

# Planning for FAIR data

Joy Davidson, DCC

With contributions from DCC colleagues

# Overview of session

- What is FAIR data?
- How do research data management, FAIR and Open intersect?
- Drivers for FAIR
- Practical messages about making data FAIR

# What is FAIR data?

## Findable

F1. (meta)data are assigned a globally unique and eternally persistent identifier.

F2. data are described with rich metadata.

F3. (meta)data are registered or indexed in a searchable resource.

F4. metadata specify the data identifier.

## Interoperable

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.

## Accessible

A1. (meta)data are retrievable by their identifier using a standardized communications protocol.

A1.1 the protocol is open, free, and universally implementable.

A1.2 the protocol allows for an authentication and authorization procedure, where necessary.

A2. metadata are accessible, even when the data are no longer available.

## Reusable

R1. meta(data) have 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 their provenance.

R1.3. (meta)data meet domain-relevant community standards.

**Want to learn more about FAIR? Try FAIR-Aware tool for researchers and support staff.**

**<https://fairaware.dans.knaw.nl/>**

Data Archiving and Networked Services  
**DANS**



Your first step towards your FAIR data(set)

Do you work with data? Are you looking to make it future-proof? The FAIR Principles can help you.

These principles stand for the Findability, Accessibility, Interoperability and Reusability of data(sets). Applying these principles to your data(set) will help others to find, cite and reuse your data more easily.

FAIR-Aware helps you assess your knowledge of the FAIR Principles, and better understand how making your data(set) FAIR can increase the potential value and impact of your data.

The tool is discipline-agnostic, making it relevant to any scientific field. You can use this tool at any point during your research before depositing your data(set) in a data repository. It is also good to keep in mind that many FAIR-related decisions can already be made in the research planning phase, so you may want to use FAIR-Aware early on to help you make those decisions. Also, if you are a trainer, you can use FAIR-Aware to assess the knowledge of FAIR of your course participants.

The self-assessment consists of 10 questions with additional guidance texts to help you become more aware of what you can do to make your data(set) as FAIR as possible. The assessment will take between 10-30 minutes, after which you will receive an overview of your awareness level and additional tips on how you can further improve your FAIR skills.

If you would like to use FAIR-Aware in your own training, you can find instructions on the trainer functionality here [🔗](#)  
You can contact the FAIR-Aware development team for any questions or comments via e-mail. ✉️

# What is Open Data?

ODI defines Open Data as those that anyone can access, use and share.

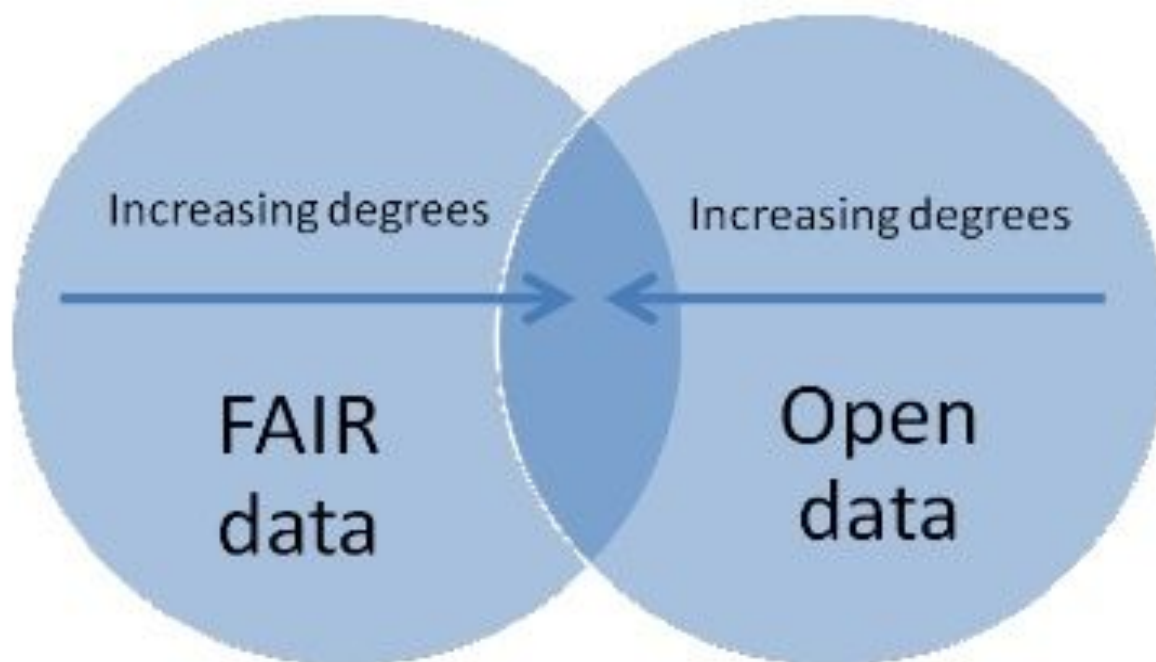
According to the ODI, open data must be licensed to make clear that anyone can use the data in any way they want, including transforming, combining, and sharing it with others, even for commercial purposes.

