



**FAIRsFAIR**  
Fostering Fair Data Practices in Europe

# Planning for FAIR data

Joy Davidson, DCC

With contributions from DCC and FAIRsFAIR colleagues Sarah Jones and Marjan Grootveld

30<sup>th</sup> June 2020



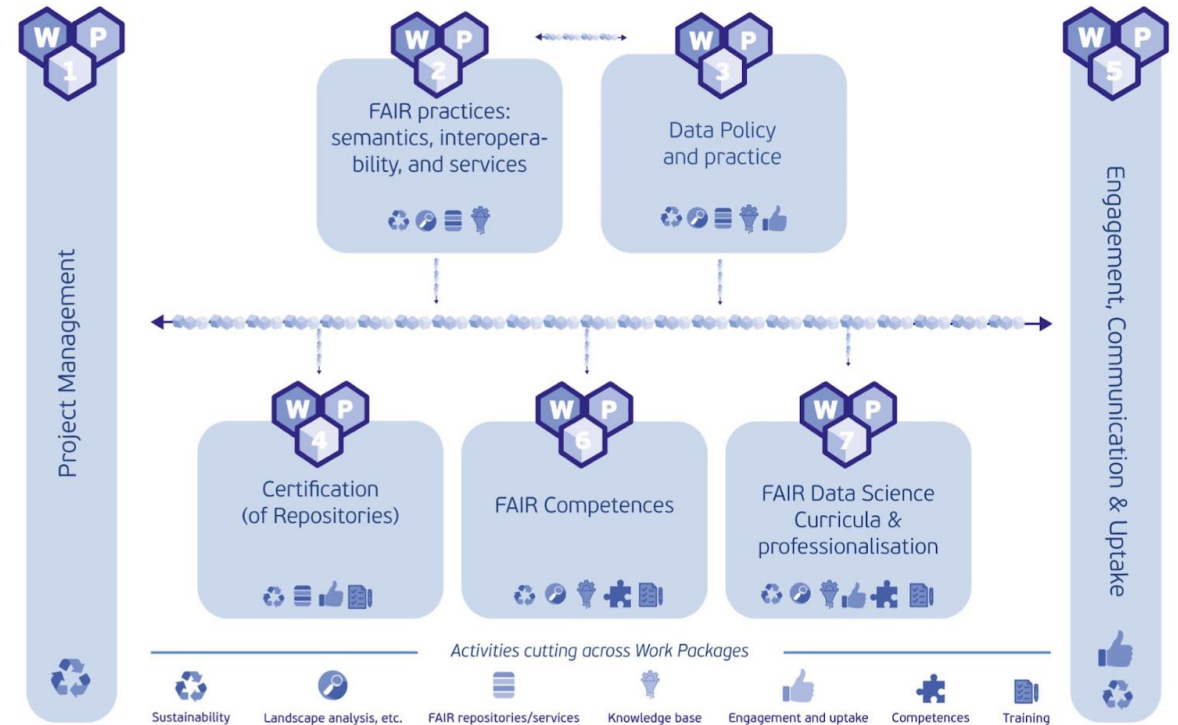
## What is Digital Curation Centre?

*a centre of expertise in digital information curation with a focus on building capacity, capability and skills for research data management and open science*



Training | Events | Tools | Advocacy | Consultancy | Guidance | Publications | Projects

## FAIRsFAIR



<https://www.fairsfair.eu>

This talk is aimed at both researchers and support staff and aims to:

- Introduce key terms in a practical sense
- Show that data management planning can help them to make data FAIR
- Aimed at both researchers and support staff





## Open, FAIR and RDM – setting FAIR in context

Slide from 'What it means to be FAIR', Sarah Jones <https://www.slideshare.net/sjDCC/what-it-means-to-be-fair?>

Image Richard Balog <https://unsplash.com/photos/P6FgiDNe6W4>



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/>



# FAIR principles

## 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

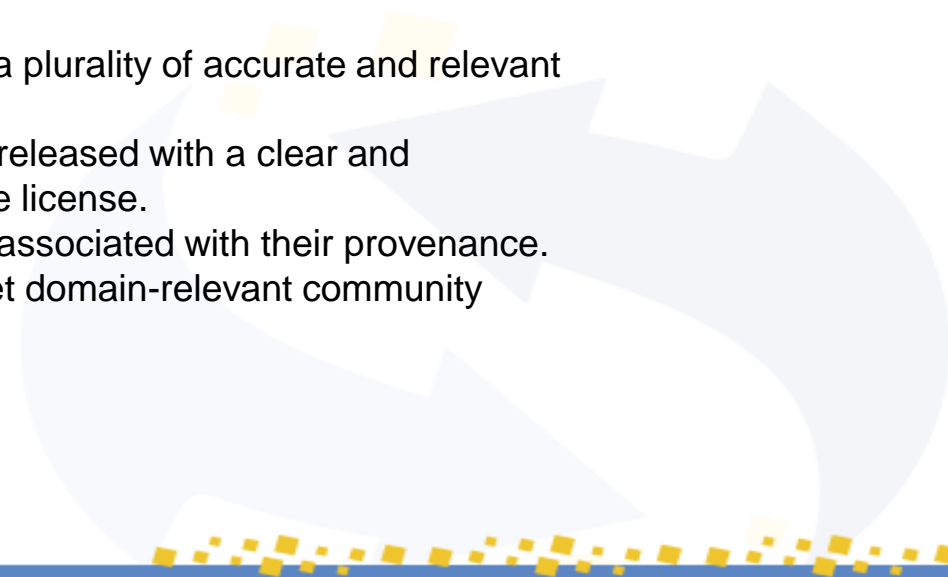
- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

## 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.



# FAIR is nothing new

- Various research communities have been sharing their data in a 'FAIR' way long before the term emerged
- Meaningful and memorable articulation of concepts
- Natural desire to want to be 'fair'
- FAIR is gaining significant international traction

Slide from 'What it means to be FAIR', Sarah Jones <https://www.slideshare.net/sjDCC/what-it-means-to-be-fair?>



## 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 sharing and FAIR data.



## FAIR doesn't just apply to big data



Often longer tail is less well catered for so more institutional may be support needed

[This Photo](#) by Unknown Author is licensed under [CC BY-NC](#)



