



FAIRification is a Team Sport

Susanna-Assunta Sansone, PhD

ORCID: 0000-0001-5306-5690

Twitter: @SusannaASansone

ELIXIR

Interoperability Platform ExCo



elixir-europe.org

Professor of Data Readiness

Associate Director, Oxford e-Research Centre



Group: datareadiness.eng.ox.ac.uk



Founding

Academic Editor

scientific data

SPRINGER NATURE

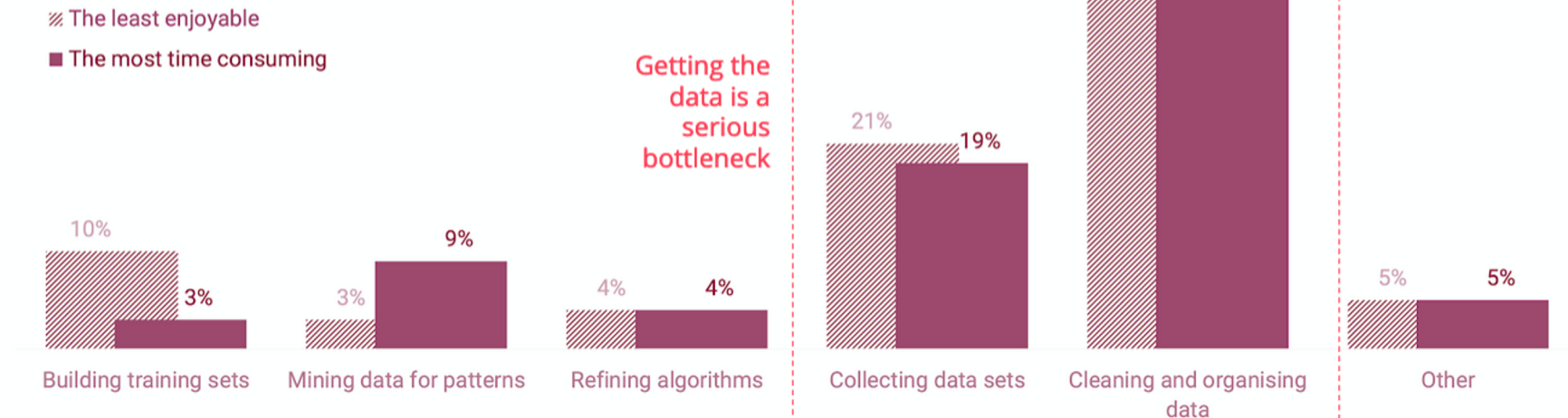
nature.com/sdata

Science is about continuity

Discoveries are made using shared data and this requires data that are:

- Cited and stored to be discoverable
- Retrievable and structured in standard format(s)
- Richly described to be understandable

Data preparation accounts for about 80% of the work of data scientists



Use of data at scale by humans and machines

SCIENTIFIC DATA

SPRINGER NATURE

OPEN

Comment: The FAIR Guiding Principles for scientific data management and stewardship

SUBJECT CATEGORIES

- » Research data
- » Publication characteristics

Mark D. Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Boume, Jildau Bouwman, Anthony J Brookes, Tim Clark, Mercè Crosas, Ingrid Dillo, Olivier Dumon, Scott Edmunds, Chris T Evelo, Richard Finkers, Alejandra Gonzalez-Beltran, Alasdair J G Gray, Paul Groth, Carole Goble, Jeffrey S. Grethe, Jaap Heringa, Peter A.C. 't Hoen, Rob Hooft, Tobias Kuhn, Ruben Kok, Joost Kok, Scott J. Lusher, Maryann E. Martone, Albert Mons, Abel L. Packer, Bengt Persson, Philippe Rocca-Serra, Marco Roos, Rene van Schaik, Susanna-Assunta Sansone, Erik Schultes, Thierry Sengstag, Ted Slater, George Strawn, Morris A. Swertz, Mark Thompson, Johan van der Lei, Erik van Mulligen, Jan Velterop, Andra Waagmeester, Peter Wittenburg, Katherine Wolstencroft, Jun Zhao, and Barend Mons

SCIENTIFIC DATA | 3:160018 | DOI: 10.1038/sdata.2016.18

Access & Citations

451k

Article Accesses

3173

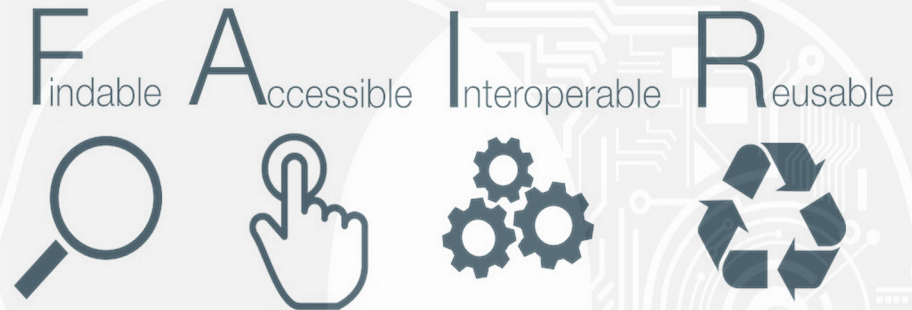
[Web of Science](#)

3993

[CrossRef](#)