ODI provides a great introduction to all aspects of Open Data in their Open Data Essentials course.  
<http://accelerate.theodi.org/>

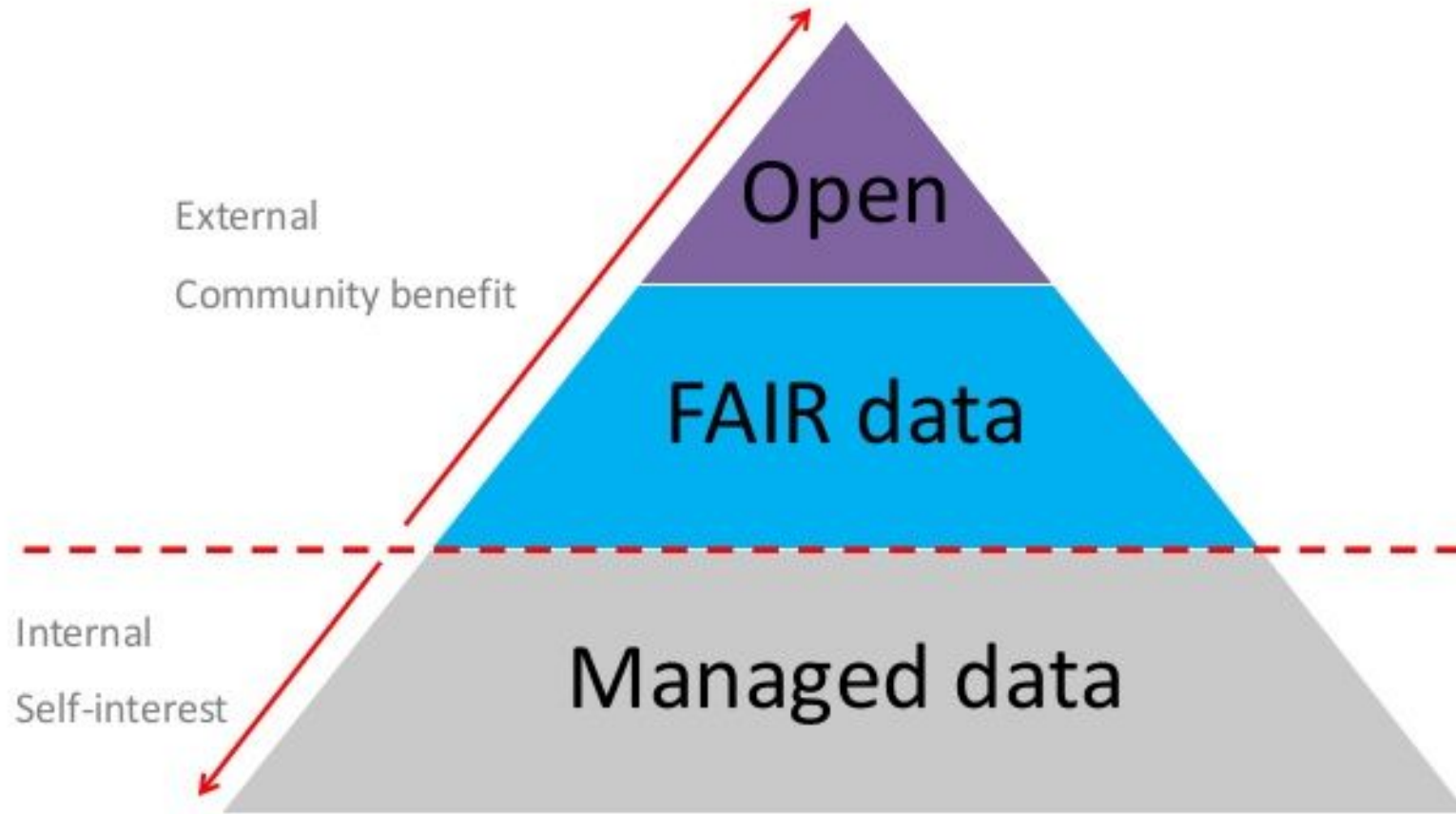


# Open and FAIR data are not the same thing!

- The greatest potential reuse comes when data are both FAIR and Open
- Align and harmonise FAIR and Open data policy