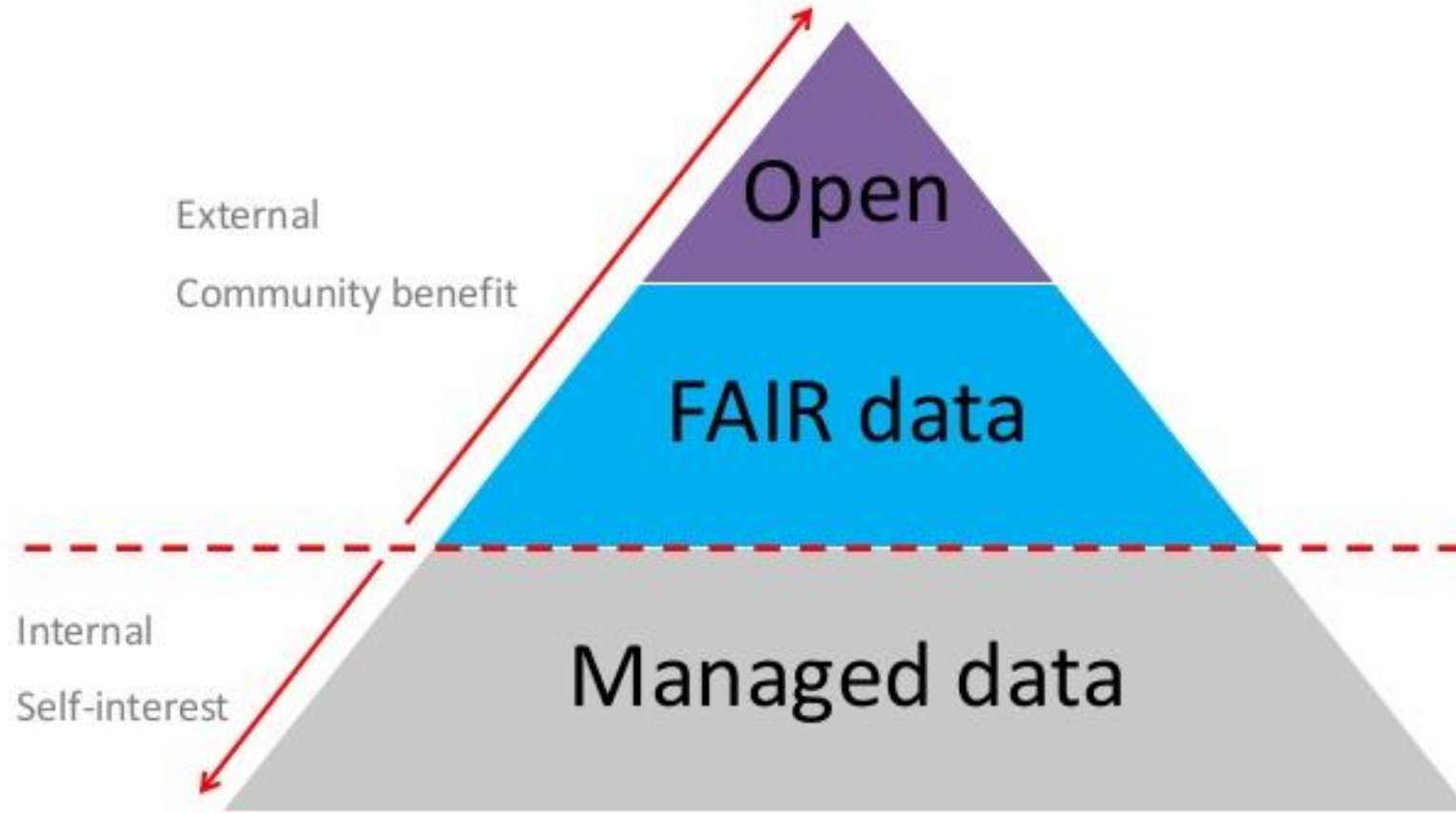
# FAIR doesn't just happen - data management planning helps!

- What data will be created (format, types, volume...)
- Standards and methodologies to be used (incl. metadata)
- How ethics and Intellectual Property will be addressed
- Plans for data sharing and access
- Strategy for long-term preservation

***A DMP is a plan to share!***



## How do Open, FAIR & RDM intersect?

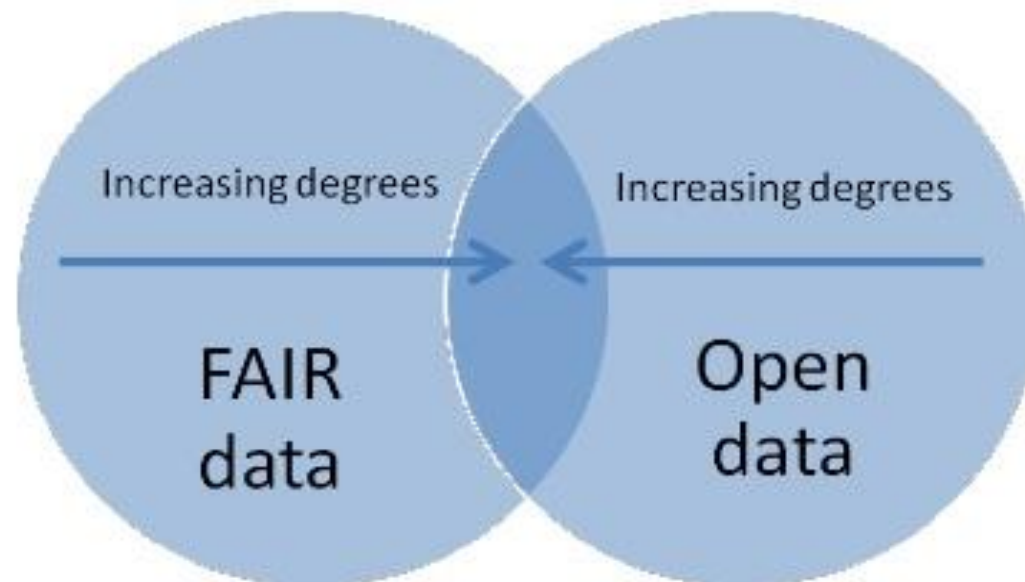


Slide from 'What it means to be FAIR', Sarah Jones <https://www.slideshare.net/sjDCC/what-it-means-to-be-fair?>

