



# FAIRsFAIR

Fostering Fair Data Practices in Europe

## Planning for FAIR data

Joy Davidson, DCC

With contributions from DCC and FAIRsFAIR colleagues

10 June 2021



FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558



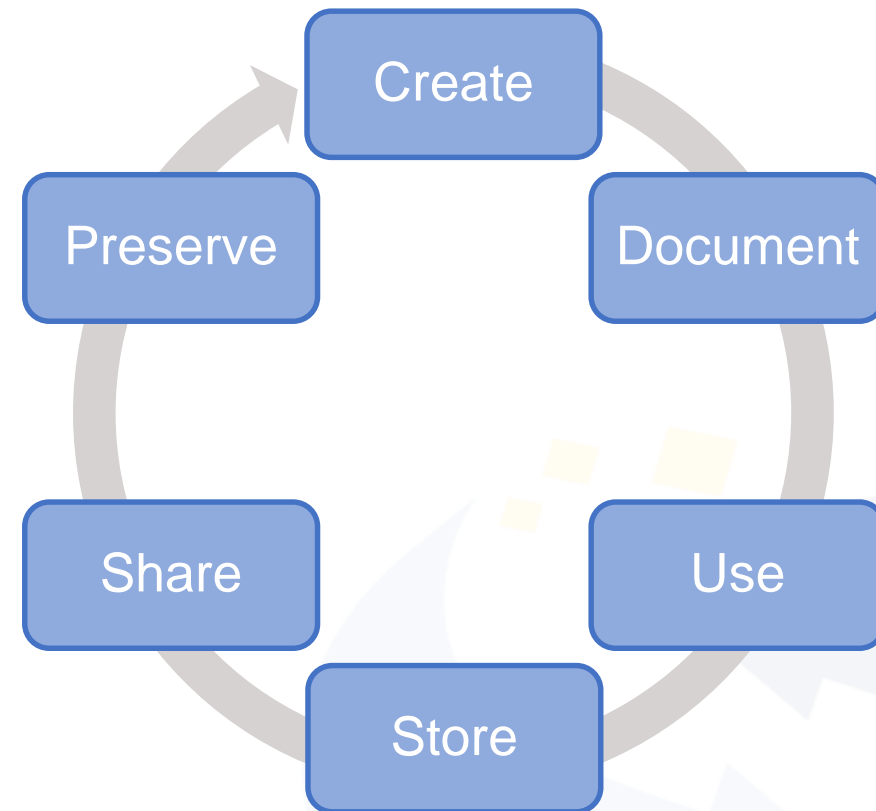
# Overview

After this session, you will:

- Know what a Data Management Plan (DMP) is
- Understand the difference between FAIR and Open Data
- Be aware of what should be included in a DMP
- Know about free tools to help you write DMPs

# What is Research Data Management?

- Data Management Planning
- Creating data
- Documenting data
- Accessing / using data
- Storage and backup
- Selecting what to keep
- Sharing data
- Data licensing and citation
- Preserving data