Attention Score



Globally unique and persistent **identifiers**

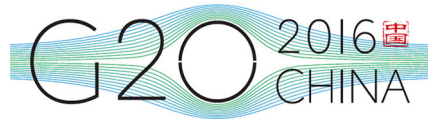
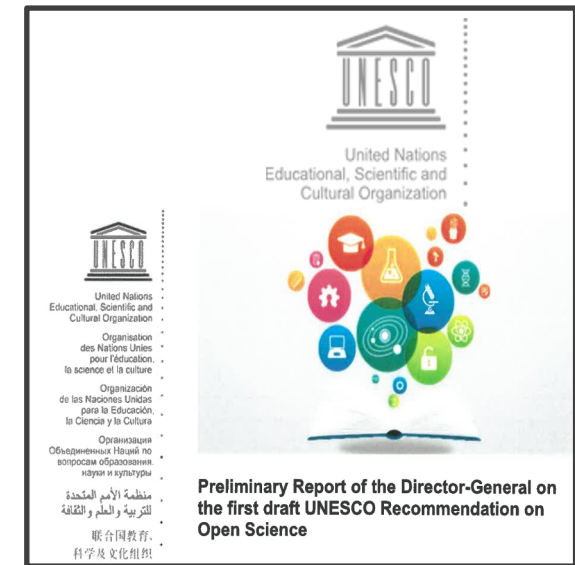
Detailed **provenance**

Community defined **terminologies**

Terms of **access**

Terms of **use**

Community defined descriptive **metadata**



www.fair-access.net.au

doi.org/10.1787/25186167

ark:/48223/pf0000374837

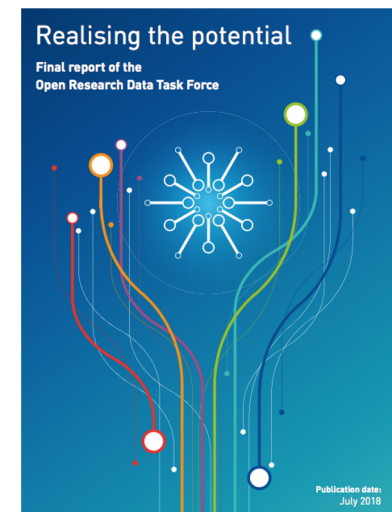
FAIR has aligned the broad community around common guidelines



www.turing.ac.uk/research/impact-stories/changing-culture-data-science



doi.org/10.7486/DRI.tq582c863



www.gov.uk/government/publications/open-research-data-task-force-final-report



doi.org/10.2777/986252

FAIR Principles: **a continuum of features, attributes and behaviours**

Fairness does not mean everyone gets the same. Fairness means everyone gets what they need.

Rick Riordan

quote fancy

FAIRsharing.org



Recommended
by national and
international
funders, e.g.:

 FAIR cookbook



Horizon Europe (HORIZON)

Programme Guide

Version 1.2

04 October 2021



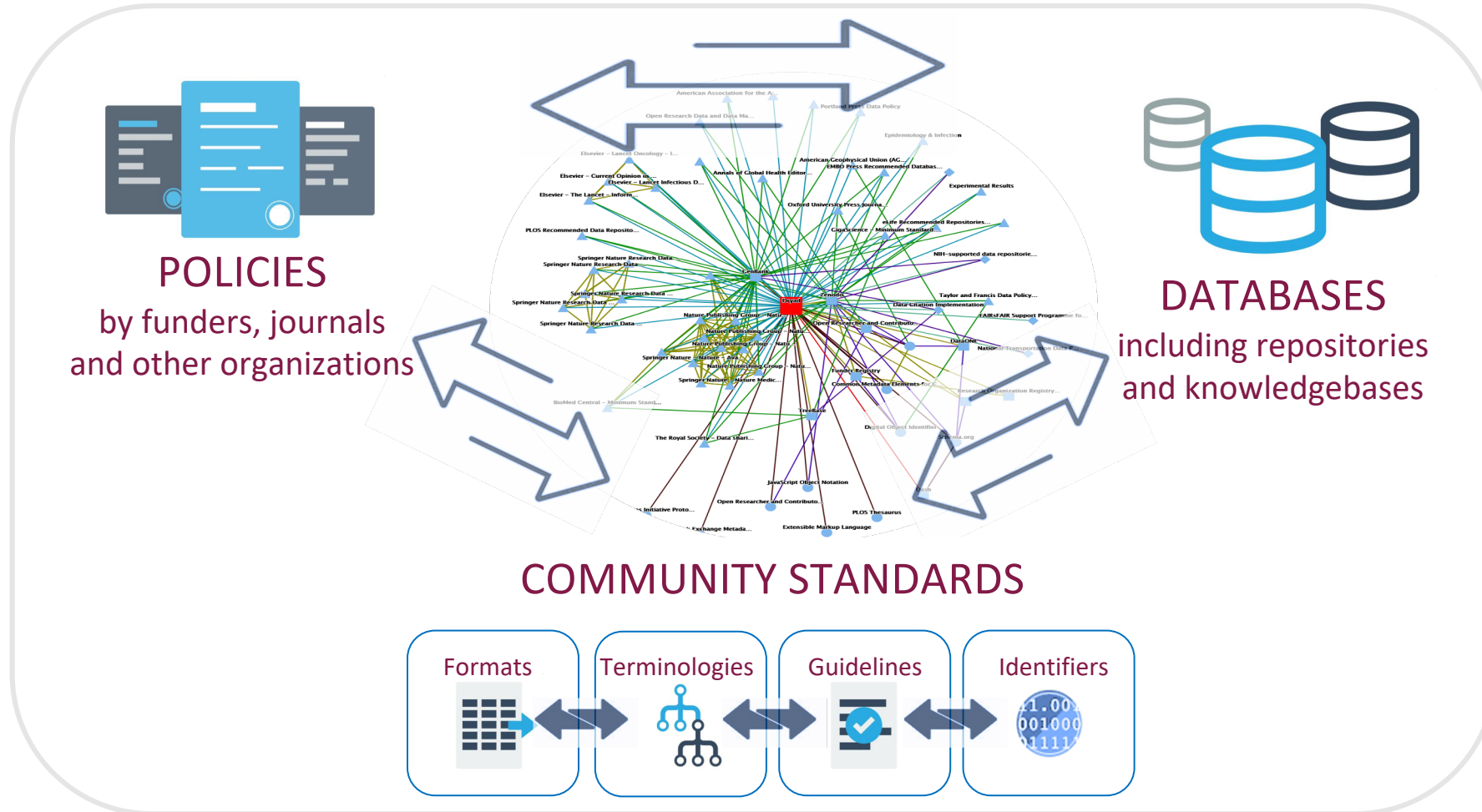
IMI2 project guidelines for
open access to publications
and research data