# FAIR and Open

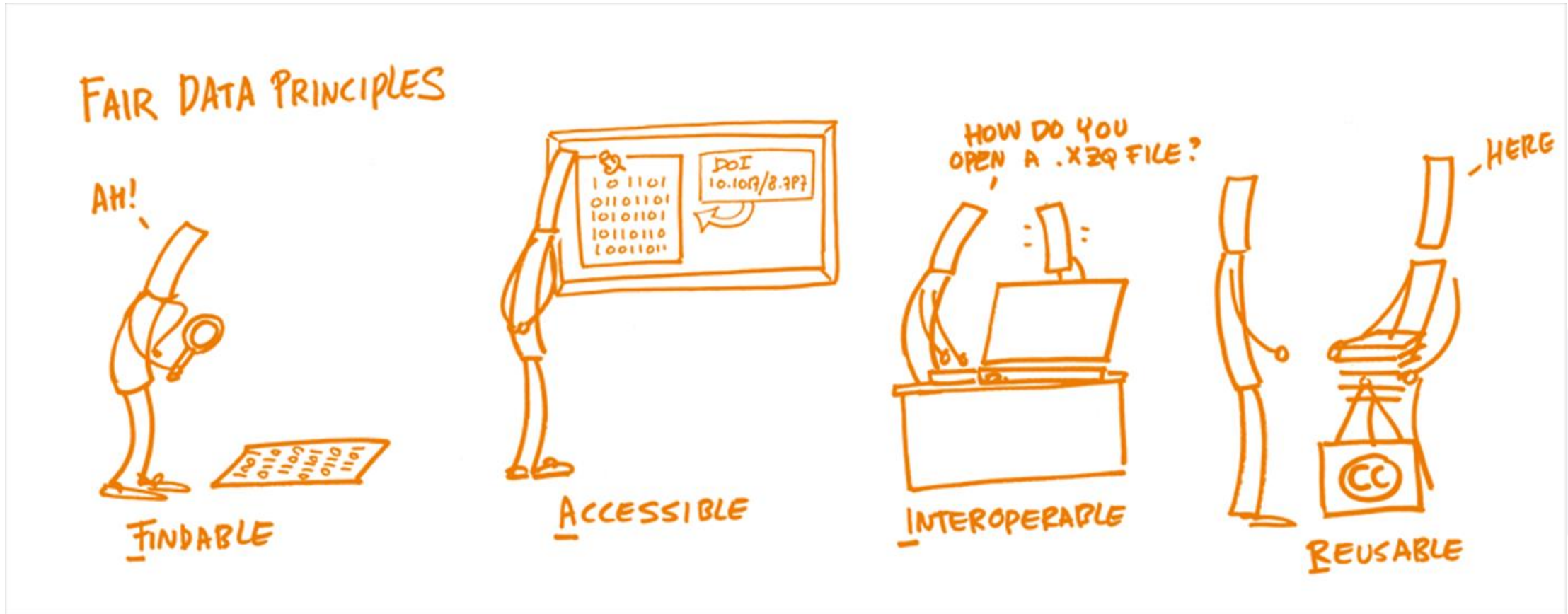
Concepts of FAIR and Open should not be conflated.  
Data can be FAIR or Open, both or neither

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



Slide from 'What it means to be FAIR', Sarah Jones <https://www.slideshare.net/sjDCC/what-it-means-to-be-fair?>

# Some practical messages about making data FAIR





# Make use of a data repository



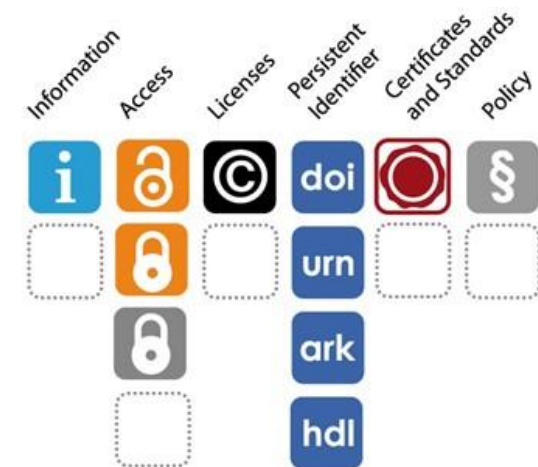
Preferred repositories:

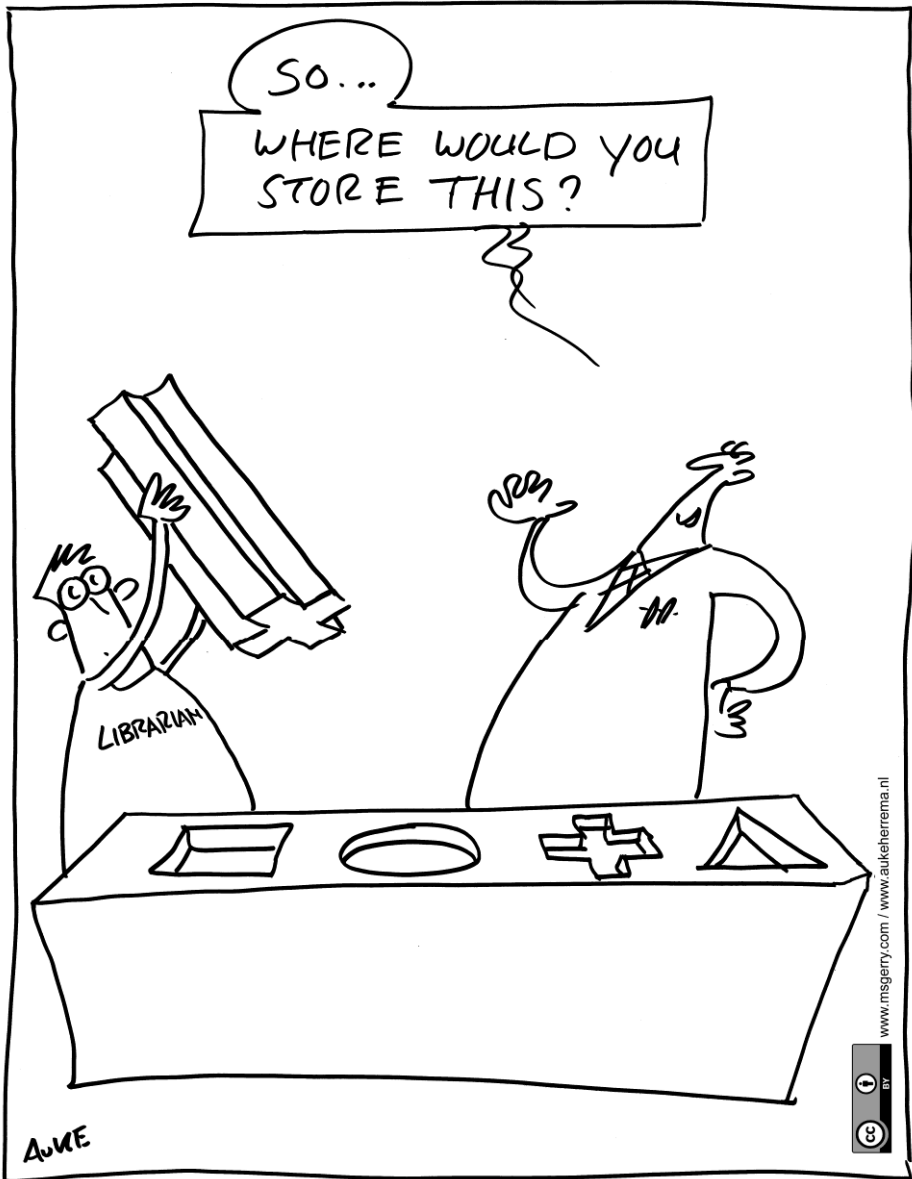
1. Domain specific
2. Institutional (Apollo)
3. Generalist (Zenodo, figshare)

**Try to choose a FAIR aligned repository**

## re3data.org

- Global registry of research data repositories
- Funded by the German Research Foundation (DFG)
- Filtered search and browse options





LEARNING HOW TO ARCHIVE DATA

**FAIRsharing.org**  
Standards, databases, policies

Search all of FAIRsharing | Standards | Databases | Policies | Collections | Add/Claim Content | Stats | Log in or Register

A curated, informative and educational resource on data and metadata standards, inter-related to databases and data policies.

