



# Designing Training for FAIR and Open Data

Kim Ferguson & Maaïke Verburg  
Research Data Management Specialists at DANS

29 June 2023 – Open Science Conference

# Agenda

## Introduction

Introducing DANS, your trainers, FAIR assessment, and FAIR-Aware | 10 min – Maaïke

## Getting to know FAIR-Aware

Short hands-on exercise to discover the tool | 15 min – Maaïke

## Designing a training using FAIR-Aware

Exercise to consider all elements when designing a training | 70 min (including comfort break) - Kim

## Discussion and wrap-up

Plenary discussion on main takeaways, lessons learned, questions, etc. | 10 min - Maaïke

# Data Archiving and Networked Services



Centre of Expertise



Versatile repository



Active collaborator

# Who are you?

Go to <https://partici.fi/63422455>

Or go to <https://partici.fi/> and type  
Room number 6342 2455

Part 1: Introduction



# FAIR and Open data

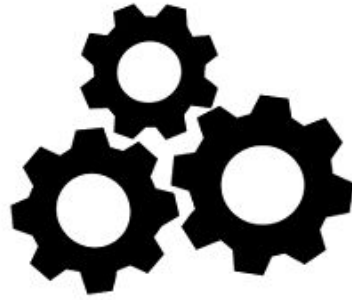
**F**  
Findable



**A**  
Accessible



**I**  
Interoperable



**R**  
Reusable



# Different FAIR assessment tools

Expectations → Monitoring and evaluation → Metrics and tools

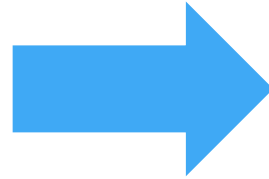
- Different aims, purposes, target objects, audiences, execution types, etc.
- Assessment of FAIR-enabling qualities | Assessment of holdings | Educating

[FAIRassist.org](https://FAIRassist.org)

# Background



- March 2019 - February 2022
- Goal: Practical solutions for the use of FAIR principles throughout the research data life cycle



- June 2022 - May 2025
- Goal: Support the implementation of FAIR-enabling practices across scientific communities and research outputs
- Takes forward FAIRsFAIR outputs