efpia

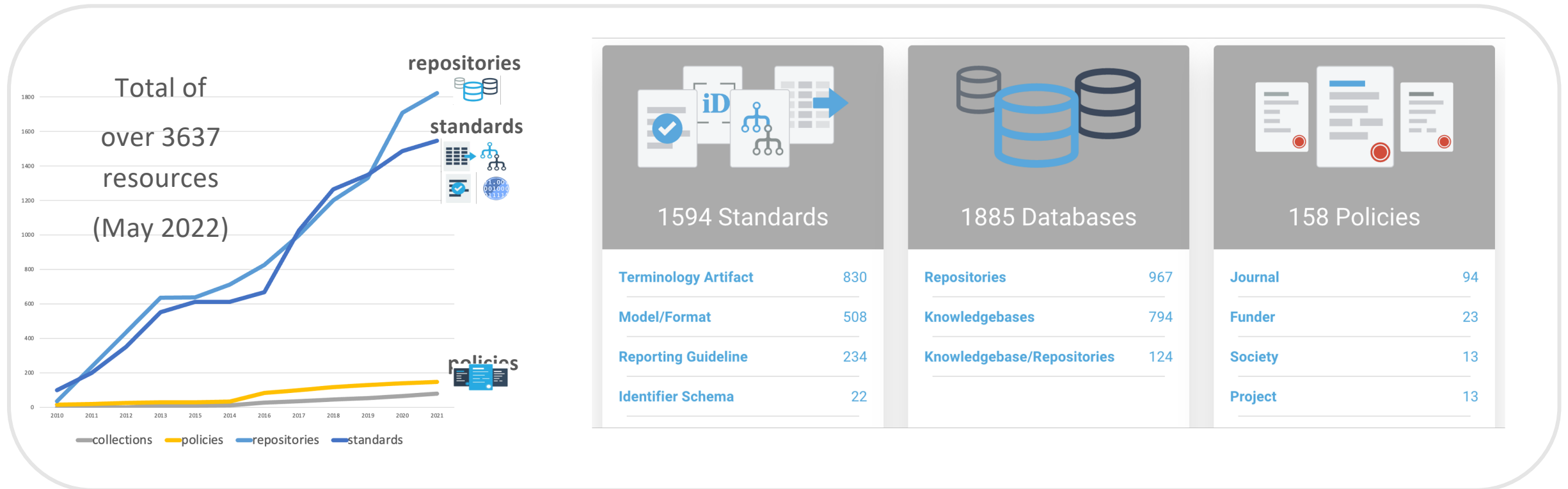
European Federation of Pharmaceutical
Industries and Associations

An informative and educational resource



It provides **curated, community-vetted descriptions** and **relationship graphs** of standards, databases and policies in all disciplines

Promoting databases, standards, policies



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

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

Detailed descriptions that help you decide what resource is right for your case

GENERAL INFORMATION

ChEMBL (ChEMBL)

doi 10.25504/FAIRsharing.m3jtpg

Type Knowledgebase

Registry Database

Description ChEMBL is an open, manually-curated, large-scale bioactivity database containing information from medicinal chemistry literature. It brings together chemical, bioactivity and genomic data to aid the translation of genomic information into effective new drugs. Information regarding the compounds tested (including their structures), the biological or physicochemical assays performed on these and the targets of these assays are recorded in a structured form, allowing users to address a broad range of drug discovery questions.

Homepage <https://www.ebi.ac.uk/chembl>

Year of Creation 2009

Maintainers [agaulton](#)

Countries developing this resource Austria, Belgium, Croatia, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Lithuania, Luxembourg, Malta, Montenegro, Netherlands, Norway, Portugal, Slovakia, Spain, Sweden, Switzerland, United Kingdom

Subjects Drug Discovery, Biochemistry, Genomics, Medicinal Chemistry

Domains Chemical Structure, Drug, Chemical Entity, Bioactivity, Small Molecule, Genome, Tissue

Taxonomic Range All

User Defined Tags None



[VIEW RELATION GRAPH](#)



How to cite this record

FAIRsharing.org: ChEMBL; ChEMBL, DOI: [10.25504/FAIRsharing.m3jtpg](https://doi.org/10.25504/FAIRsharing.m3jtpg), Last Edited: Thursday, March 17th 2022, 10:51, Last Editor: allysonlister, Last Accessed: Thursday, March 17th 2022, 10:51



Publication for citation

ChEMBL: towards direct deposition of bioassay data.: Mendez D,Gaulton A,Bento AP,Chambers J,De Veij M,Felix E,Magarinos MP,Mosquera JF,Mutowo P,Nowotka M,Gordillo-Maranon M,Hunter F,Junco L,Mugumbate G,Rodriguez-Lopez M,Atkinson F,Bosc N,Radoux CJ,Segura-Cabrera A,Hersey A,Leach AR (2018) [10.1093/nar/gky1075](https://doi.org/10.1093/nar/gky1075)

Record ID 1561 | Record created at Tuesday, November 4th 2014, 15:23 | Record updated at Thursday, March 17th 2022, 10:51

License



Life cycle status



Record Status



Standard(s)



Maintainer(s)



Database(s)



API



Show the relations between databases, standards, and policies

GENERAL INFORMATION



ChEMBL (ChEMBL)

doi: 10.25504/FAIRsharing.m3jtpg

Type

Knowledgebase

Registry

Database

Description

ChEMBL is an open, manually-curated, large-scale bioactivity database containing information from medicinal chemistry literature. It brings together chemical, bioactivity and genomic data to aid the translation of genomic information into effective new drugs. Information regarding the compounds tested (including their structures), the biological or physicochemical assays performed on these and the targets of these assays are recorded in a structured form, allowing users to address a broad range of drug discovery questions.