**Principles apply equally to data and metadata  
(i.e., the description that provides context about the data).**

# 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.

**Check out FAIR-Aware tool to get some tips on how to make your data FAIR.**

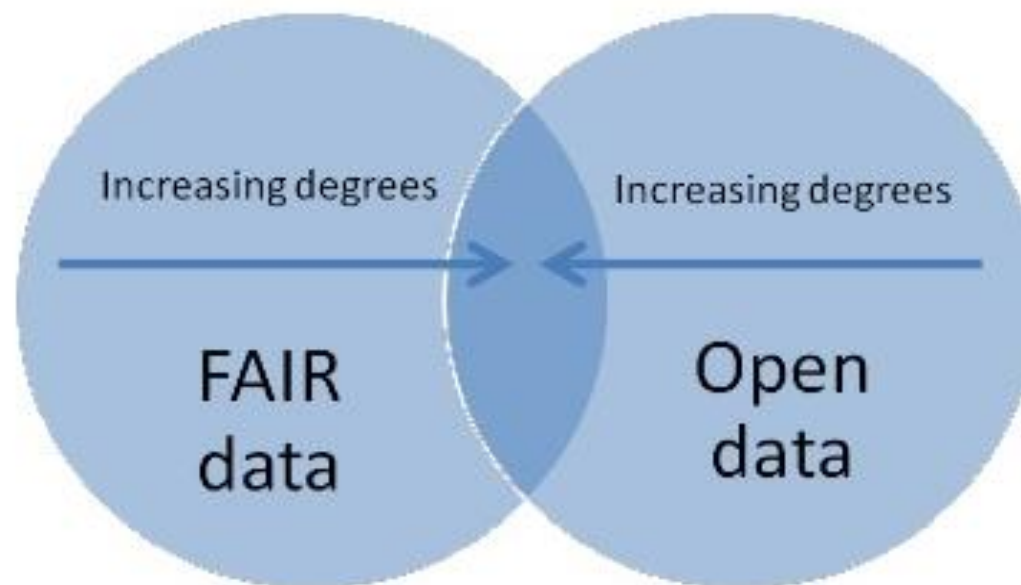
<https://www.fairsfair.eu/fair-aware>



# 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?>



# Research data under Horizon Europe

- The governing principle is to manage research data **responsibly, in line with FAIR**
  - At proposal stage, beneficiaries will be **evaluated** on **preliminary research data and research output management considerations**
  - All projects that **generate (and/or re-use)** research data will have to establish and regularly update a **Data Management Plan** (living document)
  - Beneficiaries will have to **deposit data in a trusted repository** and link data to publications they underpin, if applicable
    - For some actions, additional obligation that the data repository is federated under EOSC.
  - Beneficiaries to ensure **open access ASAP and within the deadline set up in the DMP** under CC BY or CC0 (or equivalent), unless exceptions apply (justified in the DMP), **following the principle “as open as possible, as closed as necessary”**

# What is a data management plan (DMP)?

- 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!***



## Consider: where will you store the data during your research?

- Your own laptop
- University systems
- Cloud storage
- Combination

Your decision will be based on how sensitive your data are, how robust you need the storage to be, who needs access to the data, and when they need access to the data!



# Consider: How will you name your files?

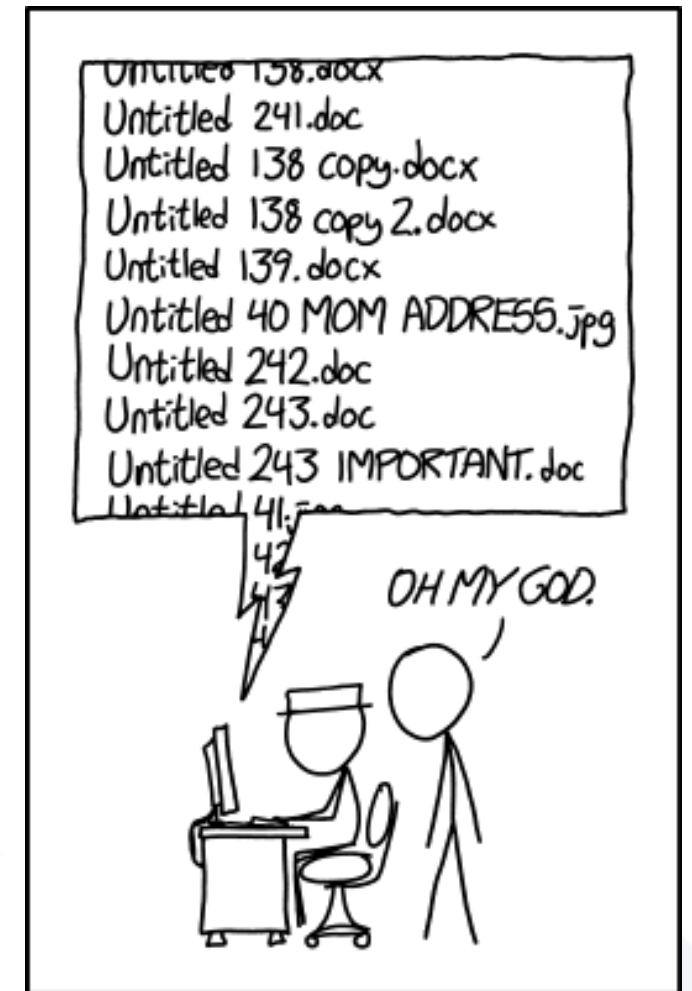
- Keep it simple!
- Include dates and times is necessary
- Avoid special characters
- Use hyphens or underscores not spaces
- Make sure you agree approach with your research partners