HOW CAN WE HELP?

We guide consumers to discover, select and use these resources with confidence, and producers to make their resource more discoverable, more widely adopted and cited.

**Researchers in academia, industry and government**  
Identify and cite the standards, databases or repositories that exist for your discipline when creating a data management plan, releasing data or submitting a manuscript to a journal...  
[\[read more\]](#)

Researchers | Developers & Curators | Journal Publishers | Librarians &...

<https://fairsharing.org>

**Data Deposit Recommendation Service**  
for humanities researchers

Find a suitable digital repository to deposit your research data or to include in your data management plan by answering the questions below

In which country are you based as a researcher?  
Select one

What is your disciplinary field?  
Select one

Clear selection

**DARIAH-EU** Digital Research Infrastructure for the Arts and Humanities  
**HaS** Humanities at Scale  
**re3data.org** REGISTRY OF RESEARCH DATA REPOSITORIES

financing from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 675570

<https://ddrs-dev.dariah.eu/ddrs/>

About Search

**Repository Finder**  
Find a repository to upload your data.

Repository Finder, a pilot project of the **Enabling FAIR Data Project** led by the American Geophysical Union (AGU) in partnership with DataCite and the Earth, space and environment sciences community, can help you find an appropriate repository to deposit your research data. The tool is hosted by DataCite and queries the re3data registry of research data repositories.

As part of the **FAIR4FAIR** project, which aims to supply practical solutions for the use of the FAIR data principles throughout the research data life cycle, the Repository Finder is extended to query for repositories relevant to FAIR4FAIR Project.

Search re3data for a repository to upload your data

Type to search...

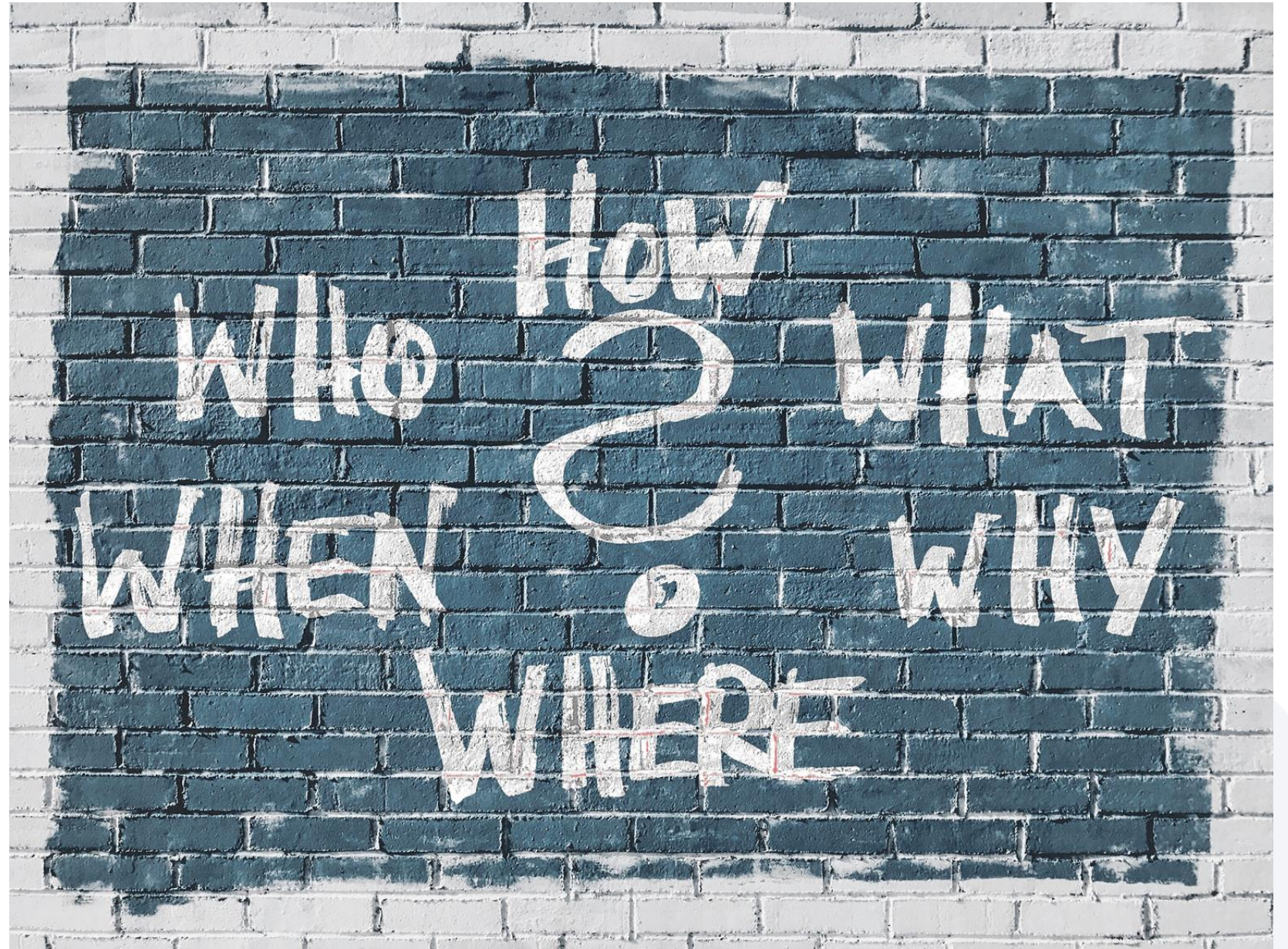
[See the repositories](#) in re3data that meet the criteria of the Enabling FAIR Data Project.

[See the repositories](#) in re3data that meet the criteria of the FAIR4FAIR Project.

<https://repositoryfinder.datacite.org/>



**Provide context!**





# What documentation is needed?

## Documentation

Think about what is needed in order to evaluate, understand, and reuse the data.