# Projects' FAIR assessment tools



How can you make new data more FAIR?  
<https://www.fairsfair.eu/fair-aware>

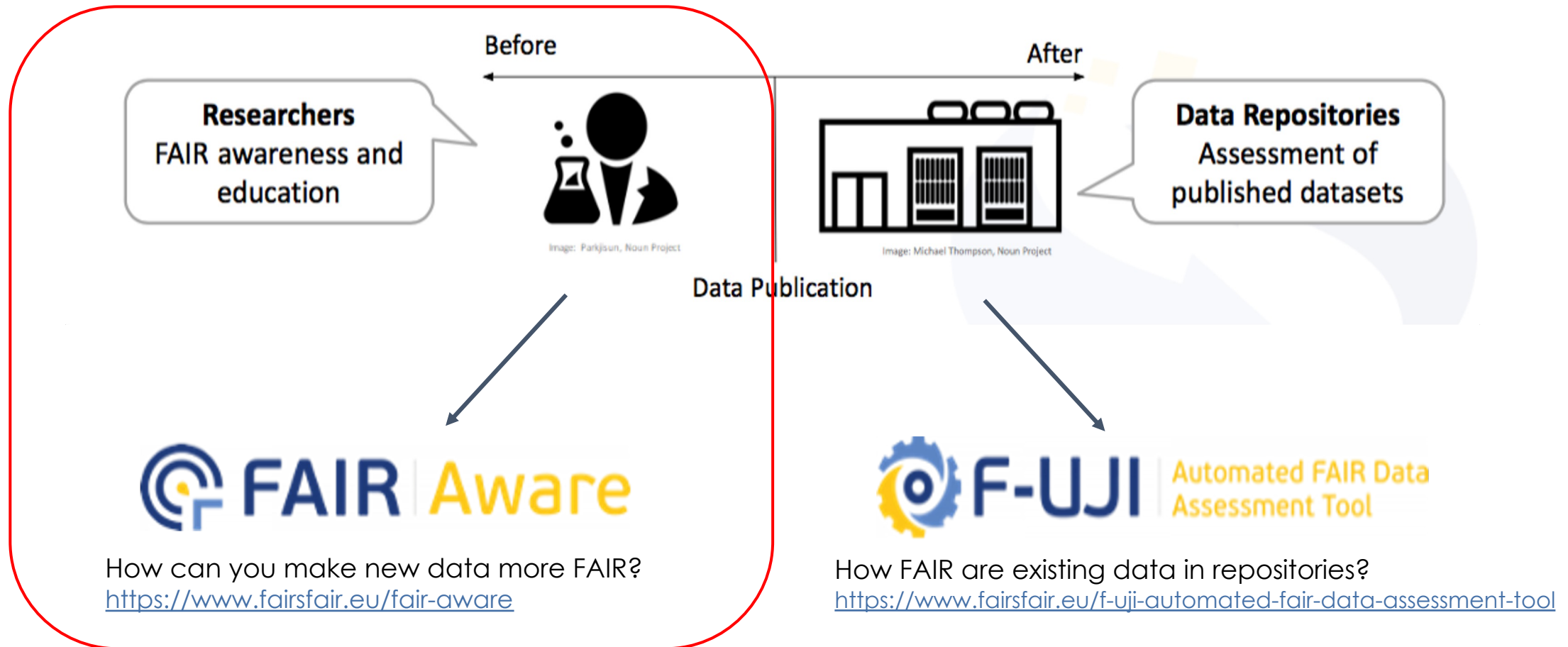


How FAIR are existing data in repositories?  
<https://www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool>




# Projects' FAIR assessment tools

## Awareness



Aim: help researchers and data professionals create more FAIR data before depositing in a repository

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

FAIR questions 

**F**

**A**

**FINDABLE**

1. Are you aware that a data(set) should be assigned a globally unique persistent and resolvable identifier when deposited with a data repository?  Yes  No
2. Are you aware that when you deposit a data(set) in a data repository, you will need to provide discovery metadata in order to make the data(set) findable, understandable and reusable to others?  Yes  No
3. Are you aware that the data repository providing access to your data(set) should make the metadata describing your data(set) available in a format readable by machines as well as humans?  Yes  No

**ACCESSIBLE**

4. Are you aware that access to your data(set) may need to be controlled and that metadata should include licence information under which the data(set) can be reused?  Yes  No
5. Are you aware that metadata should remain available over time, even if the data(set) is no longer accessible?  Yes  No

**INTEROPERABLE**

6. Are you aware that the metadata describing your data(set) should use controlled vocabularies?  Yes  No

**1. Are you aware that a data(set) should be assigned a globally unique persistent and resolvable identifier when deposited with a data repository?**

**What does this mean?**

A **persistent identifier** is a long-lasting reference to a resource. The **data(set)** you deposit in a **data repository** should be assigned a globally unique, persistent and resolvable identifier (PID) so that both humans and machines can find it. Persistent identifiers are maintained and governed so that they remain stable and direct the users to the same relevant object consistently over time. Examples of PIDs include Digital Object Identifier (DOI), Handle, and Archival Resource Key (ARK).

**Why is this important?**

If your data(set) or metadata does not have a PID, you run the risk of "link rot" (also known as "link death"). When your data(set) or metadata is moved, updated to a new version, or deleted, older hyperlinks will no longer refer to an active page. Without a PID, others will not be able to find or reuse your data(set) or metadata in the long-term.

**How to do this?**

When you upload your data(set) or metadata to a data repository, the data repository (or other service providers) usually assigns a PID. Repositories ensure that the identifier continues to point to the same data or metadata, according to access terms and conditions you specified.