# How do Open, FAIR data and RDM intersect?





## Open Science across the programme

### Open Science

Better dissemination and exploitation of R&I results and support to active engagement of society

**Mandatory Open Access to publications:** beneficiaries shall ensure that they or the authors retain sufficient intellectual property rights to comply with open access requirements

**Open Access to research data ensured:** in line with the principle "as open as possible, as closed as necessary"; Mandatory Data Management Plan for FAIR (Findable, Accessible, Interoperable, Re-usable) and Open Research Data

- Support to researcher skills and reward systems for open science
- Use of European Open Science Cloud

Funders have expectations about FAIR data and data sharing



**ES** ● Spanish  
Reproducibility Network

**FAIR data  
supports  
Research  
Integrity and  
Reproducibility**

En lo que se refiere al nodo español, tras participar como observadores durante el año 2023, se tomó la decisión de iniciar su andadura en una reunión celebrada en septiembre de dicho año entre varias personas que constituyen el OpenScienceLab de la Universidad Carlos III de Madrid, estando en fase de planeamiento las actividades que se llevarán a cabo.

Se puede contactar al grupo en el siguiente correo electrónico: [info@sprn.es](mailto:info@sprn.es).

<https://sprn.es/>

# European Open Science Cloud (EOSC) vision depends on FAIR data!



**EUROPEAN OPEN  
SCIENCE CLOUD**

- EOSC brings together institutional, national and European stakeholders, initiatives and data infrastructures to develop an inclusive open science ecosystem in Europe.
- Aim is to develop a trusted, virtual, federated environment that cuts across borders and scientific disciplines to store, share, process and re-use research digital objects following FAIR principles.



<https://www.fairsfair.eu/>



<https://fair-impact.eu/>

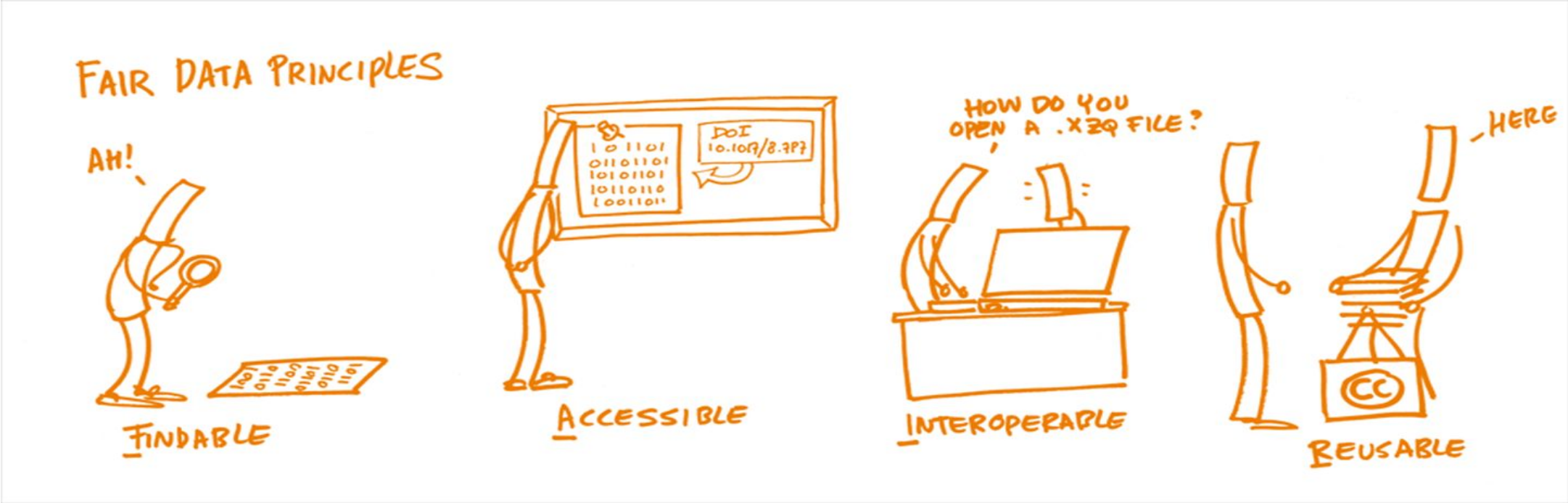


<https://www.skills4eosc.eu/>

# FAIR data doesn't just happen. You need to plan for it!



# Some practical messages about making data FAIR



# Make use of a data repository

Where to deposit FAIR data?