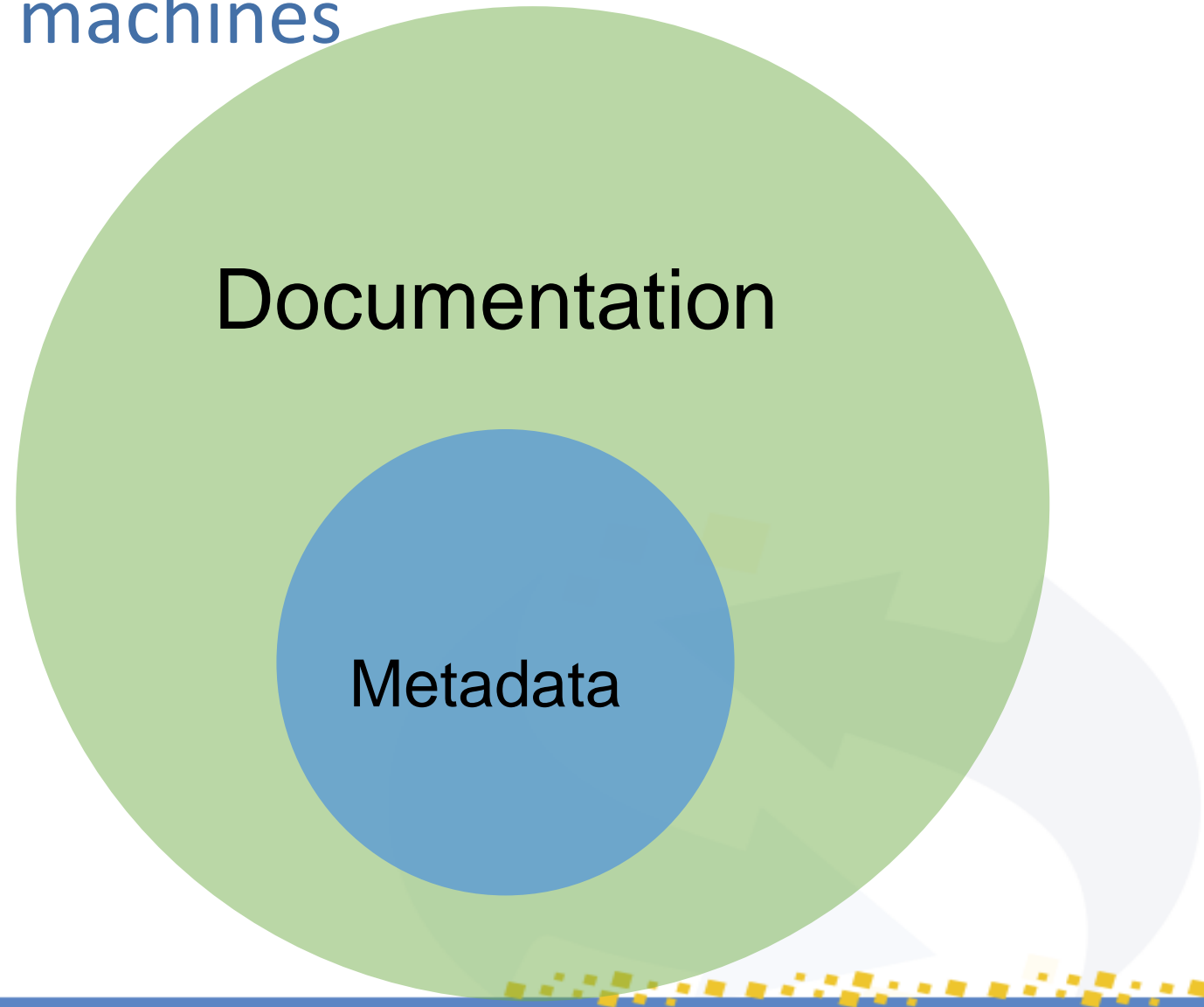
- Why was the data created?
- Have you documented what you did and how?
- Did you develop code to run analyses? If so, this should be kept and shared too.
- Important to provide wider context for trust

Slide 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)



# Descriptions for humans and machines

- Metadata
  - Standardised
  - Structured
  - Machine and human readable



# What is the minimum metadata required?

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



How much more  
will you need to  
provide?

# Use domain specific standards for descriptions

## Search by Discipline



Biology



Earth Science



General Research Data



Physical Science



Social Science & Humanities

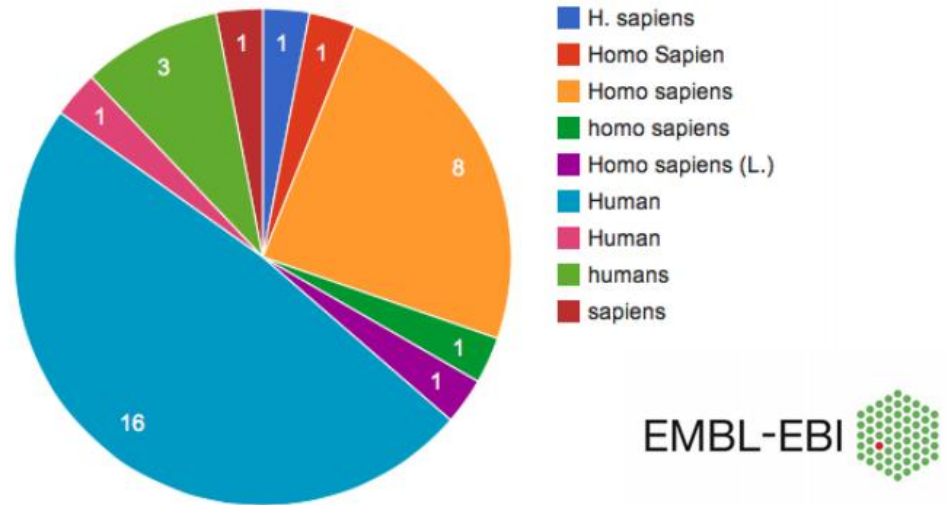


# Metadata and interoperability

Make use of controlled vocabularies, ontologies, and thesauri

## 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

### ► Organism A

- Term A1
- Term A2
- Term A3
  - Term B1
  - Term B2
- Term C4
- .
- .
- .
- Term n

