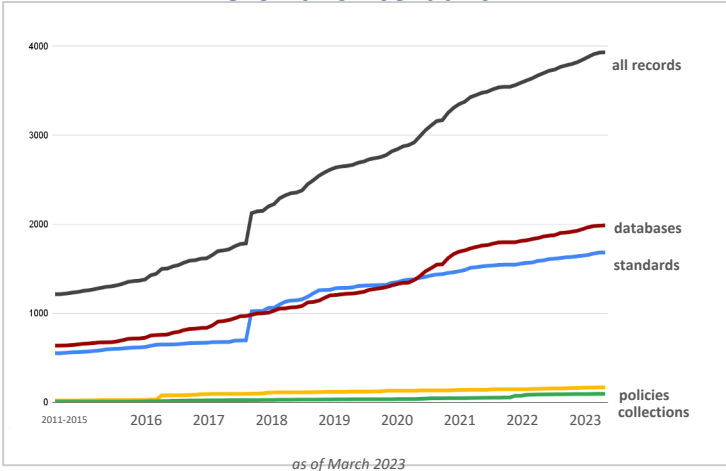
2000 Databases

Repositories	1034
Knowledgebases	825
Knowledgebase/Repositories	141

165 Policies

Journal	87
Funder	33
Project	14
Society	13

Growth since launch

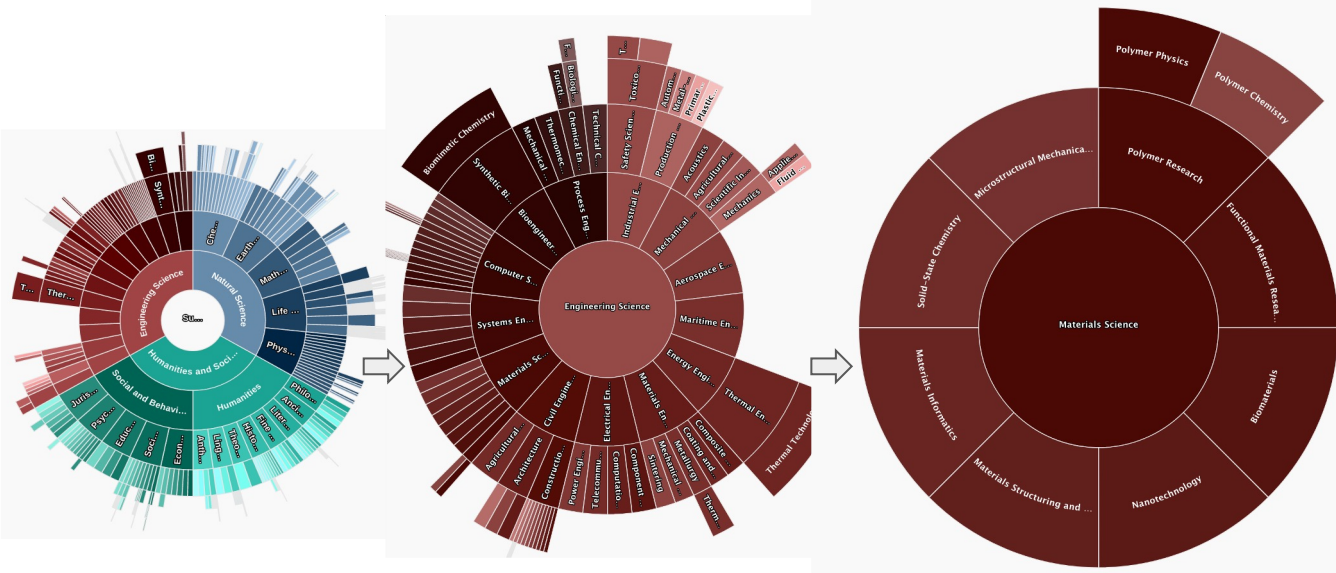
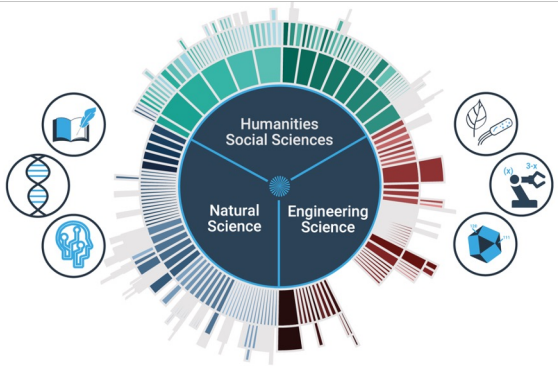


