

The Journey to FAIR

Tools to aid you in your open data adventure

Kim Ferguson (RDM Specialist)

w/ materials from:

Maaïke Verburg (RDM Specialist) & Cees Hof (Data Station Manager Life Sciences)

23/05/2024 – NWO Life 2024 – Egmond aan Zee

DOI: 10.5281/zenodo.11260359



Who am I?

Kim Ferguson

→ Research Data Management Specialist at DANS-KNAW

DANS - Data Archiving & Networked Services

Databases including Data Stations Life Sciences

→ PhD in insect genetics, postdoc in metabarcoding

→ Open Science advocate

[depositing data sets,](#)

[publishing preprints,](#)

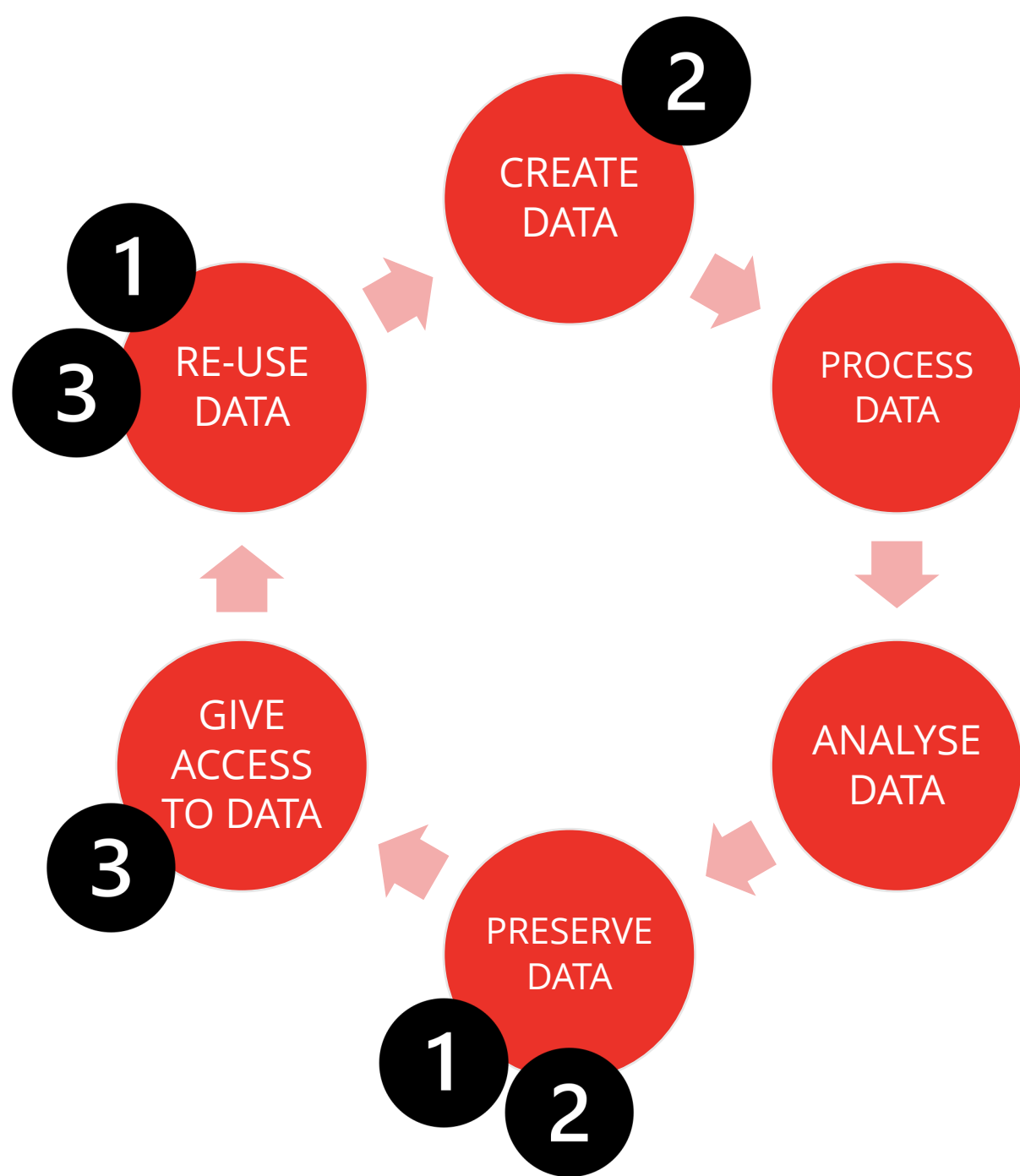
[open access publishing,](#) and

[sharing posters & presentations](#)



The (Simplified) Research Data Life Cycle

Three tools today:
1. Re3Data
2. FAIRsharing.org
3. FAIR-Aware



Data repositories:
good for finding data, good for storing data

What is archiving? What about publishing a data set?