Workshop\_report\_200820\_final.doc

Workshop\_report\_200820\_final\_jdedits.doc

Workshop\_report\_200820\_final\_FINAL.doc

Workshop\_report\_200820\_final\_FINAL\_210820. doc



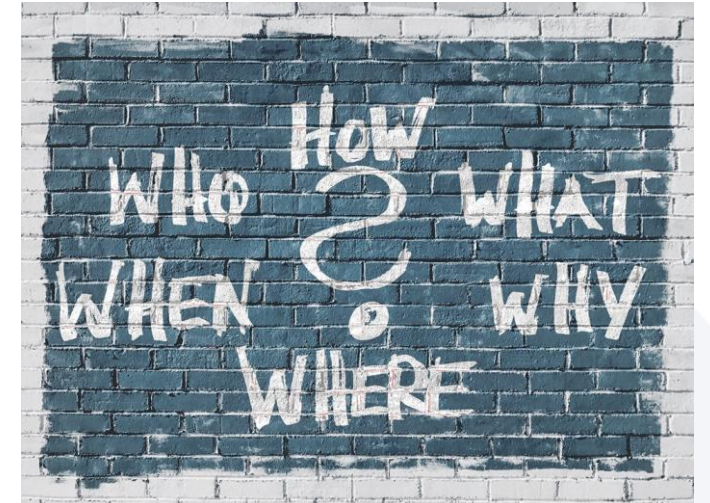
PROTIP: NEVER LOOK IN SOMEONE  
ELSE'S DOCUMENTS FOLDER.

Image source: <http://sxnkd.com/1459/>

# What documentation is needed – for you and for others?

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)