### ► Organism B

- Term A1
- Term A2
- Term A3
  - Term B1
  - Term B2
- Term C4

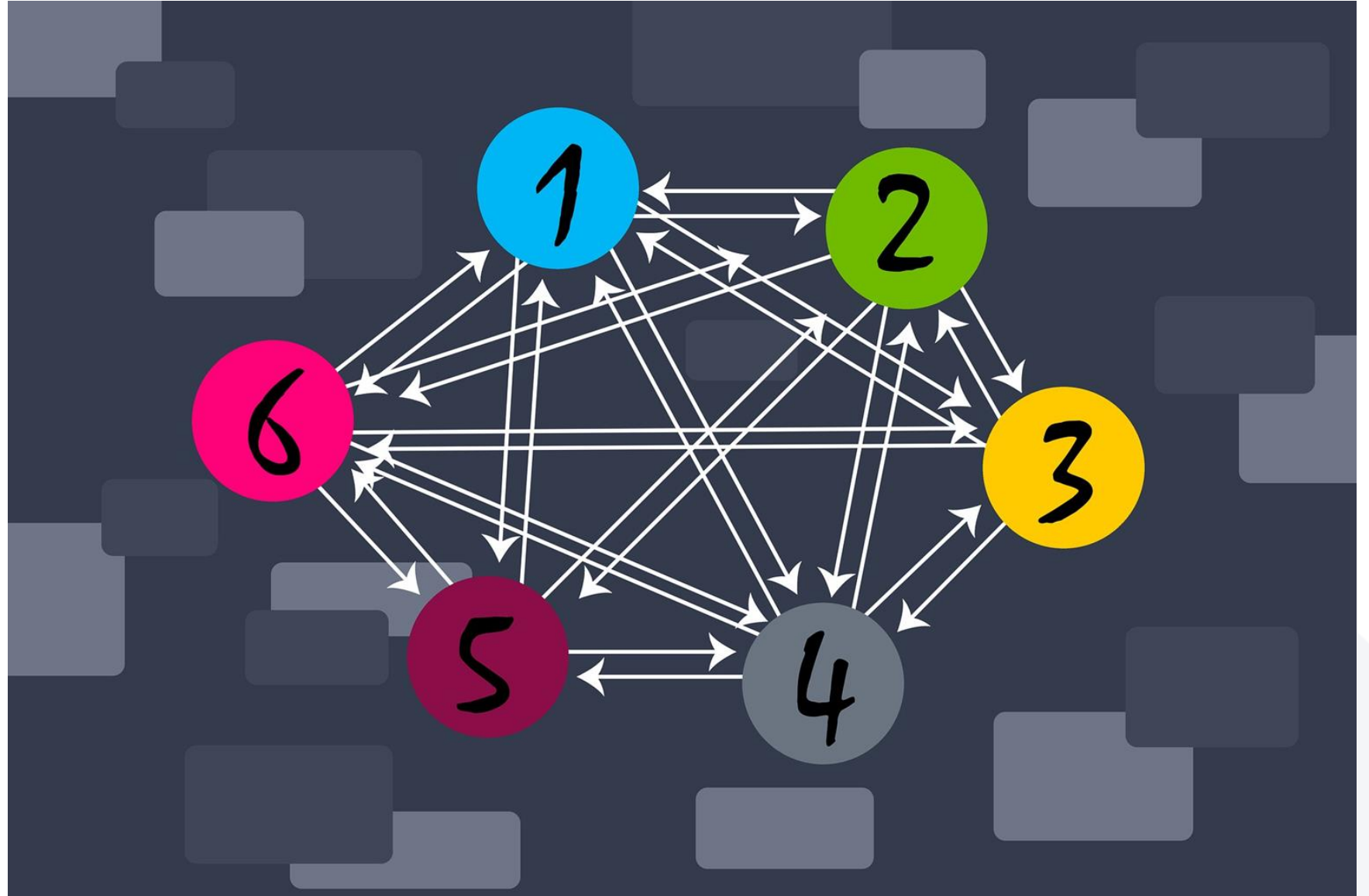


Slides from 'An Introduction to Research Data Management, FAIR and Open Data', S. Venkataraman.

[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-broadly-applicable-language-knowledge-representation/>

Use identifiers and  
metadata to  
link to  
related outputs

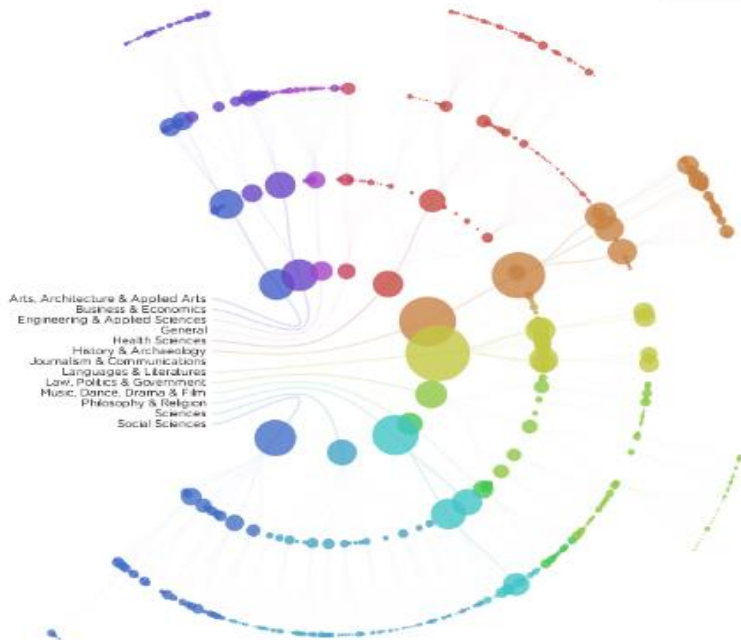




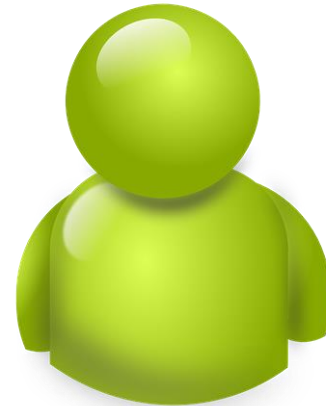
# Not only data and publications - software, models, physical samples and people!

Books: Aggregate holdings, per subject

Total holdings    Total circulations    Relative activity



[http://spatialinformationdesignlab.org/project\\_sites/library/catalog.html](http://spatialinformationdesignlab.org/project_sites/library/catalog.html)



[http://www.ukcrcepm.org.uk/Coventry\\_Warwick\\_CRF/PublishingImages/Tissue%20Bank%201.jpg](http://www.ukcrcepm.org.uk/Coventry_Warwick_CRF/PublishingImages/Tissue%20Bank%201.jpg)



Software Management Plan Service  
Prototype

[Home](#) [About](#) [Help](#)

Welcome.

Software Management Plan Service has been developed by the **The Software Sustainability Institute** to help you write software management plans.