There are many different types of PIDs, each with their own advantages, disadvantages, and disciplines they are typically used in. Generally speaking, the data repository will have thought about these aspects before deciding which PID type to use. In case you have to choose the PID type yourself, you can visit the Knowledge Hub on the PID Forum for guidance. Some disciplines or organisations also provide tools to help you make this choice, see for example this Persistent Identifier Guide for cultural heritage researchers. Once you have chosen a PID type, you can search for data repositories providing that specific PID in registries such as Re3data or FAIRsharing (see related databases).

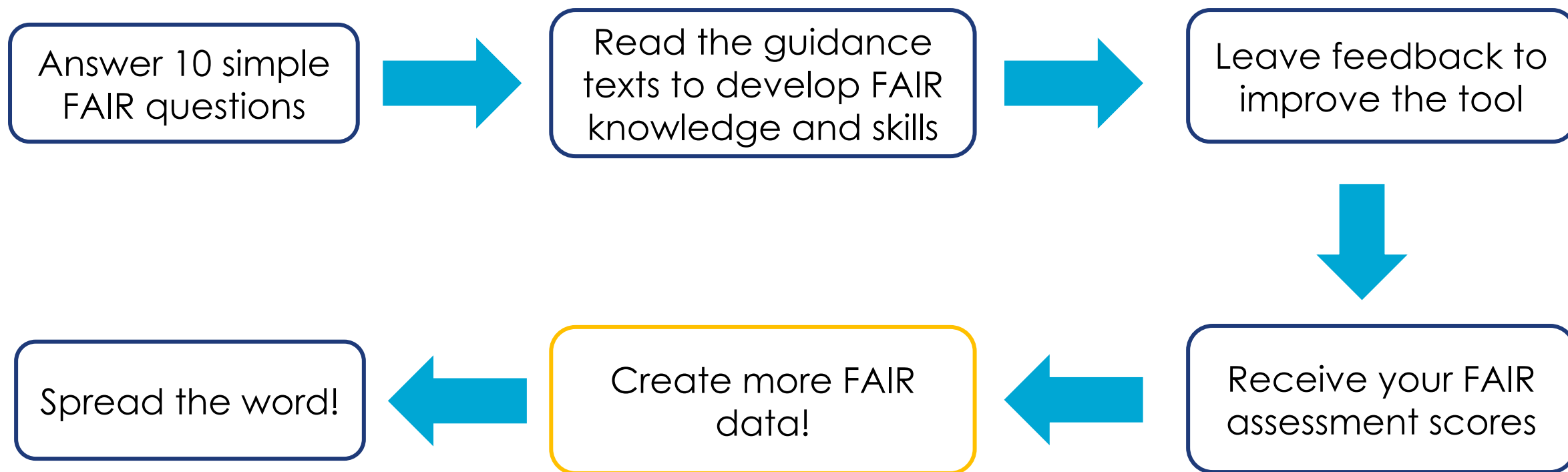
Not all data you produce during your research will need a PID. In general, those that underpin published findings or have longer term value are worth assigning a PID. If in doubt about which data should be allocated a PID, speak to your local research data management support team or the data repository.

