

Plan ahead: practical tools to make your data and software more FAIR

Workshop organised for the NWO BioPhysics 2022 on October 10 2022

[Link to the slides](#)

[Link to the collaborative google document](#)

Workshop Organizers:

- [Esther Plomp](#), TU Delft, Data Steward
- [Mateusz Kuzak](#) Training Programme Lead, [Lourens Veen](#) and [Jaro Camphuijsen](#), Research Software Engineers, from the Netherlands eScience Center
- [Kim Ferguson](#), DANS-KNAW, Research Data Management Specialist

Introduction by Esther

This workshop is split in two parts: the first part by Esther introduces the FAIR principles (Findable, Accessible, Interoperable and Reusable) and Data/Software Management Plans (**DMP**). After this introduction, we will have three sub sessions. These sessions will be introduced through short pitches, after which you can choose one session. All materials are available in the link to the slides.

Data Management Plan (DMP)

A Data Management Plan is a tool to structure the management of your data/software. A plan can help you to increase the quality of documentation, as well as prevent data loss and unfindable files (with the help of, for example, [file naming conventions](#)). The plan will also help you to consider how you can make your data/software follow the FAIR principles, which will be introduced later. In the plan you will also describe how you will share your software and data, so that others can validate your findings and to increase the impact of your research. Open data/code makes it easier for others to collaborate with you on projects, and results in [increased citations](#) to your work.

You can learn more about [Data Management Plans and their content via The Turing Way](#).

FAIR Principles

FAIR is an abbreviation for Findable, Accessible, Interoperable, and Reusable.

Findable

Your data/code is findable when you deposit your data in a **data repository** with **metadata** and a **persistent identifier**.

Data repository

A data repository is an online platform that allows you to share your data. Data repositories will curate research datasets and provide long term access to the finalised datasets (for at least 10-15 years). There are many different repositories. You can use the following resources to check whether your discipline has a recommended data repository:

- [Recommended Repositories](#) (nature)
- [Registry of Research Data Repositories](#)
- [Fairsharing.org](#)

You can always use one of the general repositories (such as [Zenodo](#), [DANS](#)).

Metadata

Metadata is information about data, for example:

- Contextual information
- Title, author, keywords
- When was the data generated? For what purpose?
- What is the expected size of the data?

There might be discipline specific metadata standards that you could use. Not every discipline has these type of standards. Use the following resources to check if there are any metadata standards that you could use:

- [FAIRsharing.org](#)
- [Research Data Alliance metadata directory](#)
- [Digital Curation Centre](#)

Persistent identifier

A persistent identifier is a long-lasting reference to a file, web page, or other object. Research articles generally have such a link, and you used a persistent identifier to get a hold of this text file. Another example is [ORCID](#), a persistent identifier for researchers.

Accessible

For your data to be accessible you will first have to consider what you can share. Not all research data can be openly shared. If you want to share personal data, data that can identify a human individual, you will need to first obtain consent and make sure you performed a risk assessment. You can still share your data using a data repository, but it might be needed to restrict access and have a protocol in place that re-users can follow to obtain access. You can also choose to only share your metadata publicly. A dataset that has public metadata instead of the full dataset is still following the Accessible principle.

Interoperable

For your data to be interoperable it is recommended to use open or common data formats, consistent vocabulary, and make use of discipline specific metadata standards where possible.

Reusable

For your data/code to be reusable, you need to apply a license. A licence specifies how others can reuse the data/code, see for example the [4TU.ResearchData license explanations](#). You will also need to document your data/code.

Licenses for data

- [Creative Commons](#) (Overview)
- [Creative Commons License Chooser](#)

Licenses for software

- [Choose a License](#)
- [tl;dr Legal](#)

FAIR is not:

The fair principles are not a standard, but guiding principles. Fair is also not the same as open data, as explained under the Accessible principle. While fair focuses on the usefulness of data, it does not say anything about the quality of the data. Learn more about the [fair principles via The Turing Way](#).

Pitches

1. Software licensing - Lourens Veen

This sub session will go into what software licenses are and which one you can choose to apply to your code/software.

2. Software citation - Jaro Camphuijsen

This sub session will teach you how to supply citation metadata and how to get a persistent identifier or DOI for your software. You will learn why you should make your software citable, how you can do this using the citation.cff file, and how you can automatically archive software releases using GitHub and Zenodo. You can find more information on [FAIR-software.nl](#).

3. FAIR-Aware tool - Kim Ferguson

In this sub session you will take your first steps towards your fair data(set). You will learn what FAIR practices can help you to create more fair data. You will assess this using 10 questions that contain some tips and tricks to help you, focusing on biological data and elements to keep in mind.

Contact information of your local research support

You can find information and contact details of your local research support in the [collaborative google document](#).

Thanks to NWO for supporting this workshop, in particular Naomi Chrispijn & Renée Calon!