It is powered by **DMPonline** developed by the

Sign in

Email address \*

Password \*

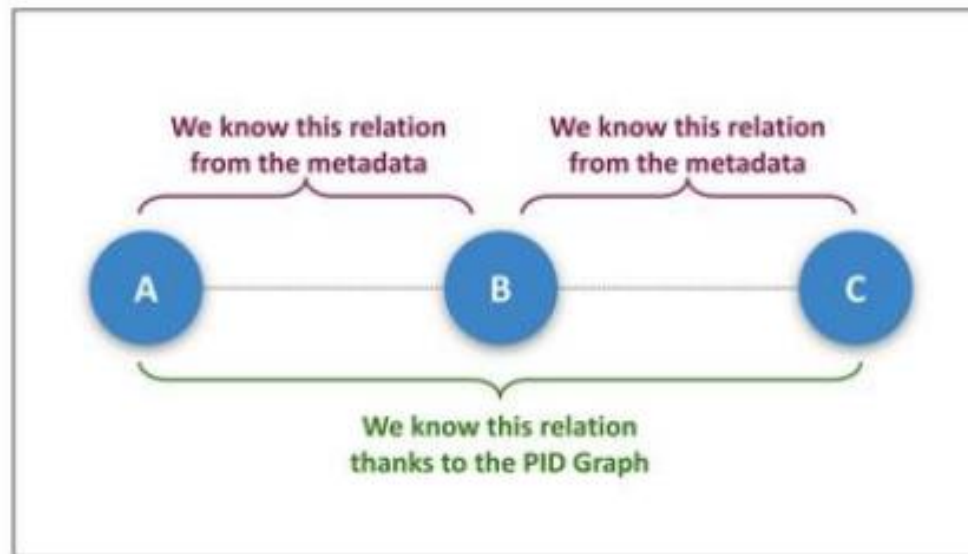
[Forgot your password?](#)

Remember me

<https://ssi-dev.epcc.ed.ac.uk/>

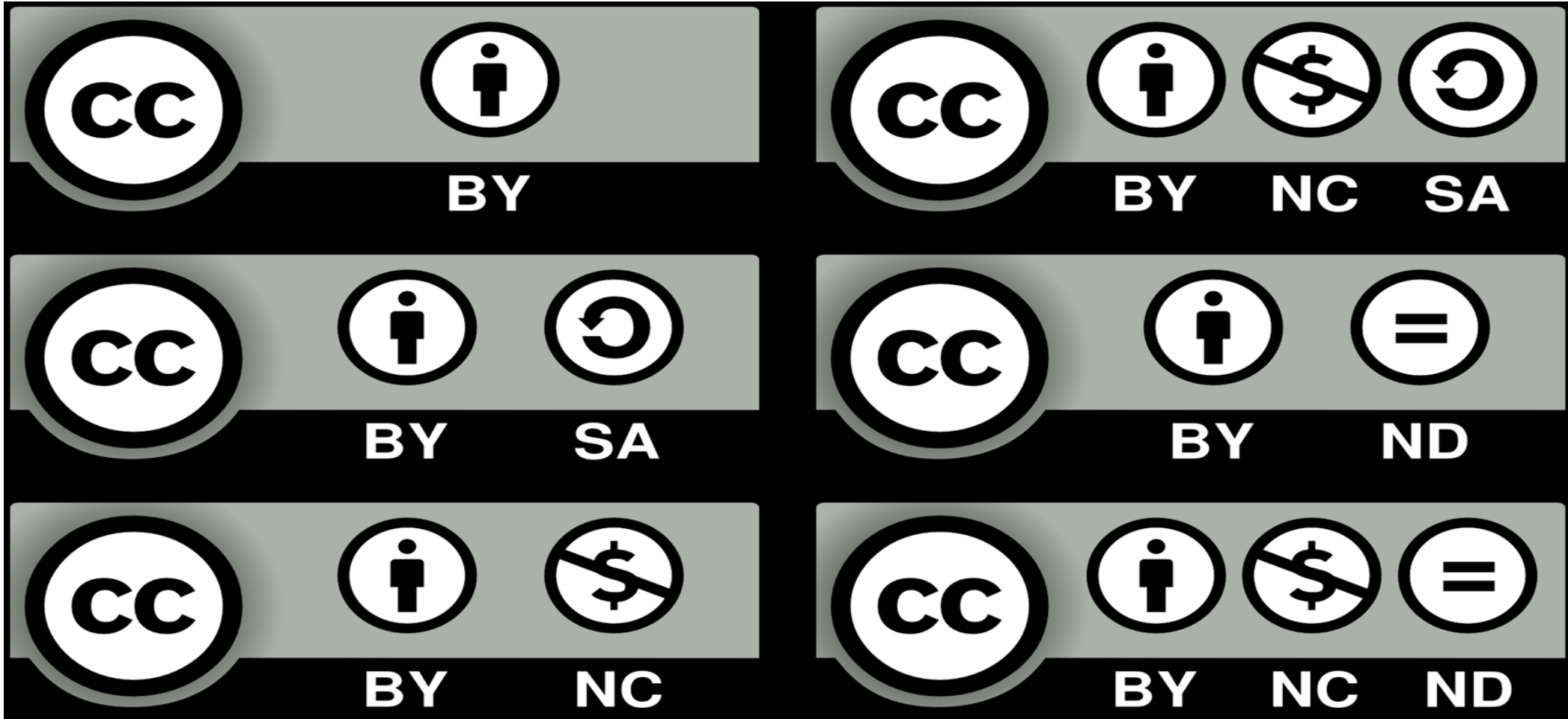
# PID Graphs – the next level

- If you have a collection of PIDs describing different objects, these can be joined together in a graph to form relationships
- These graphs can aid in workflows and provenance



Slide 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)

# Licensing information should be included in metadata



## Remember - FAIR doesn't necessarily mean open



Data should be made as open as possible, but as closed as necessary





## Level of openness should reflect:

- Funding body requirements
- Personal sensitivities
- Commercial sensitivities





## FAIR closed data

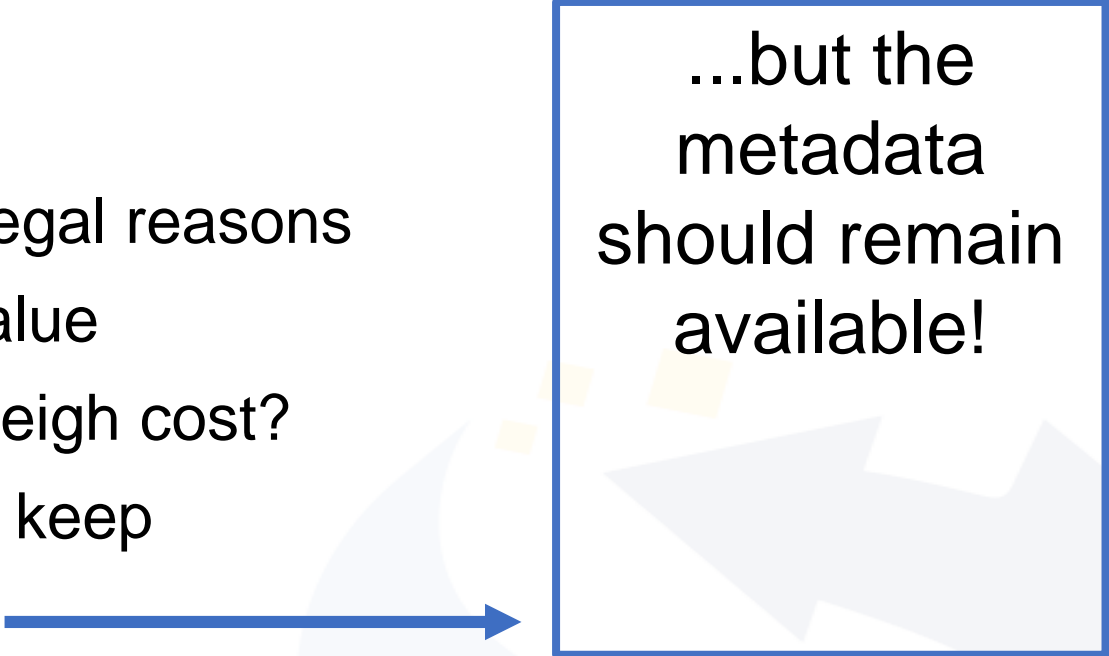
- Authentication process
- Safe havens or institutional data vault
- Metadata should be FAIR