**Want to know more?**

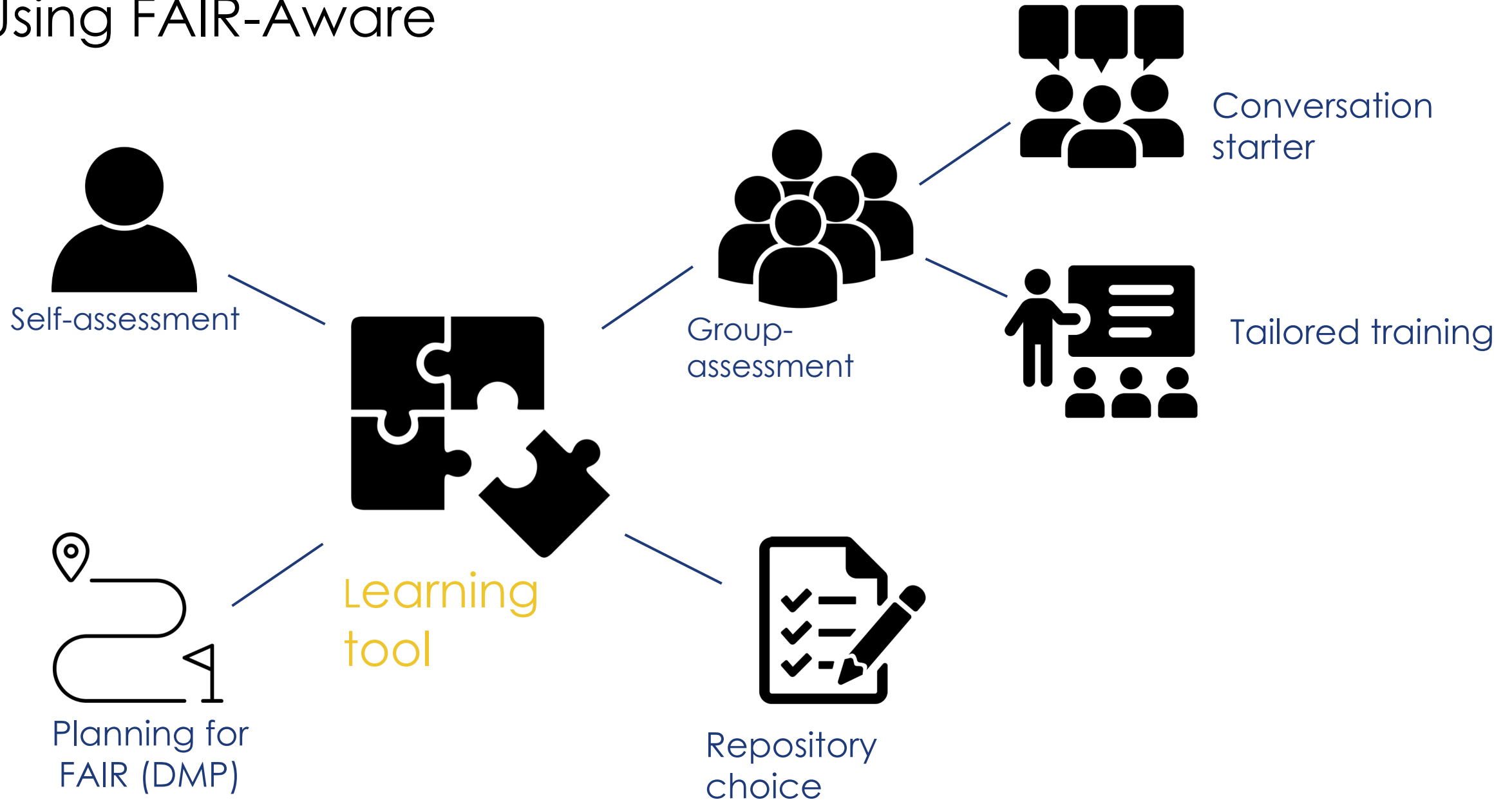
Did you know that a PID can refer to any kind of resource? Besides publications or datasets, a PID can also refer to, for example, a person, a scientific sample, a funding body, a set of geographical coordinates, an unpublished report, or a piece of software. Depending on what you find important to link to, you might want to consider using a PID for one or more of these resource types.

Persistent identifiers may point to a data file, a web service response that contains data values, or ideally to an online page that contains metadata for context and the link to access the actual data or details about how to request access. The technical process of translating the identifier to a location is called 'resolving' an identifier.