Disciplines

- ▶ Natural Science 8523
- ▶ Humanities and Social Science 706
- ▶ Subject Agnostic 303
- ▶ Engineering Science 1444

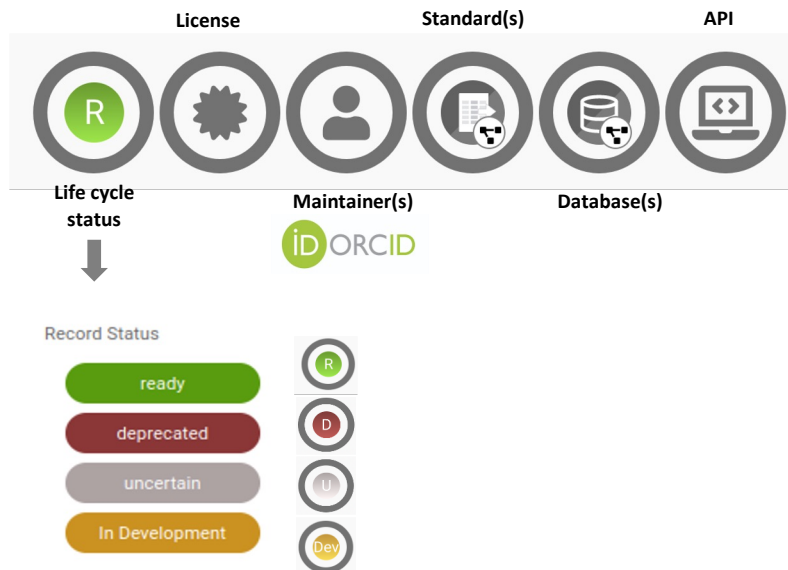
Look for resources relevant to material science

- ▶ 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!

Manually curated descriptions of each record,
by in house team



GENERAL INFORMATION

Matter Transfer Ontology (TRANSMAT)
 10.25504/FAIRsharing.hve0ac

Type Terminology artefact

Registry Standard

Description The Matter Transfer ontology is dedicated to matter transfer (eg O2, CO2, H2O) and mechanical properties of materials especially with respect to food, bacteria and packaging fields.

Homepage <https://www6.inrae.fr/cati-icat-atweb/Ontologies/Transmat>

Year of Creation 2015

Maintainers [lovinra](#)

Countries developing this resource [France](#)

Subjects Engineering Science Materials Science

Domains Food

Taxonomic Range Not applicable

User Defined Tags None

[VIEW RELATION GRAPH](#)

How to cite this record

FAIRsharing.org: TRANSMAT; Matter Transfer Ontology, DOI: [10.25504/FAIRsharing.hve0ac](https://doi.org/10.25504/FAIRsharing.hve0ac), Last Accessed: Tuesday, April 4th 2023, 19:01

FAIRsharing's FAIR content powers 3rd party tools

Findability



Sitemap.xml, JSON



Markup with Schema.org for search indexes



DOI unique persistent identifiers for each record



ORCID for author credit and authentication

Accessibility



read/write REST API



read OAI-PMH

Interoperability



JSON markup



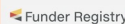
Standardized semantics



Cross-links to or import from records in other registries



ROR for organizations (*ongoing*)



FundRef for funders (*ongoing*)

Reusability



CC BY 4.0 license




JSON export

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:

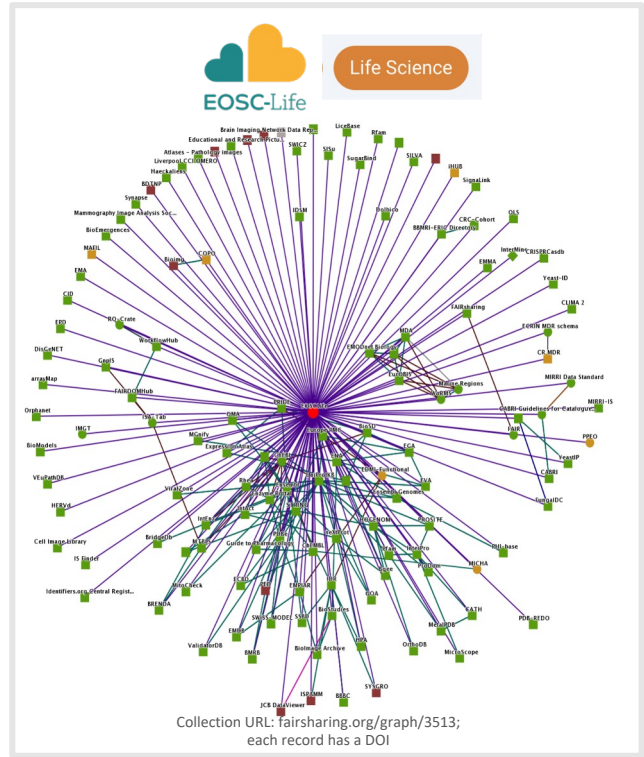
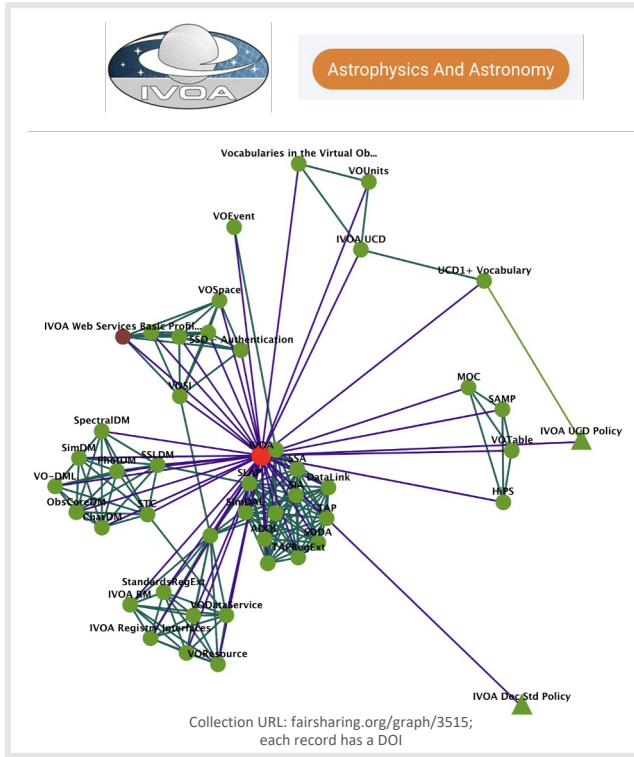
- creation of data management plans
- enrichment of guidance and training material
- assessment of FAIRness 



FAIR Metrics and Data Quality Task Force

FAIR Assessment Tools: Towards an “Apples to Apples” Comparisons

Showing organizations' "FAIR profiles" matters!



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


Record Status



- ready
- deprecated
- uncertain
- In Development

Record Relationships

- Collects
- Related To
- Deprecates
- Deprecated by
- Implements
- Implemented by
- Shares Data With
- Collected by
- Extended by
- Extends

GENERAL INFORMATION

 **Monarch Disease Ontology (MONDO)**

 [10.25504/FAIRsharing.b2979t](https://doi.org/10.25504/FAIRsharing.b2979t) 