The Carlos III University of Madrid puts at your disposal e-cienciaDatos, the data repository of the Madroño Library Consortium (of which UC3M is a member) for the deposit and dissemination of FAIR data.

We recommend that you consult the support documentation and FAQ made by the Madroño Consortium.

Infographic prepared by the Madroño Consortium: Disseminate your research data in e-scienceData 



[https://www.uc3m.es/ss/Satellite/Biblioteca/en/TextoMixta/1371325489441/Deposito\\_en\\_repositorios\\_de\\_datos\\_de\\_investigacion](https://www.uc3m.es/ss/Satellite/Biblioteca/en/TextoMixta/1371325489441/Deposito_en_repositorios_de_datos_de_investigacion)

Preferred repositories:

1. Domain specific
2. Institutional
3. Generalist (Zenodo, [esciencesDatos](#))

**Consider trustworthy and certified repositories**

# Use registries to find repositories

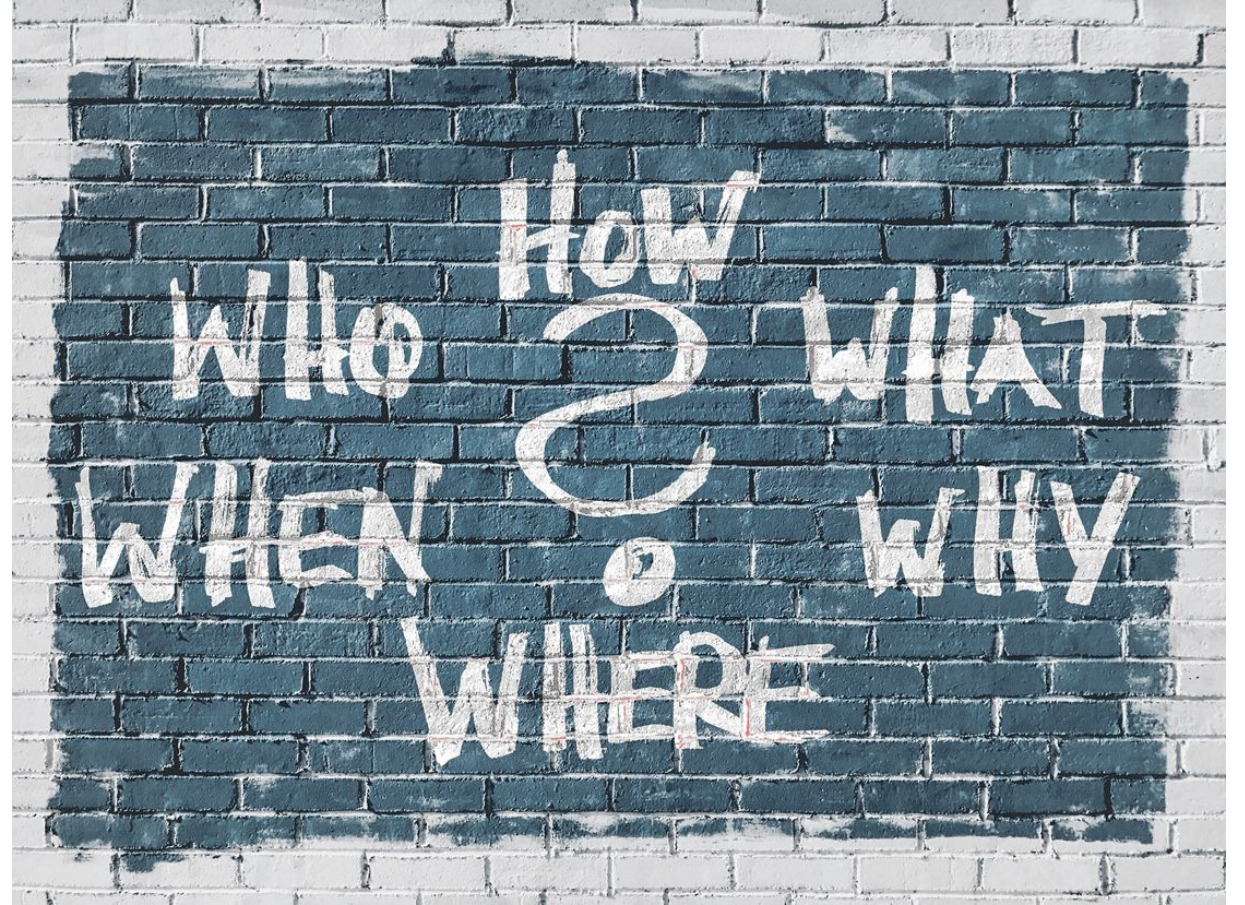
- Filtered search on domain, sub domain, and content type
- Also by licenses, PIDs and certification