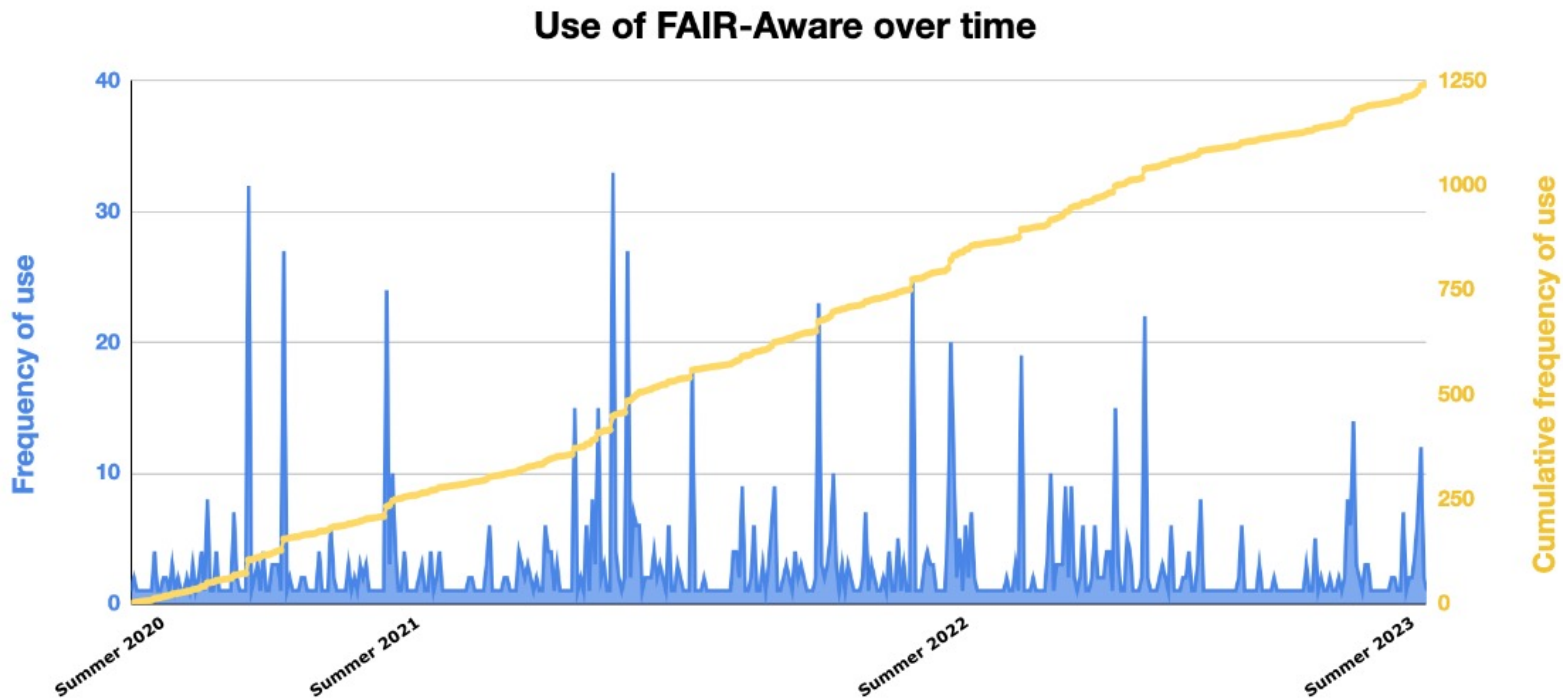
Close



# Using FAIR-Aware



# Results so far



“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

Over 1200 users with a 81% usability rating!

# Exercise: Get to know FAIR-Aware

Plan for FAIR data - Use FAIR-Aware guidance for DMP writing!

Where in FAIR-Aware can you find help for the following DMP elements?

1. In what format will you collect / archive your data?
2. What metadata will be provided to help others discover the data?
3. Who will be responsible for data management (specifically archiving)?



<https://fairaware.dans.knaw.nl/> | Science Europe DMP template ([Link](#))

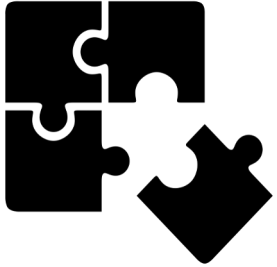
Share your answer at <https://partici.fi/63422455> (Part 2: FAIR-Aware)

10 minute break!

When we reconvene at \_\_\_\_\_, we will work together on

Designing a training using FAIR-Aware

# Elements to consider when designing a training (1)



In advance...