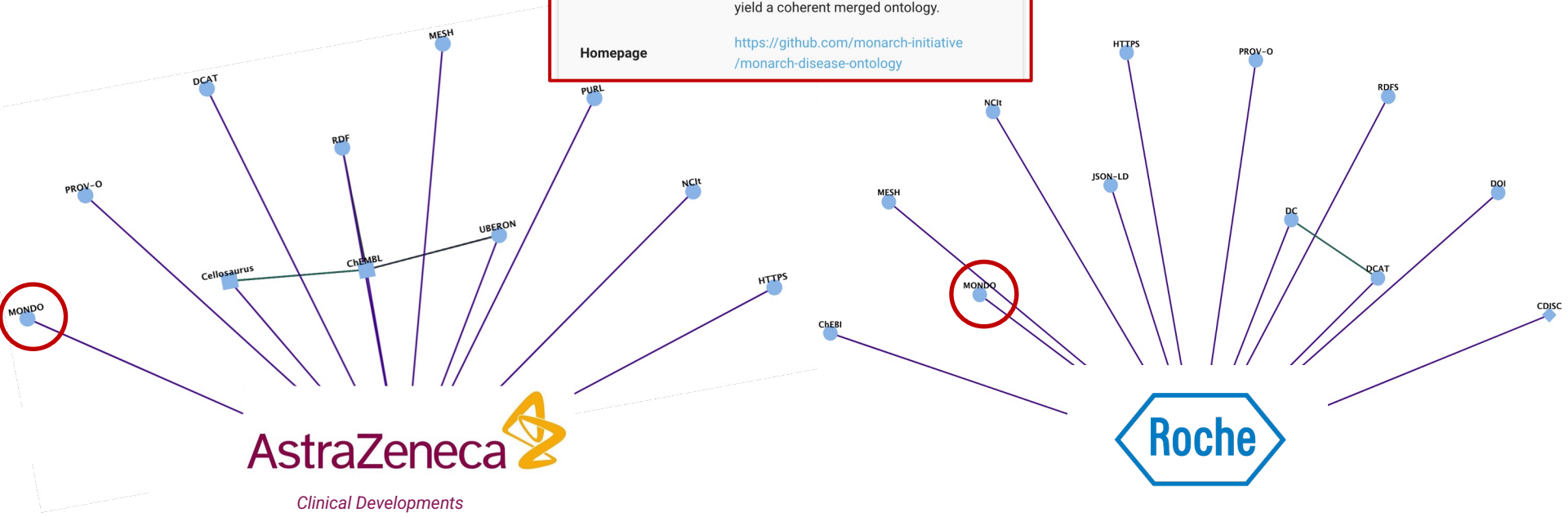
Type Terminology artefact

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

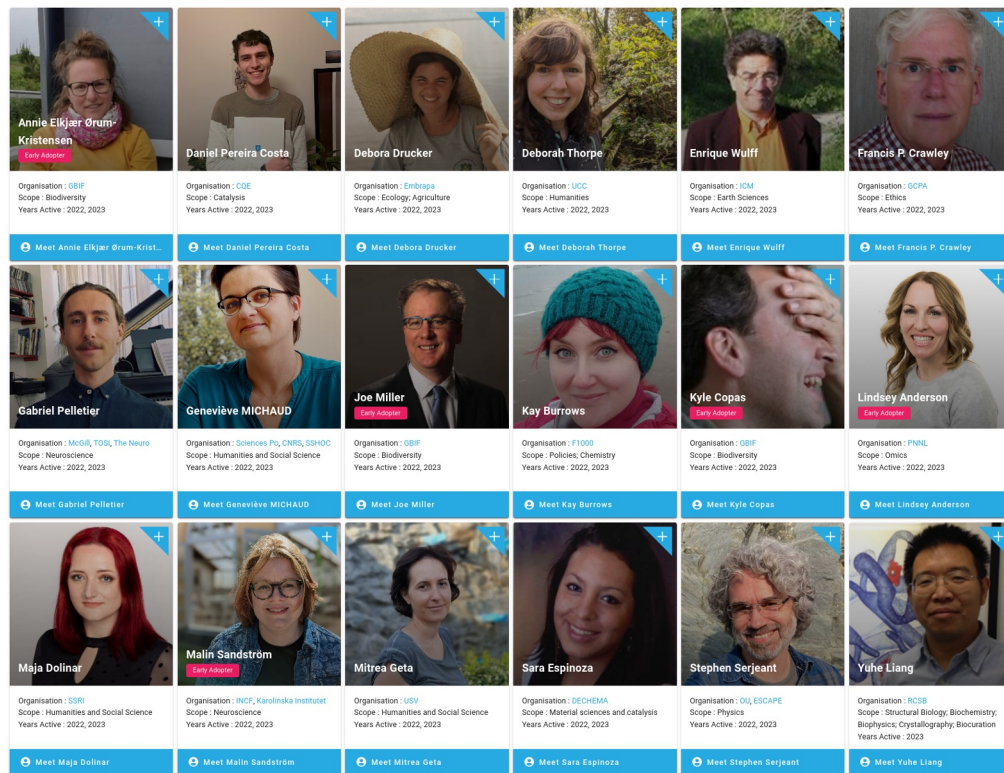


AstraZeneca 

Clinical Developments

Roche

Engaging the expertise within community is essential



FAIRsharing Community Champion Programme

Domain experts who

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



Building educational material collaboratively



Educational

Infographics and factsheets by the FAIRsharing Team and our Community Champions, covering content and functionalities for all stakeholders