Homepage

<https://www.ebi.ac.uk/chembl>

Year of Creation

2009

Maintainers

[agaulton](#)

Countries developing this resource

Austria, Belgium, Croatia, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Luxembourg, Malta, Montenegro, Netherlands, Norway, Portugal, Slovakia, Spain, Sweden, Switzerland, United Kingdom

Subjects

Drug Discovery Biochemistry Genomics Medicinal Chemistry

Domains

Chemical Structure Drug Chemical Entity Bioactivity Small Molecule Genome Tissue

Taxonomic Range

All

User Defined Tags

None

[VIEW RELATION GRAPH](#)


How to cite this record

FAIRsharing.org: ChEMBL; ChEMBL, DOI: [10.25504/FAIRsharing.m3jtpg](https://doi.org/10.25504/FAIRsharing.m3jtpg), Last Edited: Thursday, March 17th 2022, 10:51, Last Editor: allysonlister, Last Accessed: Thursday, March 17th 2022, 10:51



Publication for citation

ChEMBL: towards direct deposition of bioassay data.: Mendez D, AP,Chambers J,De Veij M,Felix E,Magarinos MP,Mosquera JF,Mut M,Gordillo-Maranon M,Hunter F,Junco L,Mugumbate G,Rodriguez-L F,Bosc N,Radoux CJ,Segura-Cabrera A,Hersey A,Leach AR (2018) [10.1093/nar/nkz001](https://doi.org/10.1093/nar/nkz001)

Record ID 1561 | Record created at Tuesday, November 4th 2014, 15:23 | Record updated at Thursday, March 17th 2022, 10:51

RELATED STANDARDS (18)

RELATED DATABASES (16)



BindingDB database of measured binding affinities

ChEMBL related to BindingDB data



Information system for G protein-coupled receptors

ChEMBL related to Information system for G protein-coupled receptors



IUPHAR/BPS Guide to PHARMACOLOGY

ChEMBL related to IUPHAR/BPS Guide to PHARMACOLOGY

IN POLICIES (9)



Wellcome Open Research Data Guidelines

Wellcome Open Research Data Guidelines recommends ChEMBL



F1000Research Data Policy

F1000Research Data Policy recommends ChEMBL



HRB Open Research

HRB Open Research recommends ChEMBL

Visualize these relations to guide you

GENERAL INFORMATION

ChEMBL (ChEMBL)
 doi: 10.25504/FAIRsharing.m3jtpg

Type: Knowledgebase
 Registry: Database

Legend and configuration

The graph's centre is shown in red. The registry of each record is as follows:

- Database
- Standard
- Policy
- Collection

Click on any point to re-draw the graph with that point as the centre. Click on the centre to view the record.

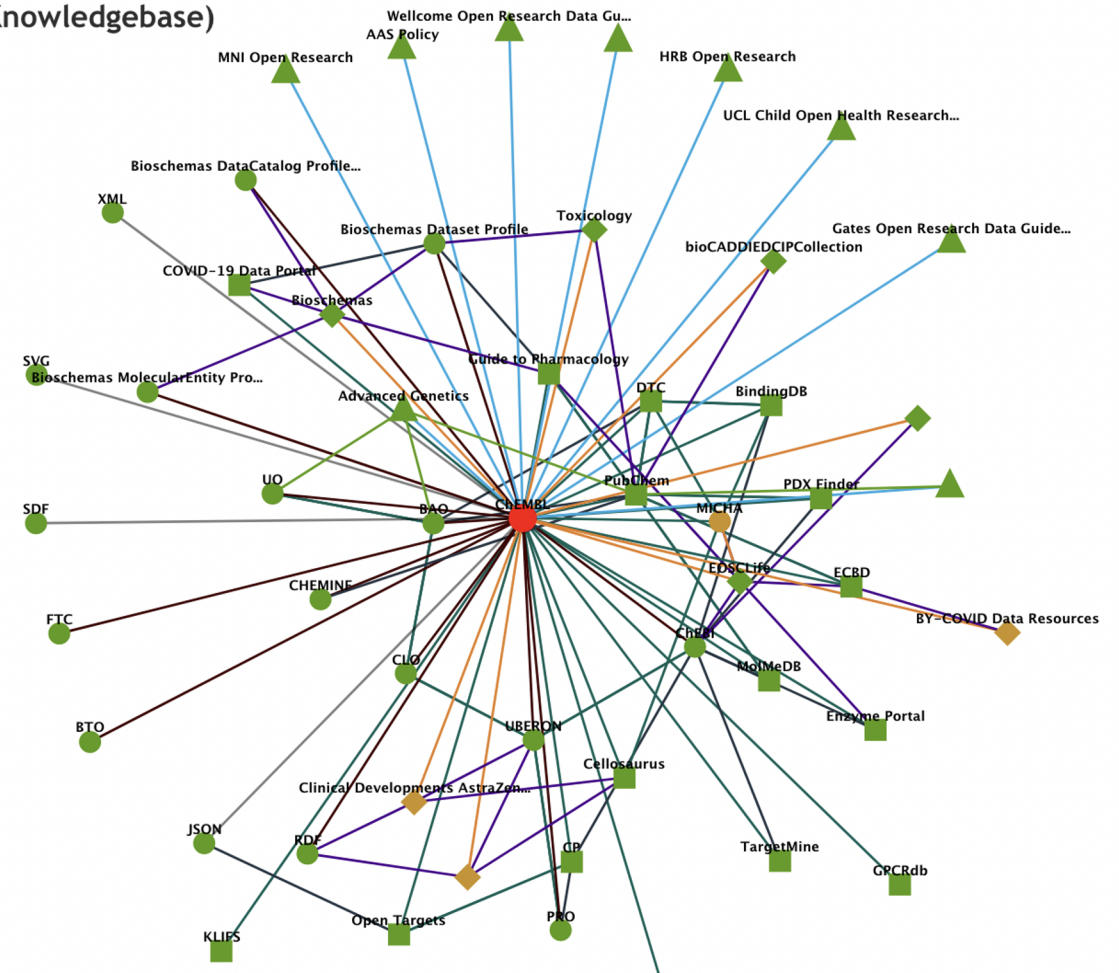
Record Status

- ready
- deprecated
- uncertain
- In Development

Record Relationships

- Related To
- Implements
- Implemented by
- Recommended by
- Recommends
- Collected by
- Collects