The screenshot shows the re3data.org website interface. On the left, a 'Filter' sidebar lists various categories: Subjects, Content Types, and Countries. Under 'Countries', 'Spain (55)' is highlighted with a red circle. The main search area includes a search bar, a 'Search' button, and a pagination control showing '1' of 3 pages. Below the search bar, it says 'Found 55 result(s)'. Two search results are visible: 'UPF Digital Repository' and 'World Values Survey'. The 'UPF Digital Repository' result shows details for Subject(s), Content type(s), and Country. A large, semi-transparent overlay is positioned over the search results, featuring a grid of icons for different registries and standards. The columns are labeled 'Information', 'Access', 'Licenses', 'Persistent Identifier', 'Certificates and Standards', and 'Policy'. The icons include 'i', 'a', '©', 'doi', a red seal, '§', 'urn', 'ark', and 'hdl'. At the bottom of the overlay, there are two rows of small icons representing various licenses and standards.

<https://www.re3data.org/>

# Provide rich metadata descriptions

- Standardised
- Structured
- Machine and human readable



# How much metadata will you provide?

- DataCite metadata
- Citation/disambiguation
  - Identifier e.g. DOI
  - Creator
  - Title
  - Publisher
  - Publication Year
- Licencing/access conditions



Why should you go beyond the minimum required?

# Use discipline specific metadata standards

Search by Discipline



Biology



Earth Science



General Research Data



Physical Science



Social Science & Humanities

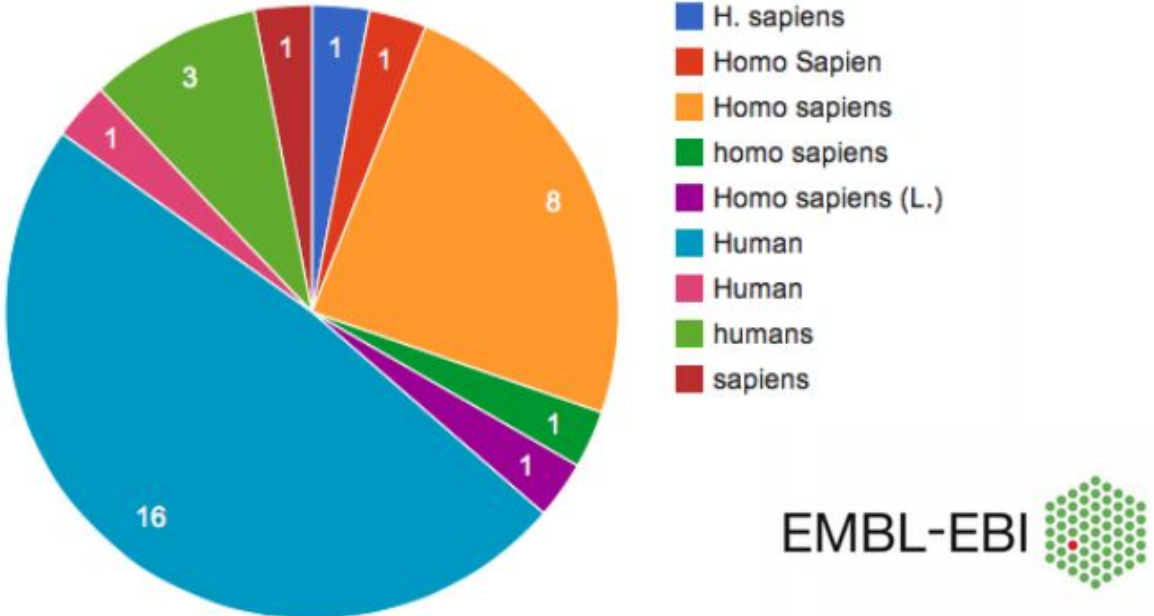


# Support interoperability

# Make use of controlled vocabularies and ontologies

## Controlled vocabularies

“MTBLS1: A metabolomic study of urinary changes in type 2 diabetes in.....”