# Not all data selected will be kept in perpetuity...

## 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



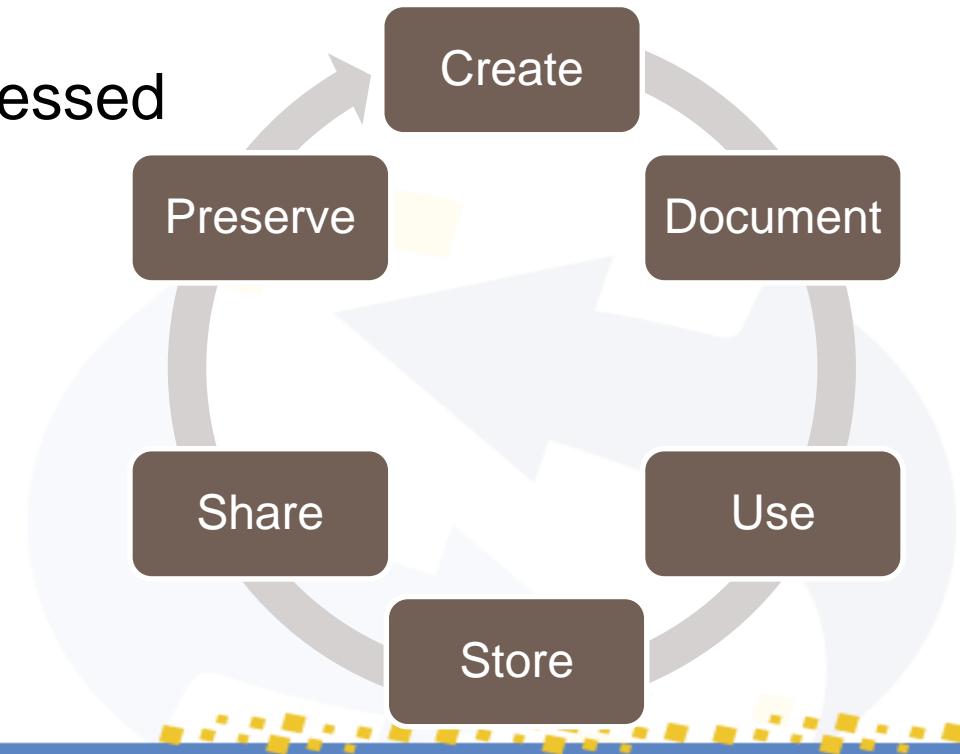
...but the metadata should remain available!

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)

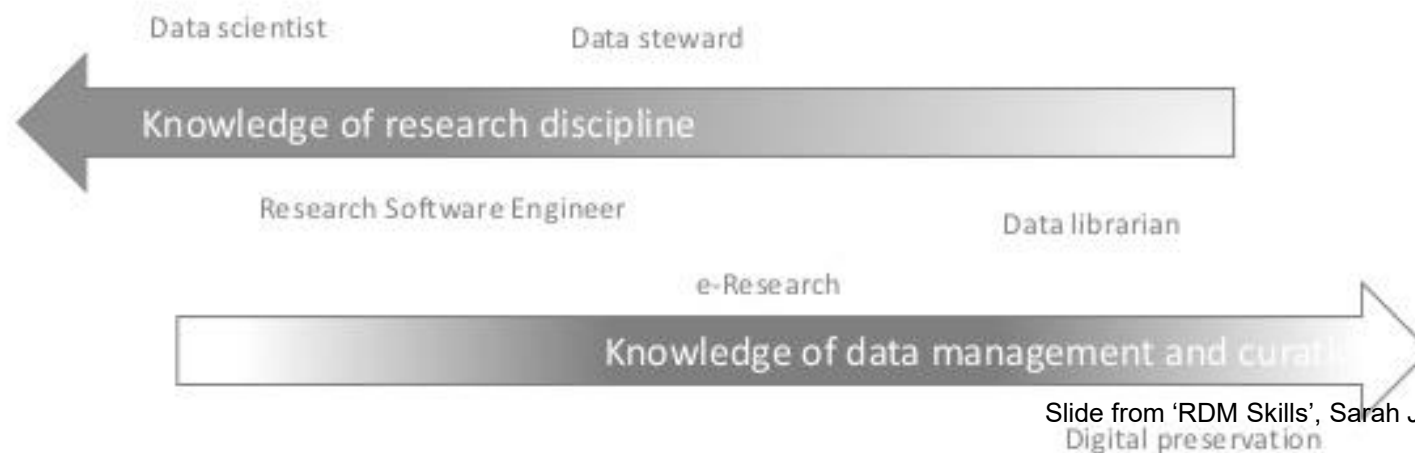
# Metadata – not just at the point of deposit!

- What data will be created (format, types, volume...)
- Standards and methodologies to be used (incl. metadata)
- How ethics and Intellectual Property will be addressed
- Plans for data sharing and access
- Strategy for long-term preservation



# Researchers can't do it all – they will need help!

## Data skills landscape



Slide from 'RDM Skills', Sarah Jones <https://www.slideshare.net/sjDCC/rdm-skills>



## 2. Recognise the value of all roles

*“FAIR data should be recognised as a core research output and included in the assessment of research contributions and career progression. The provision of infrastructure and services that enable FAIR data must also be recognised and rewarded accordingly.”*

Hodson, S, Jones, S et al. (2018) *Turning FAIR into Reality: Final report and action plan from the European Commission Expert Group on FAIR data.*  
<https://doi.org/10.2777/1524>

Slide from 'RDM Skills', Sarah Jones <https://www.slideshare.net/sjDCC/rdm-skills>



## Assessing the FAIRness of Data

<https://www.fosteropenscience.eu/learning/assessing-the-fairness-of-data/>



<https://fairsfair.eu/news/fairsfair-launches-fair-assessment-tool>

For a refresher on the issues raised and to learn about free FAIR assessment tools, take a look at the Assessing the FAIRness of Data course in the FOSTER Open Science toolkit. Please also check out the new FAIRsFAIR FAIR Aware tool which is out for comment now.

Thanks – any questions?