# Consider: Using metadata standards

## Search by Discipline



Biology



Earth Science <https://rdamsc.bath.ac.uk/>



General Research Data



Physical Science



Social Science & Humanities



<https://rdamsc.bath.ac.uk/>



## Consider: using a repository

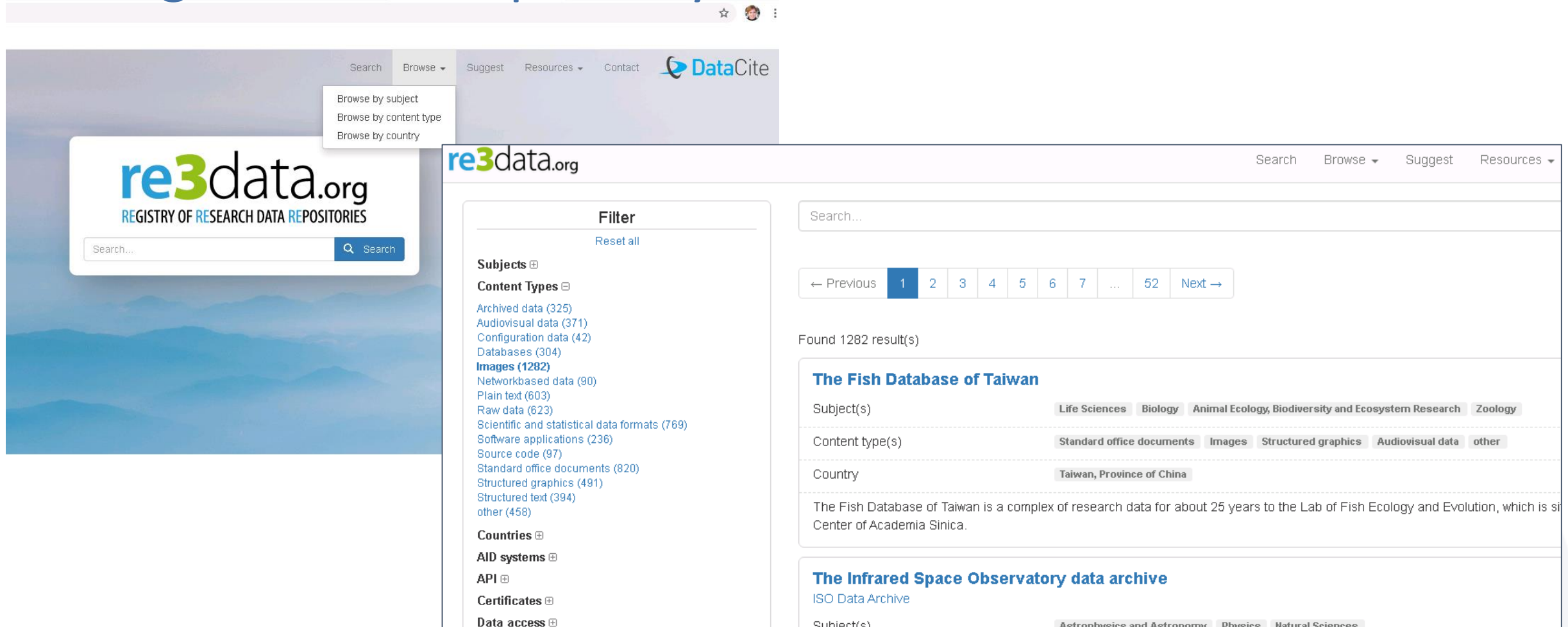


Preferred repositories:

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

**Try to choose a FAIR enabling repository**

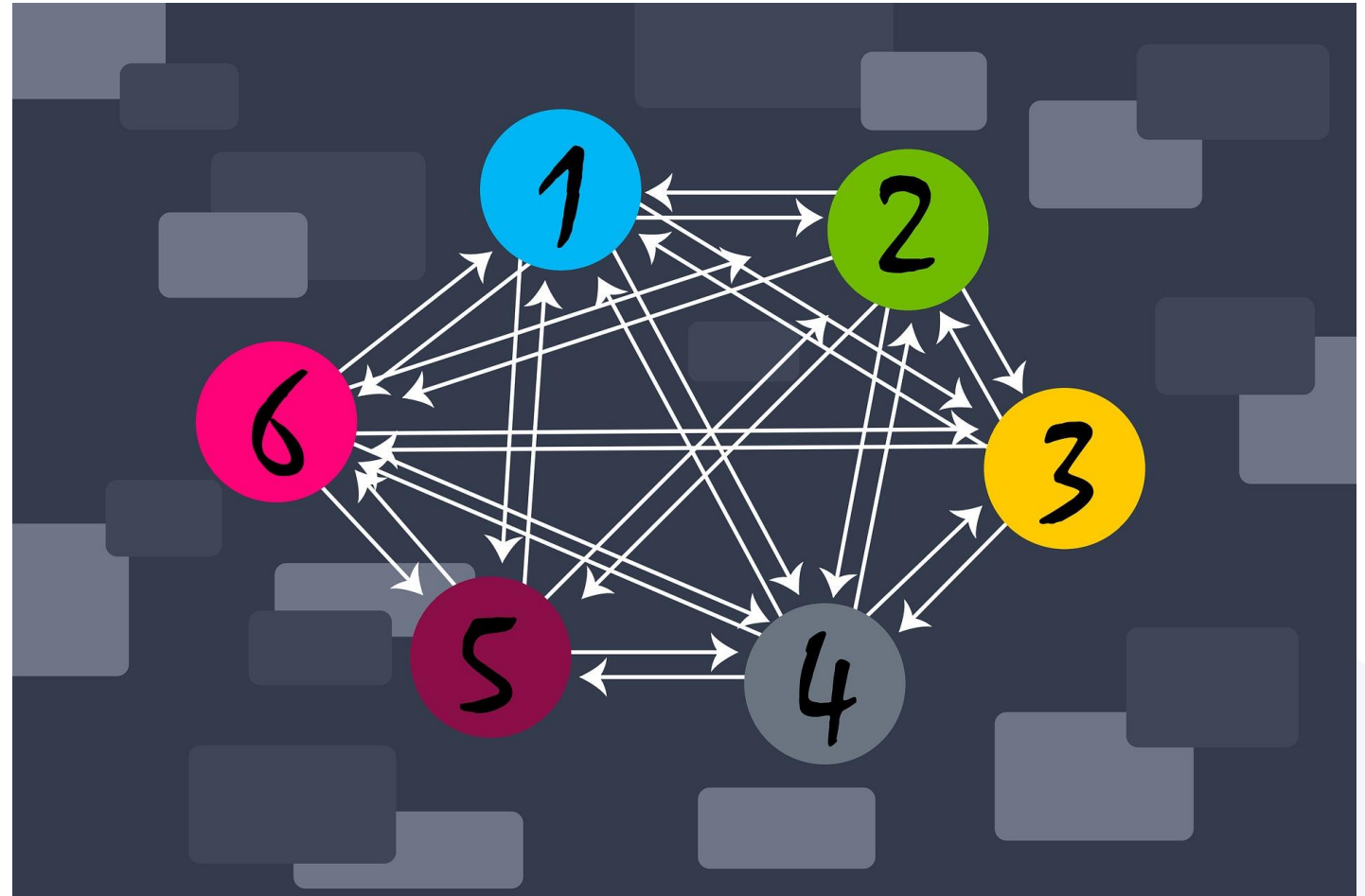
# Finding a suitable repository



The screenshot displays the re3data.org website interface. At the top, there is a navigation bar with links for Search, Browse, Suggest, Resources, and Contact, along with a DataCite logo. A dropdown menu under 'Browse' shows options: 'Browse by subject', 'Browse by content type', and 'Browse by country'. The main header features the re3data.org logo and the tagline 'REGISTRY OF RESEARCH DATA REPOSITORIES'. Below this is a search bar with a 'Search' button. The left sidebar contains a 'Filter' section with a 'Reset all' link. The filter categories include: Subjects, Content Types (with a list of data types and counts, such as 'Archived data (325)', 'Audiovisual data (371)', 'Configuration data (42)', 'Databases (304)', 'Images (1282)', 'Networkbased data (90)', 'Plain text (603)', 'Raw data (623)', 'Scientific and statistical data formats (769)', 'Software applications (236)', 'Source code (97)', 'Standard office documents (820)', 'Structured graphics (491)', 'Structured text (394)', and 'other (458)'), Countries, AID systems, API, Certificates, and Data access. The main content area shows a search bar, a pagination bar with 'Previous', '1', '2', '3', '4', '5', '6', '7', '...', '52', and 'Next', and a message 'Found 1282 result(s)'. Two repository cards are visible: 'The Fish Database of Taiwan' and 'The Infrared Space Observatory data archive'. The first card shows filters for Subject(s) (Life Sciences, Biology, Animal Ecology, Biodiversity and Ecosystem Research, Zoology), Content type(s) (Standard office documents, Images, Structured graphics, Audiovisual data, other), and Country (Taiwan, Province of China). The second card shows filters for Subject(s) (Astrophysics and Astronomy, Physics, Natural Sciences).

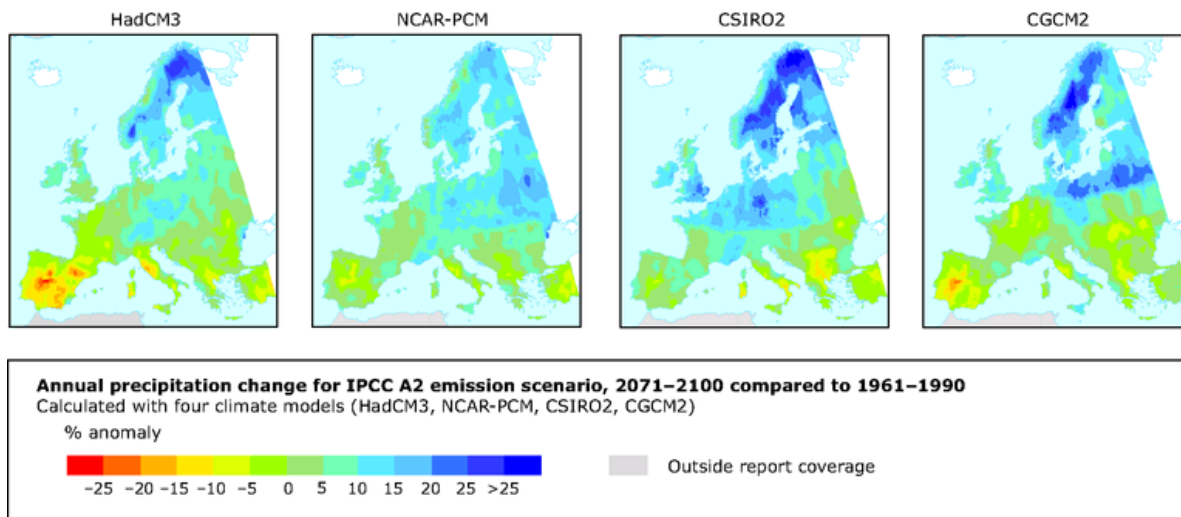
<https://www.re3data.org/>

Consider: how will you  
link to  
related outputs  
(e.g., ORCID, DOIs)