Example courtesy of Ken Haug, European Bioinformatics Institute (EMBL-EBI)

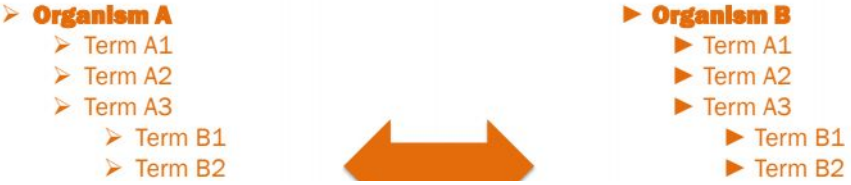
## ...and ontologies?

e.g. SNOMED CT (clinical terms) or MeSH

Include ontologies as well

Defined terms + taxonomy

Useful for selecting keywords to tag datasets

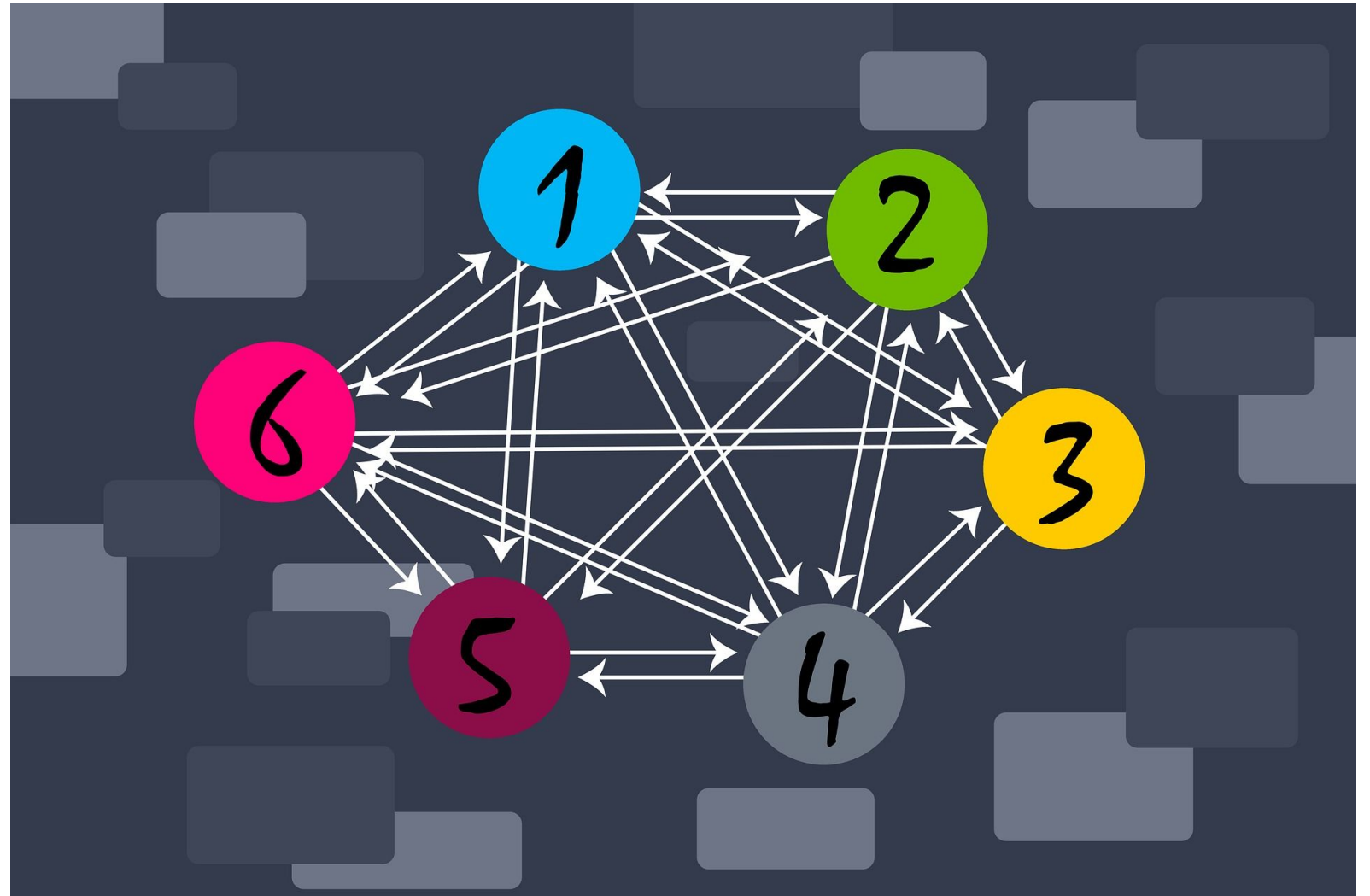


Slides from 'An Introduction to Research Data Management, FAIR and Open Data', S. Venkataraman. [https://drive.google.com/drive/folders/1\\_MXFhrzKVuKjoytVf7wh5Pndp-BAWAA1](https://drive.google.com/drive/folders/1_MXFhrzKVuKjoytVf7wh5Pndp-BAWAA1)

<https://www.go-fair.org/fair-principles/i1-metadata-use-formal-accessible-shared-br-oadly-applicable-language-knowledge-representation/>