FAIRsharing Network Graph
 ChEMBL
 (Database/Knowledgebase)

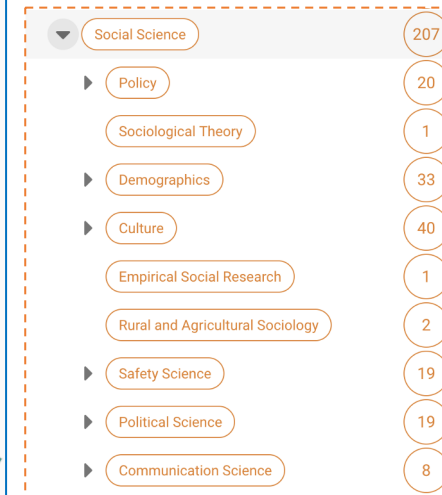
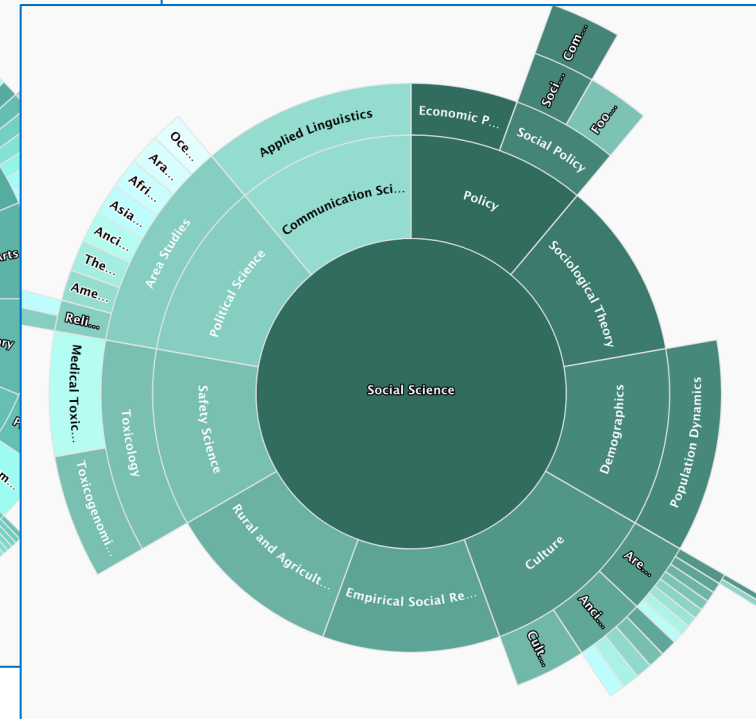
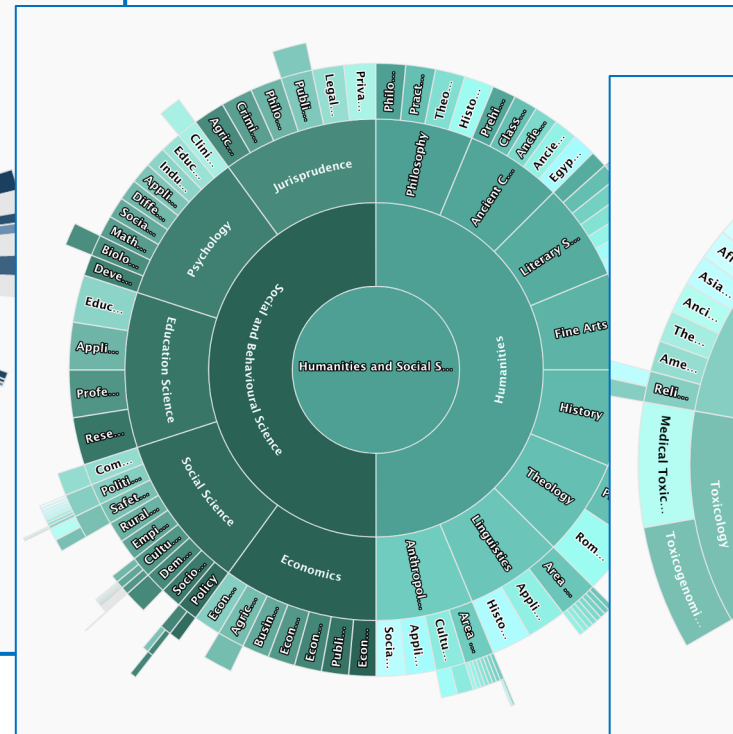
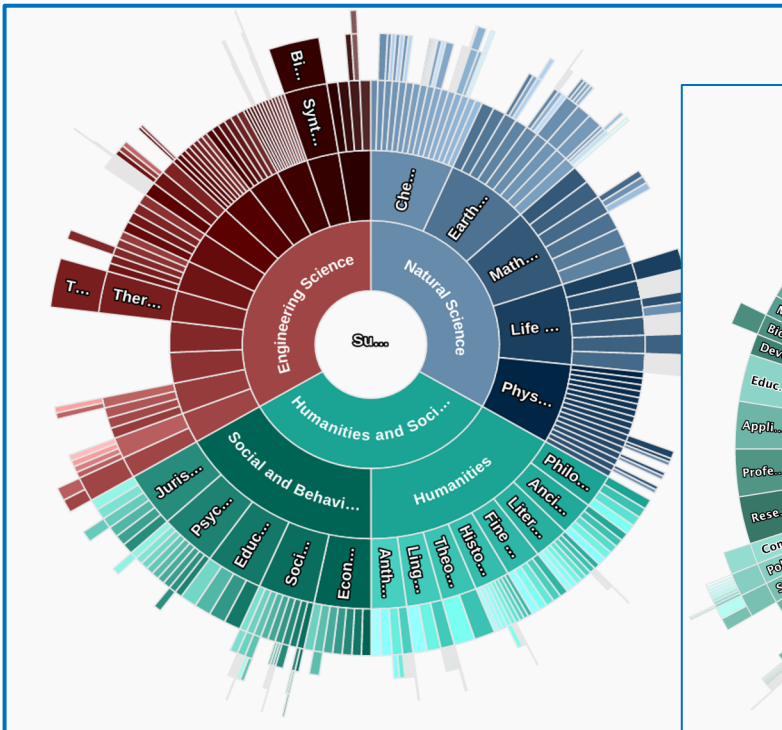
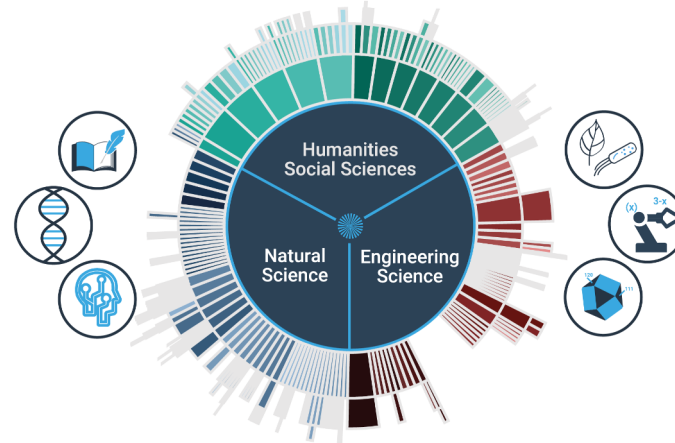
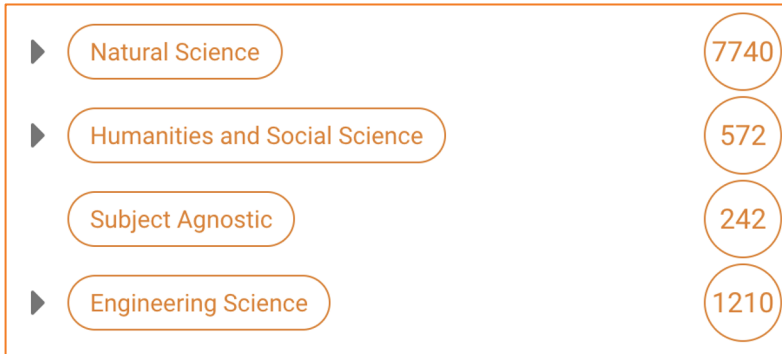