- Become familiar with the tool
- Learning outcomes/objectives
  - Essentially, what do you want your trainees to take away from the training?
  - Bloom's Taxonomy of Educational Objectives<sup>1</sup> can be helpful if you're uncertain
- Example datasets for target audiences
  - Domain specific?
  - Archive/repository specific?
  - Use cases for interdisciplinary research?

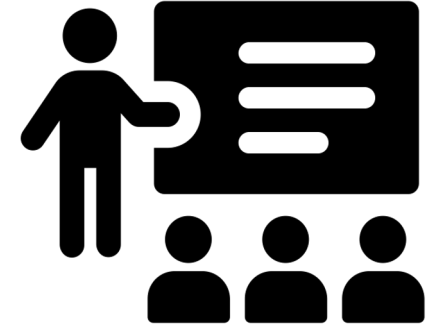
1. Stapleton-Corcoran, E. (2023). "Bloom's Taxonomy of Educational Objectives." Center for the Advancement of Teaching Excellence at the University of Illinois Chicago. Retrieved 27.06.2023 from <https://teaching.uic.edu/blooms-taxonomy-of-educational-objectives/>



# Elements to consider when designing a training (2)

## Specifics:

- Target audience and size
- Programme and technical needs
- Support needed
- Sharing materials before and/or after



What is under your control vs. What is not under your control

- Disaster planning – technical issues, logistic issues, uncertainties...

# Exercise: Designing a FAIR-Aware training (collaboration!)

60 mins - there will be a break in the middle, but take a moment when you need to.

- Collaboration element: everyone is welcome to opt-in to be a Contributor on the Zenodo record from this training!
  - Name in Column A (if you do not opt in to being a Contributor, leaving blank is fine)
  - Indicate Yes or No in Column B for being named as a Contributor
  - Affiliation in Column C
  - ORCID ID in Column D

Why collaborate? Generating ideas together, reducing the load, and different perspectives

# [Working on collaboration doc]

Spreadsheet/Google Sheet for collaboration <https://tinyurl.com/OSC-FAIRtraining>

- Approx 30 min for training design
  - Approx 20 min for disaster planning
- >> final version uploaded to Zenodo for July 6th

Resources for after:

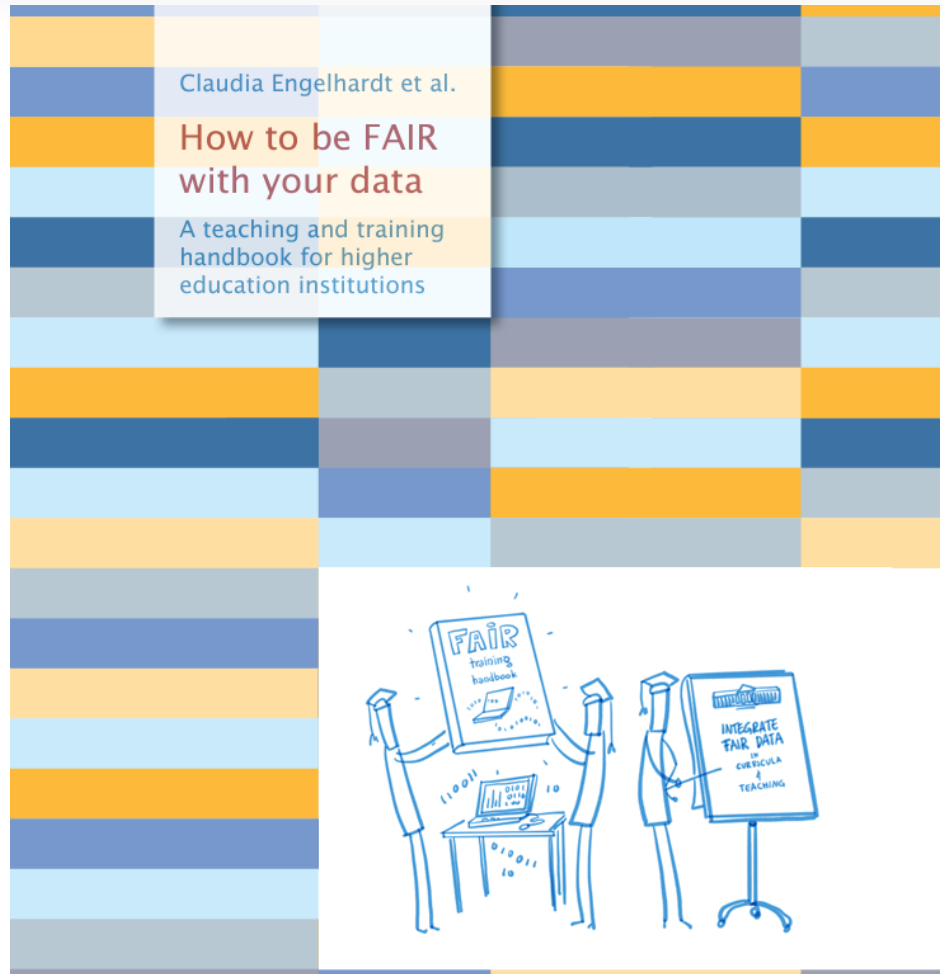
- Template of activity sheet (for your own (re-)use): <https://doi.org/10.5281/zenodo.8089632>
- Training exercise with worked examples: <https://doi.org/10.5281/zenodo.8089501>