# Use persistent identifiers

PIDs provide links  
between  
related outputs  
and actors



# Describe and link to other outputs (e.g, physical samples, software, models)



## D4.4 - Guidelines for recommended metadata standard for research software within EOSC

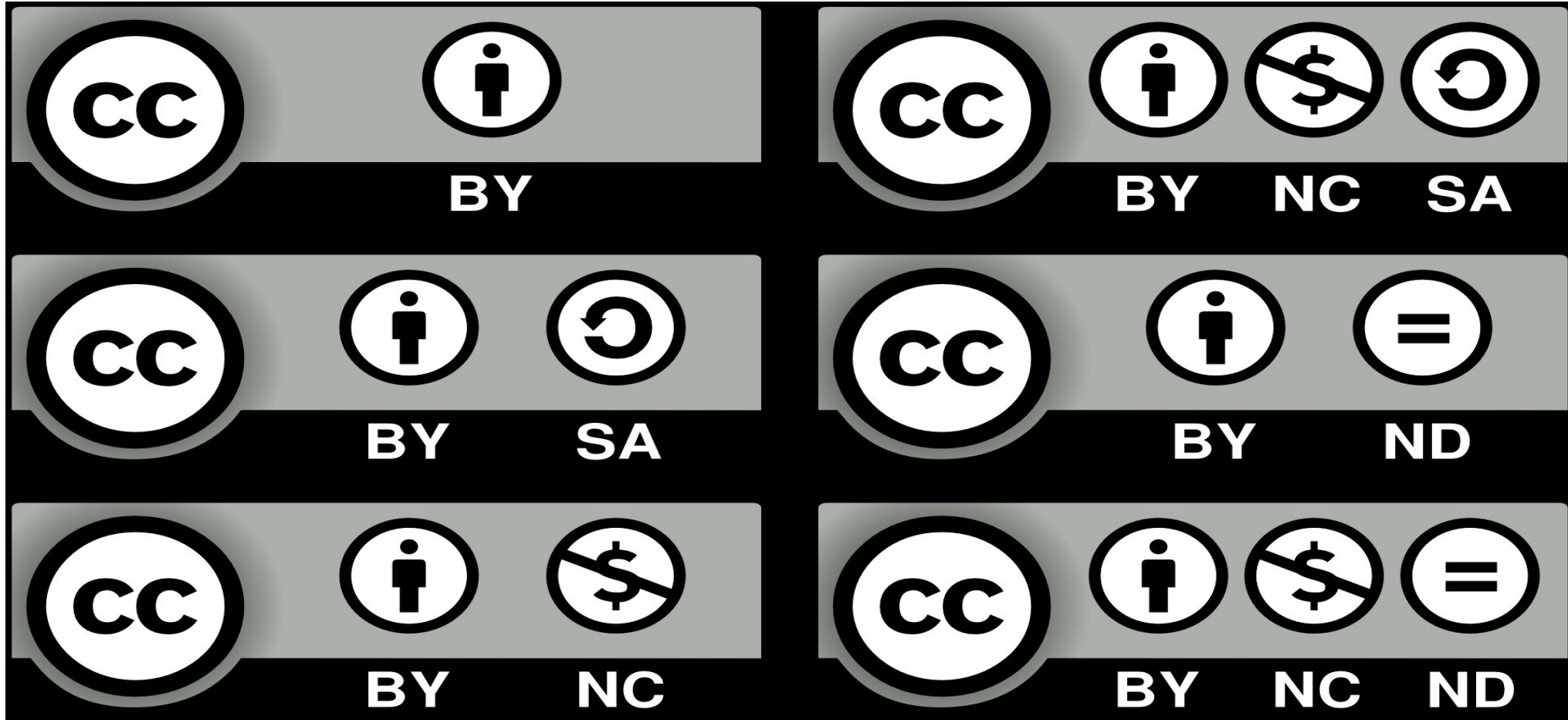
📅 Published on: 06 March 2024

👤 Gruenpeter, Morane; Granger, Sabrina; Monteil, Alain; Chue Hong, Neil; Breitmoser, Elena; Antonioletti, Mario; Garijo, Daniel; González Guardia, Esteban; Gonzalez Beltran, Alejandra; Goble, Carole; Soiland-Reyes, Stian; Juty, Nick; Mejias, Gabriela

The FAIR-IMPACT Guidelines for recommended metadata standards for research software within the EOSC present the first proposal of the Research Software MetaData (RSMD) Guidelines, developed by Task 4.3 (T4.3), Standard metadata for research software, as part of Work Package 4, Metadata and ontolo... [Read more](#)

👁 1054 views    ⬇ 908 downloads

# License your data



# Restricted data can still be FAIR data

- Authentication processes can be used
- Safe havens or institutional data vaults are an option
- If data can't be shared at all, metadata could still be FAIR



# Not all outputs need to be made and kept FAIR

Five steps to follow

1. **Could** this data be re-used
2. **Must** it be kept as evidence or for legal reasons
3. **Should** it be kept for its potential value
4. **Consider costs** – do benefits outweigh cost?
5. **Evaluate criteria** to decide what to keep

5 steps to decide what data to keep

[www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep](http://www.dcc.ac.uk/resources/how-guides/five-steps-decide-what-data-keep)



To be FAIR, metadata should be kept accessible even if the data no longer exist

# Consult your local support and guidance!

**RESEARCH DATA**

Libra

CARRYING OUT RESEARCH FINDING INFORMATION SERVICES LIBRARY

Engl

## RESEARCH DATA

Library / Scientific Communication / Research data

English Powered by Google Translate