Internal shared drives are useful for colleagues, but aren't the same as publishing a dataset in a repository

-
-
-
-
-
-



Repositories & FAIR

A good repository:

- Provides a Persistent Identifier (PID) to your dataset = Findable
- Enables the inclusion of rich metadata = Reusable
- Allows for human and machine findable and readable metadata = Findable + Accessible
- Uses Knowledge Organisation Systems (KOS), such as vocabularies, to improve the quality of metadata = Reusable
- Provides long term and easy access to your data = Accessible + Reusable
-

Where to deposit data, or find data for re-use

Depositing sequence data:

- ENA <> GenBank <> DDBJ (share sequences, trustworthy, all use MIBBI)
- Organism domain or species/family-specific databases? In addition, never only

Depositing everything else, or finding (non-sequence) data for re-use?

- Domain-specific repositories likely have more applicable metadata or keywords
- Some institutions or universities have their own repositories, making it easier for you to deposit
- National repositories like DANS Data Station Life Sciences is a good option as well

So what else is out there? Browse-by-subject in [Re3Data is a great starting point.](#)

1

Re3Data

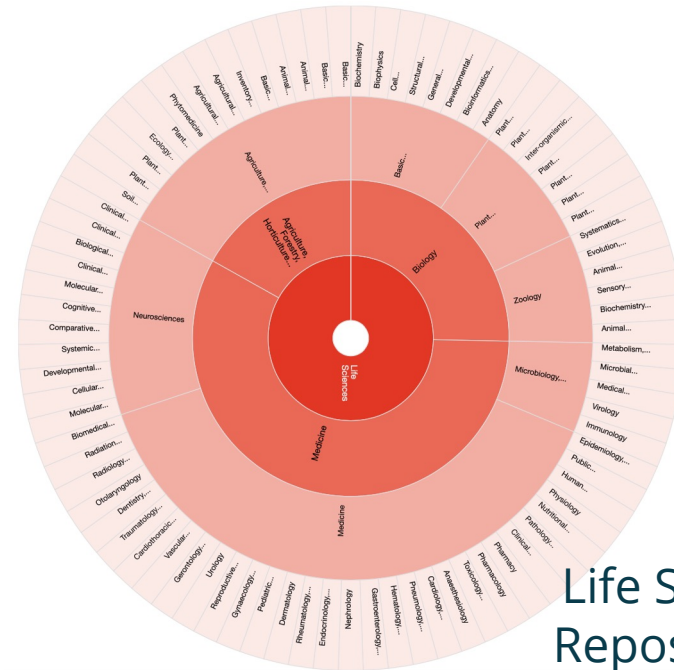
<https://www.re3data.org>

Over 2600 international repositories for a wide spread of domains

Plentiful filters to find appropriate repositories



All Categories



Life Science
Repositories
Categories

Wait a second... Metadata?

Findable

→ Assign persistent IDs, provide rich metadata, register in a searchable resource, ...

Accessible

→ Retrievable by their ID using a standard protocol, metadata remain accessible even if data aren't...

Interoperable

→ Use formal, broadly applicable languages, use standard vocabularies, qualified references...

Reusable

→ Rich, accurate metadata, clear licences, provenance, use of community standards...



2

FAIRSharing.org

What is it?

Curated resource on data and metadata standards and policies

Several ways to access:

- Browse by subject
- Use search by standard or policy
- View collections of resources

FAIRsharing.org standards, databases, policies search through all content Q SEARCH LOGIN

STANDARDS DATABASES POLICIES COLLECTIONS ORGANISATIONS ADD CONTENT STATS

A curated, informative and educational resource on data and metadata standards, inter-related to databases and data 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.
Provides humans and tools with access to trustworthy content to enable data management tasks.

RESEARCHERS DEVELOPERS & CURATORS JOURNAL PUBLISHERS LIBRARIANS & TRAINERS SOCIETIES & ALLIANCES FUNDERS

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](#)

1717 Standards

Terminology Artifact	794
Model/Format	590
Reporting Guideline	256
Identifier Schema	63

[VIEW ALL](#)

2098 Databases

Repositories	1077
Knowledgebases	857
Knowledgebase/Repositories	164

[VIEW ALL](#)

174 Policies

Journal	87
Funder	35
Society	16
Project	6

[VIEW ALL](#)

Note on FAIRSharing.org Life Cycle Status:

FAIRsharing uses indicators to show the life-cycle status of each standard

- R** **Ready** when a resource is considered suitable for use
- Dev** **In development** when a resource is being developed and may be used but may also be in a state of flux
- D** **Deprecated** when the community no longer mandates its use; this status is curated jointly with an explanation and, where available, a link to the standard that has superseded it, or been merged with it
- U** **Uncertain** when contact cannot be established with the community or owners of a resource, and therefore its current status cannot be determined; generally used on a temporary basis