# Discussion and wrap-up

- How easy or difficult was it for you to design your training?
- Which aspects were new for you to consider?
- Are you planning to deliver the training you designed?
- What are your main takeaways from today's session?

You can raise your hand / unmute to share your thoughts!  
Or share in Part 3: Final thoughts, or the Q&A section

# Additional training resource



## FAIR Teaching Handbook<sup>1</sup>

- FAIRsFAIR output, from 40+ trainers/authors

2. Engelhardt, Claudia, et al. 2022. 'How to Be FAIR with Your Data: A Teaching and Training Handbook for Higher Education Institutions'. FAIR Teaching Handbook. 27 June 2023. <https://fairsfair.gitbook.io/fair-teaching-handbook/>. (available as PDF as well)

# Thank you for your attention!



Live Feedback

Let us know what you thought of the session! <https://partici.fi/63422455>



Kim Ferguson | [kim.ferguson@dans.knaw.nl](mailto:kim.ferguson@dans.knaw.nl) | @kfergy

Maike Verburg | [maaike.verburg@dans.knaw.nl](mailto:maaike.verburg@dans.knaw.nl) | @MaikeVerburg

FAIR-Aware | [fair-aware@dans.knaw.nl](mailto:fair-aware@dans.knaw.nl) | [fairaware.dans.knaw.nl](http://fairaware.dans.knaw.nl)

DANS | [info@dans.knaw.nl](mailto:info@dans.knaw.nl) | @DANS\_knaw\_nwo | [dans.knaw.nl/en](http://dans.knaw.nl/en)

## Answers to the short exercise

Where in FAIR-Aware can you find help for the following DMP elements?

# In what format will you collect / archive your data?



## REUSABLE

7. Are you aware that provenance information about the collection and/or generation of data should be included in the metadata? ⓘ  Yes  No
8. Are you aware that metadata describing your data should follow the specifications of a community-endorsed standard? ⓘ  Yes  No
9. Are you aware that data should be deposited preferably in a file format that is open - to support reuse - and supported by the repository for long-term preservation? ⓘ  Yes  No
10. Are you aware that maintaining your dataset FAIR over time requires professional data curation and preservation? ⓘ  Yes  No

### 9. Are you aware that data should be deposited preferably in a file format that is open - to support reuse - and supported by the repository for long-term preservation? ✕

File formats refer to methods for encoding digital information. If there is no open format available, then you may also use a proprietary format. For example, CSV for tabular data, NetCDF for multidimensional data and GeoTIFF for raster imagery.

Data should be made available in a recommended file format that is accepted by the research community to enable data sharing and reuse, and by the repository to enable long-term preservation. Repositories normally have different recommended standard file formats based on data types (for example: [UK Data Service](#)) and [ISO/TR 22299 Document management - Digital file format recommendations for long-term storage](#)).