# Remember to also consider links to software, models and physical data



[https://www.eea.europa.eu/data-and-maps/figures/changes-in-annual-precipitation-for-the-ipcc-a2-scenario-2071-2100-compared-with-1961-1990-for-four-different-climate-models/chapter-3-map-3-1-belgrade-precipitation.eps/image\\_large](https://www.eea.europa.eu/data-and-maps/figures/changes-in-annual-precipitation-for-the-ipcc-a2-scenario-2071-2100-compared-with-1961-1990-for-four-different-climate-models/chapter-3-map-3-1-belgrade-precipitation.eps/image_large)

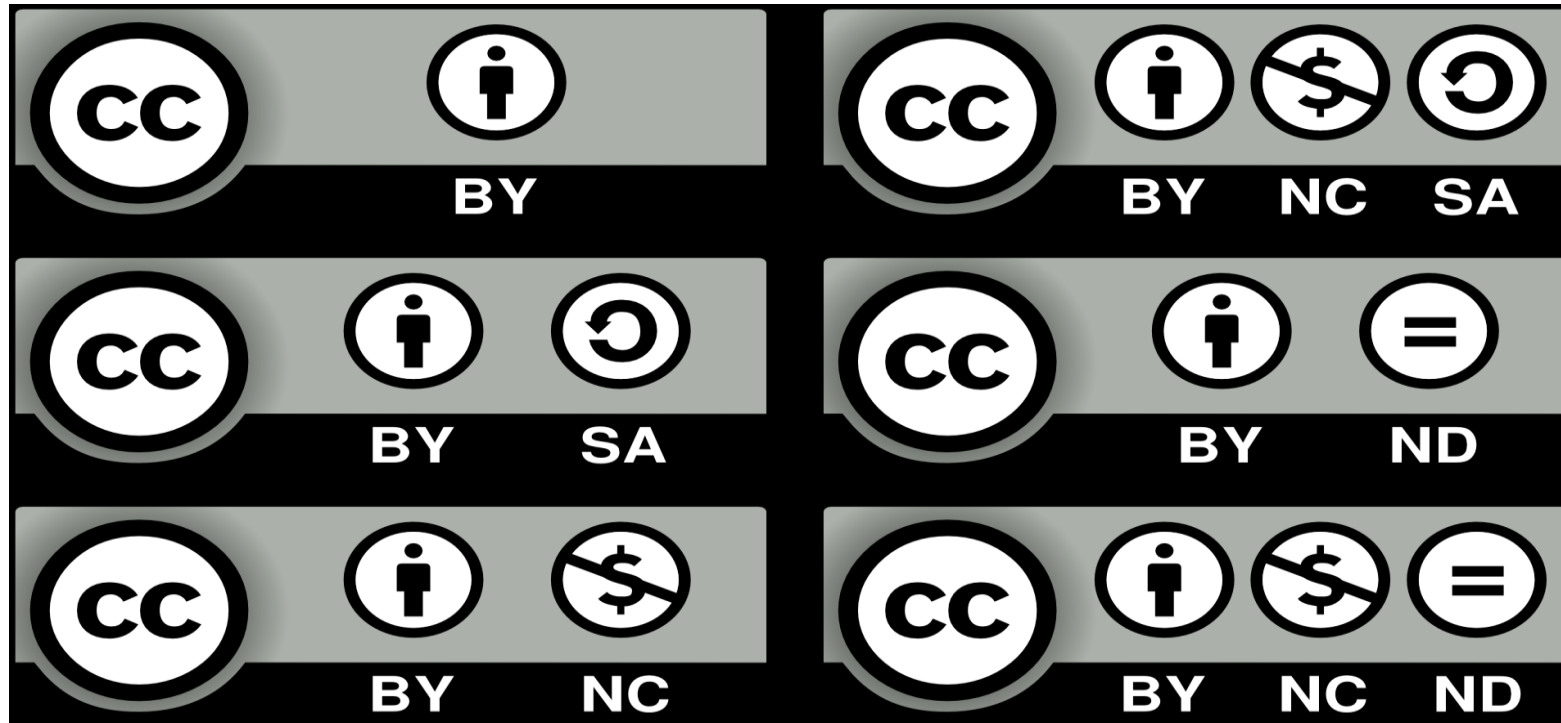


[http://www.ukcrcepxmed.org.uk/Coventry\\_Warwick\\_CRF/PublishingImages/Issue%20Bank%201.jpg](http://www.ukcrcepxmed.org.uk/Coventry_Warwick_CRF/PublishingImages/Issue%20Bank%201.jpg)



Image by Olalekan Oladipupo from Pixabay

# Consider: How will you license your data?



Creative Commons <https://creativecommons.org/>

# DMPonline <https://dmponline.dcc.ac.uk>

**Anyone can register  
for a free user  
account.**

[Home](#)[Public DMPs](#)[Funder requirements](#)[Help](#)

✓ Notice: Signed out successfully.

## Welcome

DMPonline helps you to create, review, and share data management plans that meet institutional and funder requirements. It is provided by the Digital Curation Centre (DCC).