- > Di
- > Pu
- > Of
- > Re
- > Me

- > Digital identity
- > Publish/spread
- > Open publication
- > **Research data**
- > Meet our researchers

The UC3M supports all open science practices to create an Open Science University, which also includes, in addition to publications, the deposit of other research results such as data. The Library Service provides support for the planning, management and deposit of research data in the multidisciplinary repository e-CienciaDatos maintained by the Madroño Consortium , which in turn manages InvestigaM, Open Science Portal .

- > Data management plan
- > Open access requirements for research data
- > Deposit in repositories
- > Keys to research data management

library finder

Buscar en todo

Ir al buscador completo

> R&B, Research and Library

If you want to make a query:

e-archivo@listserv.uc3m.es  
91 624 9076/9082

( f ) ( X ) ( in ) ( ↻ )

<https://www.uc3m.es/library/how-publish/research-data>

# Benefits!

- Helps prevent data loss
- Leads to more citations and impact
- Easier outputs reporting
- Supports research integrity
- Enables reproducibility
- Easier to collaborate



Artist's impression of COVID-19 open access data sharing. Credit: Spencer Phillips

## Open data sharing accelerates COVID-19 research

19 Oct 2020 - 15:51

### Summary

- Open access increases the visibility of research data and information, giving scientists the ability to build upon and react to existing research quickly
- EMBL-EBI launched the European COVID-19 Data Platform to enable rapid access to datasets and results pertaining to the SARS-CoV-2 outbreak
- Open access data sharing has greatly accelerated COVID-19 research and helps further our understanding of the biology, transmission, and spread of the SARS-CoV-2 virus

Thanks!

joy.davidson@glasgow.ac.uk  
@jd162a

eosc

FAIR-IMPACT

Expanding FAIR solutions across EOSC



@fairimpact\_eu /company/fair-impact-eu-project



Funded by  
the European Union