Recommended formats are widely used and supported by the most common software and tools ensuring that your data can be read in the future, but they will also help increase the reusability and interoperability. Using recommended formats enables data to be loaded directly into the software and tools used for data analysis. It makes it possible to easily integrate your data with other data using the same format. The use of recommended formats will allow migration of the format to a newer one, in case a preferred format becomes outdated.

Close



# What metadata will be provided to help others discover the data?



## FINDABLE

1. Are you aware that a dataset should be assigned a globally unique persistent and resolvable identifier when deposited with a data repository?  Yes  No
2. Are you aware that when you deposit a dataset with a repository, you will need to provide some details (known as discovery metadata) in order to make the data findable, understandable and reusable to others?  Yes  No
3. Are you aware that the repository providing access to your dataset should make the metadata describing your datasets available in a format readable by machines as well as humans?  Yes  No

## 2. Are you aware that when you deposit a dataset with a repository, you will need to provide some details (known as discovery metadata) in order to make the data findable, understandable and reusable to others?

The necessary metadata is descriptive information about the data object (e.g. creator, title, publisher, creation and publication date, summary and keywords describing the data) provided by the researcher when documenting the dataset for deposit in a data repository.

Since the metadata required depends on the type of users and the intended applications, the focus here is on ensuring the minimum descriptive information needed to enable data discovery and citation by potential users - even those from other domains. The metadata should include the unique, persistent and resolvable identifier (PID) for the data so that users can discover and access the data. In cases where the data cannot be shared openly for ethical, legal or commercial reasons, the metadata should make clear how a potential reuser can request legitimate access.

### Data content (relates to Reusability in the FAIR principle)

The content of the dataset should be specified in the metadata. It should be an accurate reflection of the actual data deposited. Examples of the properties specifying data content are: resource type (e.g., data or a collection of data), variable(s) measured or observed, method, data format and size. Ideally, semantic vocabularies should be used to describe data content (e.g., variable) to support interdisciplinary reuse. For example, variables should be described by providing a link with a PID to a resource (such as a landing page) containing a structured description of the concept, references and links to other relevant concepts. The access to the linked concepts should be available through the query interface (e.g., using SPARQL or GraphQL).

### Links to other research outputs (relates to Interoperability in the FAIR Principles)

# Who will be responsible for data management (specifically archiving)?



## REUSABLE

7. Are you aware that provenance information about the collection and/or generation of data should be included in the metadata? [i](#)

Yes

No

8. Are you aware that metadata describing your data should follow the specifications of a community-endorsed standard? [i](#)

Yes

No

9. Are you aware that data should be deposited preferably in a file format that is open - to support reuse - and supported by the repository for long-term preservation? [i](#)

Yes

No

10. Are you aware that maintaining your dataset FAIR over time requires professional data curation and preservation? [i](#)

Yes

No

## 10. Are you aware that maintaining your dataset FAIR over time requires professional data curation and preservation? [x](#)

Data curation is the active and ongoing management of data from the point of creation to ensure that it is fit for contemporary purpose and available for discovery and reuse. Likewise, digital preservation refers to the series of managed activities necessary to ensure continued access to and reusability of digital materials for as long as necessary (i.e FAIR).

Professional data curation and preservation require people, skills and technology. They are central to the mission of several academic institutions, such as universities, libraries, data repositories and archives. Specifically, Trustworthy Digital Repositories (TDRs) play a critical role in both making and preserving data FAIR over time. TDRs provide support and take responsibility for the curation and preservation of data with different levels of FAIRness. Digital repositories demonstrate their trustworthiness, for example, through certification with community-endorsed standards such as the CoreTrustSeal [↗](#), DIN31644/NESTOR [↗](#) and ISO163638 [↗](#). Trustworthy Data Repositories can be found on registries. For example, CoreTrustSeal certified repositories can be filtered on Re3data [↗](#).

[Want to know more?](#)

Close