Help you search by subject

Classification is powered by our **Subject Ontology of 436 terms**

URL: fairsharing.org/browse/subject

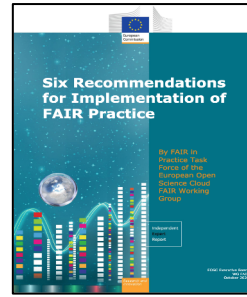
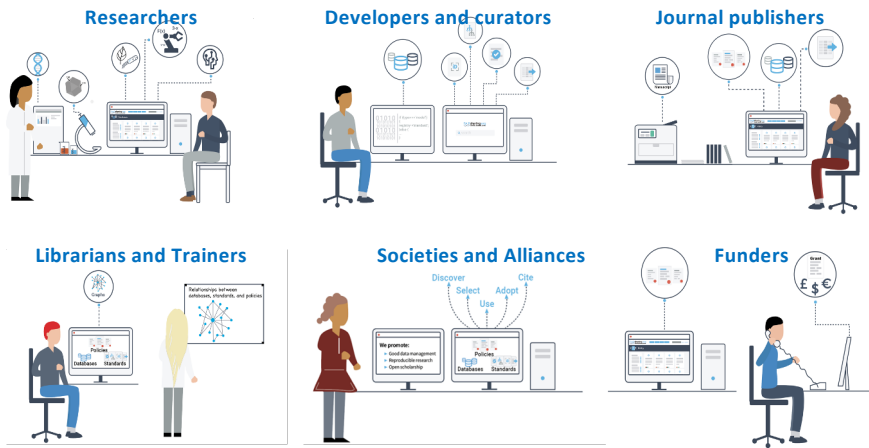
URL: github.com/FAIRsharing/subject-ontology



FAIRsharing: working with and for all stakeholders

Users, adopters and collaborators include:

Users from all stakeholder groups



WILEY
SPRINGER
NATURE



F1000
FACULTY OF 1000



EMBOpress



(GIGA)²
SCIENCE

OXFORD
UNIVERSITY PRESS

THE
ROYAL SOCIETY
PUBLISHING



Taylor & Francis

CAMBRIDGE
UNIVERSITY PRESS



Jisc



COS
CENTER FOR
OPEN SCIENCE



GO FAIR

FORCE11
The Future of Research Communications and e-Scholarship



An endorsed output of the
FAIRsharing WG (since 2015):



A recommended resource in EOSC reports



A WG (since 2015) in:



Current operational Team

- **Allyson Lister**, Content and Community Lead
- **Milo Thurston**, Technical Lead
- **Ramon Granell**, Data Enrichment & Quality Manager
- **Delphine Dauga**, Data Curator Manager
- *Hiring in progress*, Web Developer
- **Dominique Batista**, Research Software Engineer
- **Philippe Rocca-Serra**, Co-Founder
- **Susanna-Assunta Sansone**, PI and Founder
- *and many collaborators and contributors!*

Executive Advisors

- **Varsha Khodiyar**, Independent expert
- **Chris Graf**, Springer Nature
- **David Carr**, Independent expert
- **Robert Hanisch**, Director, NIST Office of Data & Informatics
- **Peter McQuilton**, FAIRsharing Founding Member, GSK

Stakeholder Advisors

- **Amye Kenall**, VP of Publishing and Product, Research Square
- **Adam Leary**, Oxford University Press
- **Catriona MacCallum**, Hindawi
- **Dagmar Meyer**, European Research Council, Executive Agency
- **Dominic Fripp**, JISC, UK
- **Emma Ganley**, Protocols.io
- **Geraldine Clement-Stoneham**, Medical Research Council
- **Helena Cousijn**, DataCite
- **Iain Hrynaskiewicz**, PLoS
- **Imma Subirats**, FAO of the United Nations
- **Kiera McNiece**, Cambridge University Press
- **Luiz Olavo Bonino**, GO-FAIR
- **Marina Soares E Silva** and **Sarah Callaghan**, Elsevier
- **Michael Ball**, Biotechnology and Biological Sciences Research Council
- **Mike Huerta**, NIH National Library of Medicine
- **Molly Cranston** and **Guillaume Wright**, F1000Research
- **Nick Everitt** and **Matthew Cannon**, Taylor and Francis
- **Scott Edmunds**, GigaScience, Oxford University Press
- **Simon Hodson**, CODATA
- **Theo Bloom**, British Medical Journal
- **Thomas Lemberger**, EMBO Press
- **Wei-Mun Chan**, eLife
- **Sowmya Swaminathan**, Springer Nature

FAIRsharing.org



Recommended
by national and
international
funders, e.g.:

 FAIR cookbook



Horizon Europe (HORIZON)

Programme Guide

Version 1.2

04 October 2021



IMI2 project guidelines for
open access to publications
and research data



efpia

European Federation of Pharmaceutical
Industries and Associations

A 'live' collection of recipes to help achieve FAIR data management



FAIRcookbook

FAIR Cookbook

FOREWORD

Introduction

Ethical values of FAIR

Glossary

RECIPES

Findability

Accessibility

Interoperability

Reusability

Infrastructure

Assessment

Applied examples

ABOUT

Community

Contribute

Platform

Powered by [Jupyter Book](#)

- 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**

faircookbook.elixir-europe.org

Recipes that cover all aspects of FAIRness

FAIRcookbook

Search this book...

FAIR Cookbook

FOREWORD

- Introduction
- Ethical values of FAIR
- Glossary

RECIPES

- Findability
- Accessibility
- Interoperability
- Reusability
- Infrastructure
- Assessment
- Applied examples

ABOUT

- Community
- Contribute
- Platform

Powered by **Jupyter Book**

The recipes

The FAIR Cookbook organizes the recipes according to the FAIR elements, audience type (your role), reading time, and level of difficulty. The FAIR Cookbook is a 'live resource'; recipes are added and improved, iteratively, in an open manner, therefore bear with us if several sections are work in progress! Below there are links to some key recipes, click on them to explore their content; otherwise use the main menu on the left hand side to browse all the current recipes.

F Findability

Exemplar recipes:

- Unique, persistent identifiers
- Search engine optimization

→ More about Findability

A Accessibility

Exemplar recipes:

- Transferring data with SFTP
- Downloading data with Aspera

→ More about Accessibility

I Interoperability

Exemplar recipes:

- Selecting terminologies and ontologies
- Creating a metadata profile

→ More about Interoperability

R Reusability

Exemplar recipes:

- Data licenses
- Declaring data's permitted uses

→ More about Reusability

Infrastructure

Applied Examples

Assessment

Over 70 recipes (May 2022) released and many more in progress!

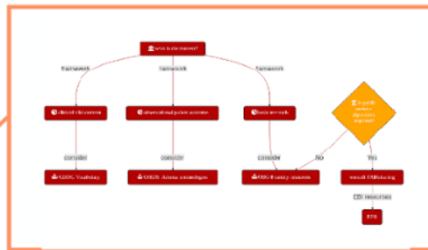
Examples in the life sciences, including omics, pre-clinical and clinical areas

But not limited to it!

Anatomy of a recipe

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

Step by step process

Guidelines, process, description

References

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

Citability of the recipes and credit to the authors

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



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

Maturity level and indicators of FAIRness

Recipe Overview

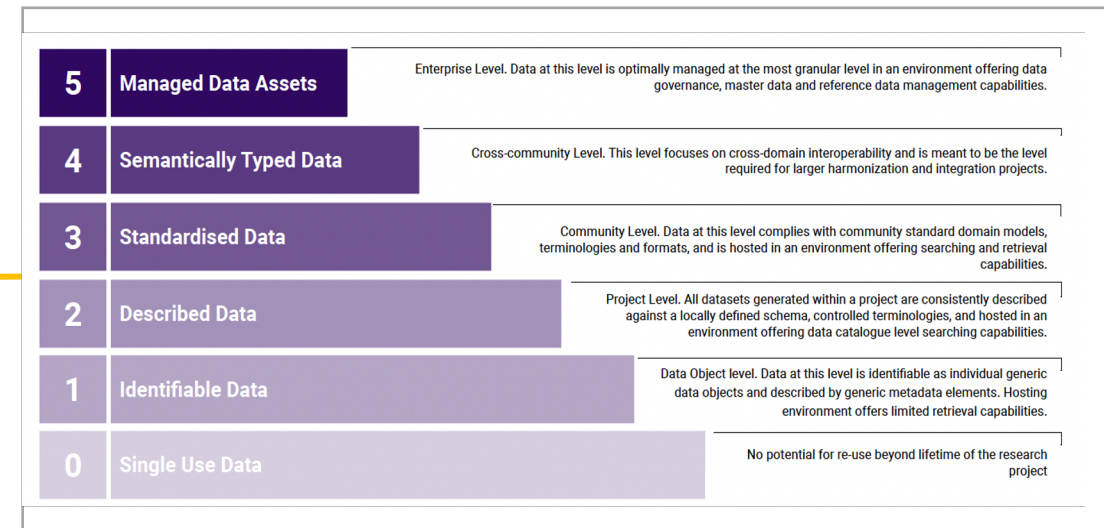
- 🕒 Reading Time
30 minutes
- ▶️ Executable Code
No
- 💧 Difficulty
👉👉👉👉