Join the growing international community that have adopted DMPonline:



17,622 Users



203 Organisations



23,083 Plans



89 Countries

Some funders mandate the use of DMPonline, while others point to it as a useful option. You can [download funder templates](#) without logging in, but the tool provides tailored guidance and example answers from the DCC and many research organisations. Why not sign up for an account and try it out?

[Sign in](#)[Create account](#)

\* Email

\* Password

[Forgot password?](#)

☐ Remember email

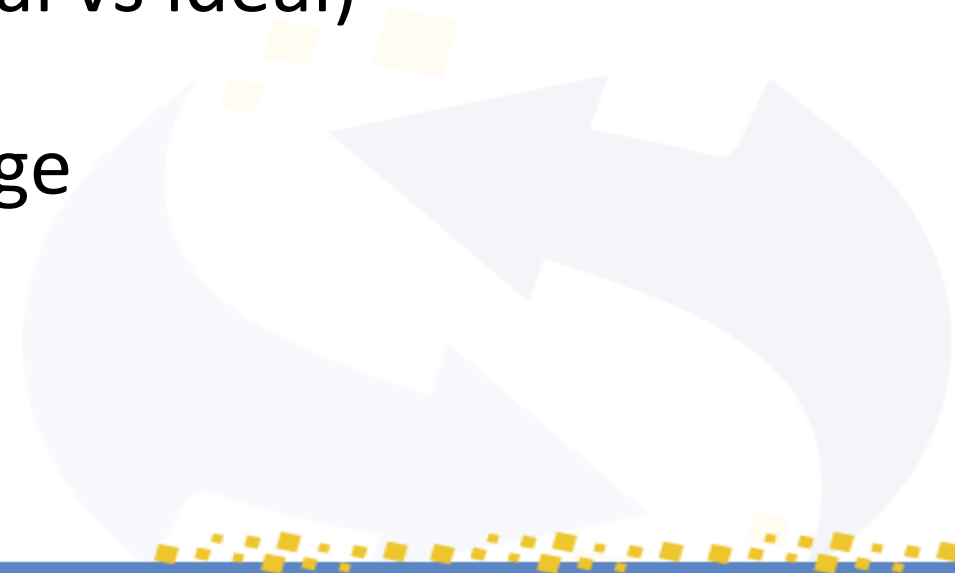
Sign in

- or -

Sign in with your institutional credentials

## Good practice when creating DMPs

- ☑ Start early in the research process
- ☑ Get guidance when writing the plan
- ☑ Be realistic about what you can do (real vs ideal)
- ☑ Update your DMP – things often change
- ☑ Use free tools





# Data management planning:

- Helps prevent data loss
- Helps you produce FAIR data and get more impact
- Supports research integrity and enables validation
- 
- Makes it easier to collaborate
- Leads to real world benefits!



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

<https://www.ebi.ac.uk/about/news/announcements/open-data-sharing-accelerates-covid-19-research>

Thanks – any questions?