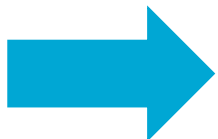
From Sandström, Lister, & Sansone 2023 [10.5281/zenodo.8186982](https://doi.org/10.5281/zenodo.8186982)

3



Your first step towards your FAIR data(set)

Answer 10 simple FAIR questions



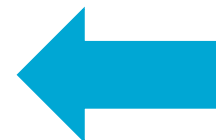
Read the guidance texts to develop FAIR knowledge and skills



Leave feedback to improve the tool



Receive your FAIR assessment scores

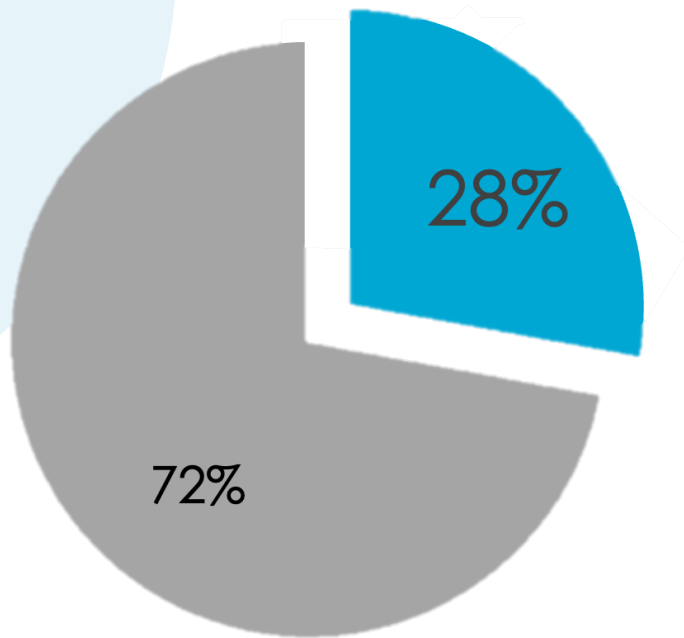


Create more FAIR data!



Spread the word!

Results so far



■ Life Sciences ■ Other Domains

Over 900 users with a 81% usability rating!

“The FAIR principles were not familiar to me at the time when I was trying to get published. The FAIR-Aware tool is informative and easy to use as a point of reference for my further publication career”

- Researcher from the life science domain

So what does it do?



Your first step towards your FAIR data(set)

→ [Let's take a look at FAIR Aware](#)

Note 1: The tool works best when you're about to deposit a data set, but can be useful to think about what you need to consider or when you're writing a data management plan

- Note 2: A lot of the elements can be taken care of by a trustworthy repository, which we will look at shortly after

But how much can we cover in 25 minutes?

[EMBL-EBI Training](#) > [On-demand training](#) > [Online tutorial](#) > **Bringing data to life**



- This online tutorial from EMBL-EBI is self-paced and can provide a great structure for learning data management, but you don't necessarily need it.

→ recommended for those who want to know more.

References/Further Reading

- Re3Data: <https://www.re3data.org/>
- FAIRsharing.org: <https://fairsharing.org/>
- FAIR-Aware: <https://fairaware.dans.knaw.nl/>

Additional resources

- Elixir RDM Kit: <https://rdmkit.elixir-europe.org/>
- BioSB - Netherlands Bioinformatics and Systems Biology research school: <https://www.dtls.nl/biosb/>
- EMBL-EBI Course on Research Data Management in Bioinformatics: <https://www.ebi.ac.uk/training/online/courses/bringing-data-to-life-data-management/>

More information

Visit our website www.dans.knaw.nl

And follow us online



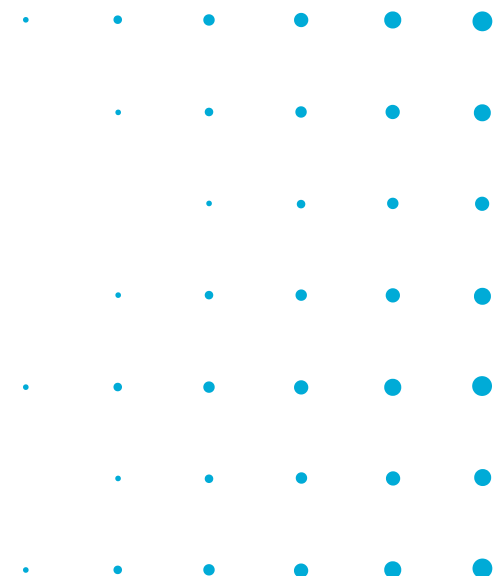
Mastodon @DANS_knaw_nwo



LinkedIn @DANS



X @DANS_knaw_nwo



- All emojis designed by [OpenMoji](#) – the open-source emoji and icon project. License: [CC BY-SA 4.0](#)