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



We are working to tag the recipes with a 'dataset maturity model'

It show the level of FAIRness you can reach by applying a specific recipe to improve a dataset

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

FAIR Cookbook: turning knowledge into recipes

Almost 100 life sciences professionals, researchers and data managers

FAIRplus partners
Industry + Academia



ELIXIR Nodes
represented



The roadmap, **your** steps



For everyone



adopt it

**Start using the
FAIR Cookbook**

Help us to help you,
this is a user-
oriented resource

**Contribute
according to
your needs**

Create and review,
or signpost gaps and
needs

leverage it

**Recommend it
in your guidance
and training
material**

use it

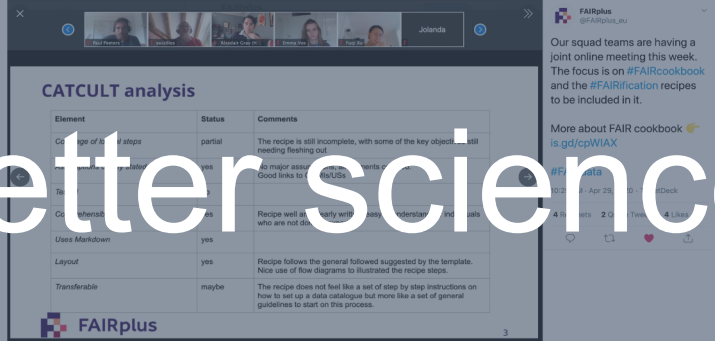
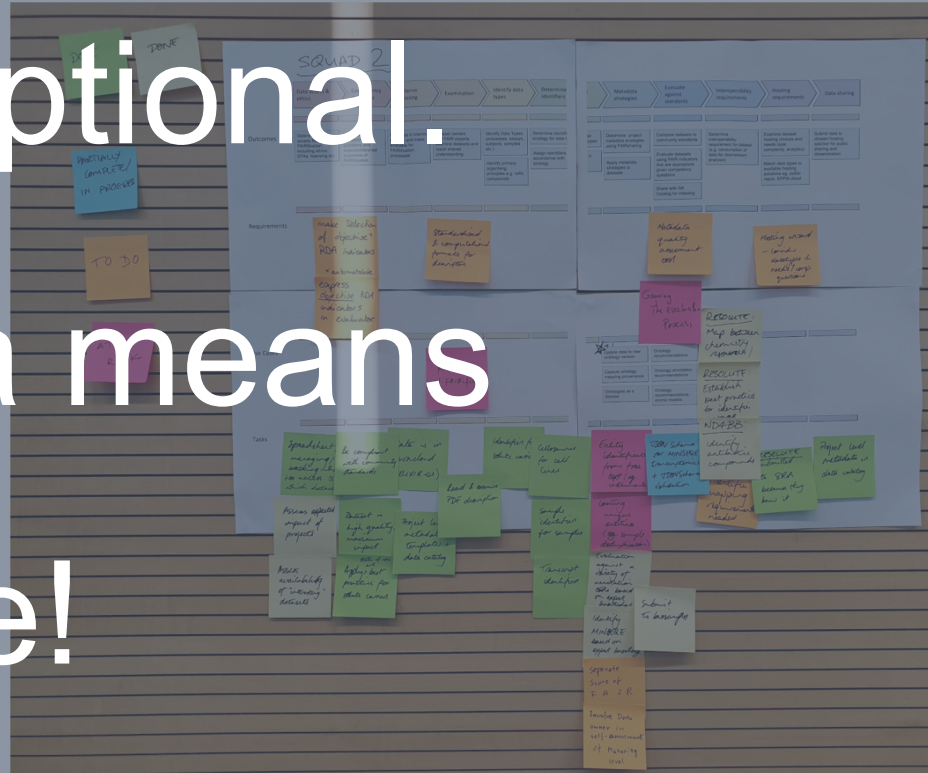
**Join a network
of FAIR experts**

Share and broaden
your expertise,
joining our journey

join it

**Benefit from the
power of the
FAIR community**

Become a recognised
expert, enlarge your
network



FAIRification is a team sport,

it takes a village,

but it is no longer optional.

Because better data means

better science!

Our squad teams are having a joint online meeting this week. The focus is on #FAIRcookbook and the #FAIRification recipes to be included in it.

More about FAIR cookbook is.gd/cpWIAx

#FAIRplus

Element	Status	Comments
Clarity of steps	partial	The recipe is still incomplete, with some of the key objects still missing. Flushing out.
Reproducibility	yes	Good links to relevant resources.
Clarity of instructions	yes	Recipe well written and easy to follow. Includes the steps who are not clear.
Use of diagrams	yes	Recipe follows the general flow diagram suggested by the template. Nice use of flow diagrams to illustrate the recipe steps.
Transferable	maybe	The recipe does not feel like a set of step by step instructions on how to set up a data catalogue but more like a set of general guidelines to start on this process.

FAIRplus