<p>FAIRsharing in a nutshell</p>	<p>FAIRsharing content: standards overview</p> <p>Definitions, types, benefits and need-to-know facts</p>	<p>FAIRsharing content: databases overview</p> <p>Definitions, types, benefits and need-to-know facts</p>	<p>FAIRsharing content: policies overview</p> <p>Definitions, types, benefits and need-to-know facts</p>	<p>FAIRsharing content: collections overview</p> <p>Coming Soon</p>
<p>FAIRsharing for you: researchers</p>	<p>FAIRsharing for you: journal publishers</p>	<p>FAIRsharing for you: funders</p>	<p>FAIRsharing for you: societies and alliances</p> <p>Coming Soon</p>	<p>FAIRsharing for you: librarians and trainers</p> <p>Coming Soon</p>
<p>FAIRsharing for you: developers and curators</p> <p>Coming Soon</p>	<p>FAIRsharing functionality: searching</p> <p>Coming Soon</p>	<p>FAIRsharing functionality: create a view using collection</p> <p>Coming Soon</p>	<p>FAIRsharing behind the scenes: capturing the resource life cycle</p> <p>Coming Soon</p>	<p>FAIRsharing behind the scenes: the FAIRness of FAIRsharing</p> <p>Coming Soon</p>

fairsharing.org/educational

FAIRsharing content: standards overview

Core to research data management good practices

FAIRsharing promotes the **value** of standards, the backbone of the **FAIR Principles**

Guides users to **discover**, select, and use standards with more **confidence**

Assists producers to make their standards more **visible**, and more widely adopted



FAIRsharing describes standards for all data, including datasets, software, materials and other digital objects

Standards ...

Are a **collectively agreed-upon** set of requirements, specifications, guidelines or characteristics that can be used for the **description, structure, harmonisation, citation, sharing, and/or preservation** of all kinds of data*

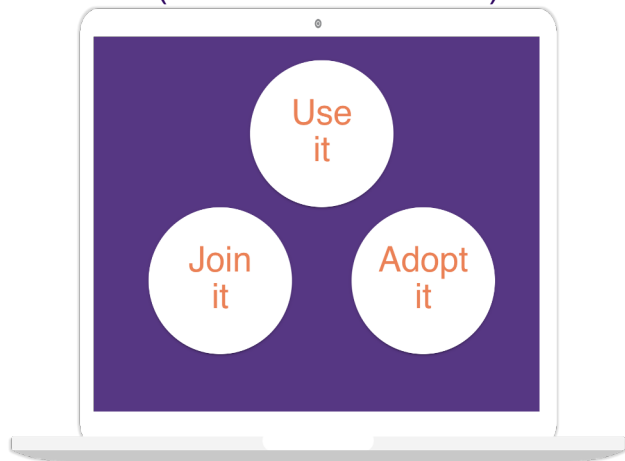
Help **machines** with computational accessibility, **interoperability**, and use of data with little/no human intervention; enable humans to understand and reuse **data at scale**

FAIRsharing categorises standards with four types:

- 1 Reporting guidelines**
Outline in narrative form the necessary and sufficient information that should be reported about data, such as in itemised, prescriptive checklists; or the features and behaviours that should be followed, such as in general guiding principles
- 2 Models and formats**
Define the representation of information for use by machines; these range from conceptual models to transmission formats, facilitating data retrieval and exchange between systems
- 3 Terminology artefacts**
Add an interpretive, semantic layer for use by machines and humans; these range from controlled vocabularies (lists of terms, often with definitions) to ontologies (complex hierarchical groupings), providing unambiguous identification of concepts and aiding data querying
- 4 Identifier schemata**
Are formal systems to identify information in a unique, machine-readable way; these persistent identifiers (PIDs), minted by recognised registries, build reliable and long-lasting links between data, people, organisations and infrastructures



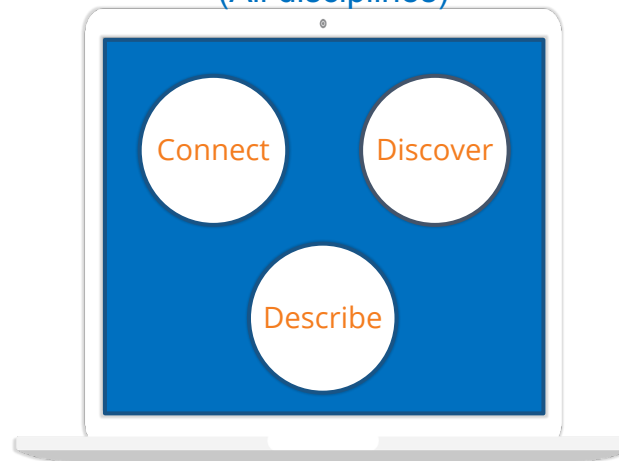
(Life Science focused)



faircookbook.elixir-europe.org
faircookbook-ed@elixir-europe.org



(All disciplines)



fairsharing.org
contact@fairsharing.org