Data Archiving and Networked Services

DANS

FAIR Aware

Designing Training for FAIR and Open Data

Kim Ferguson & Maaike Verburg
Research Data Management Specialists at DANS

29 June 2023 – Open Science Conference

10.5281/zenodo.8090163

Agenda

Introduction

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

Getting to know FAIR-Aware

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

Designing a training using FAIR-Aware

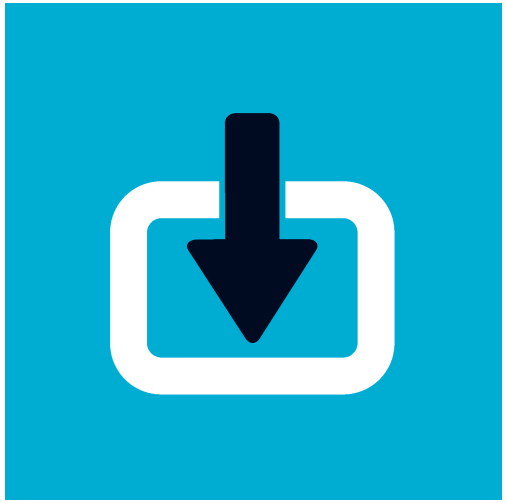
Exercise to consider all elements when desiging a training | 70 min (including comfort break) - Kim

Discussion and wrap-up

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

Data Archiving and Networked Services

Centre of Expertise



Versatile repository

Active collaborator

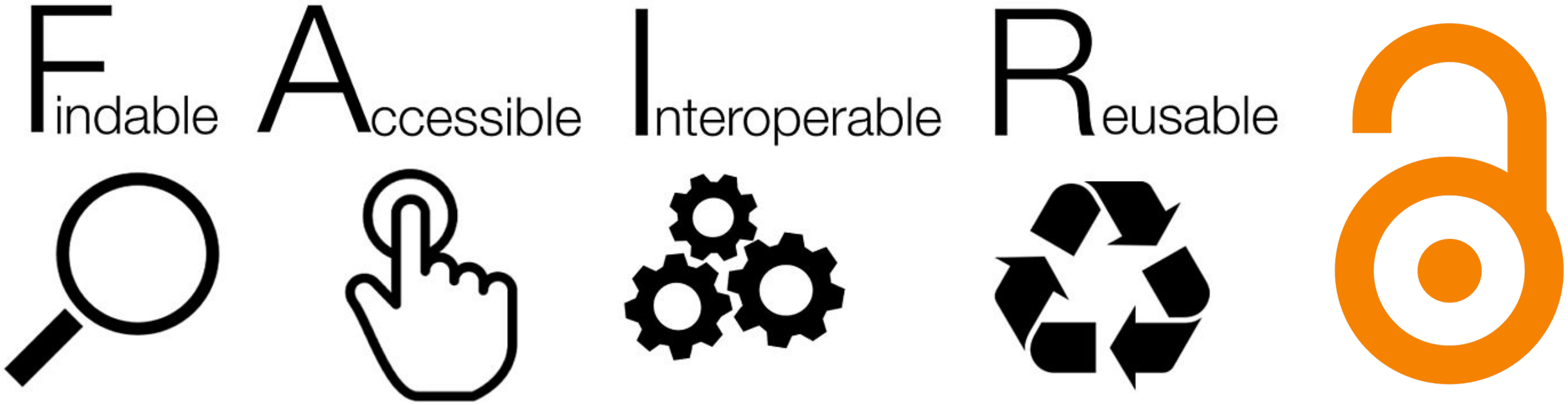
Who are you?

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



Different FAIR assessment tools

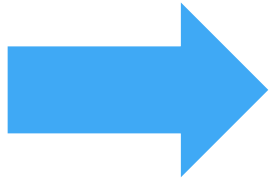
Expectations → Monitoring and evaluation → Metrics and tools

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

Background

FAIRSFAIR
Fostering Fair Data Practices in Europe

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



FAIR-IMPACT
adoption & support for FAIR implementation in EOSC

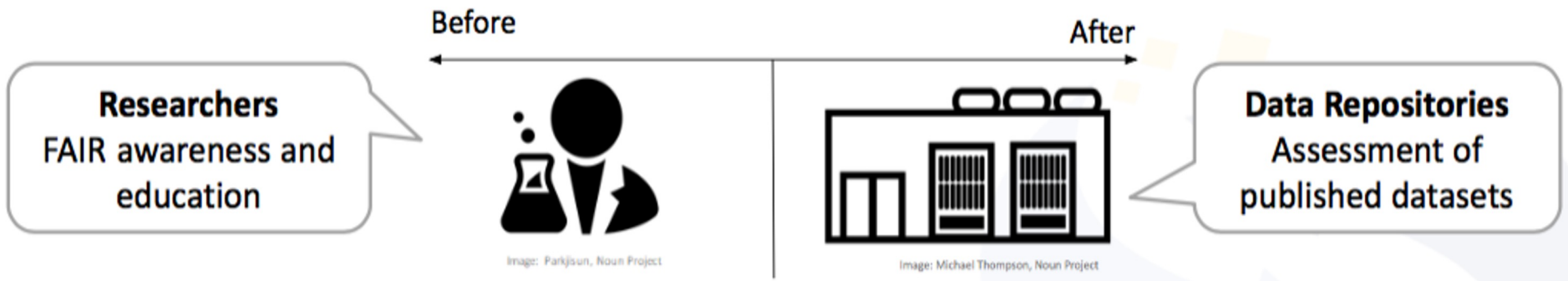
- June 2022 - May 2025
- Goal: Support the implementation of FAIR-enabling practices across scientific communities and research outputs
- Takes forward FAIRsFAIR outputs

Data Archiving and Networked Services
DANS

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.

FAIR-IMPACT is funded by the European Union

Projects' FAIR assessment tools



FAIR Aware

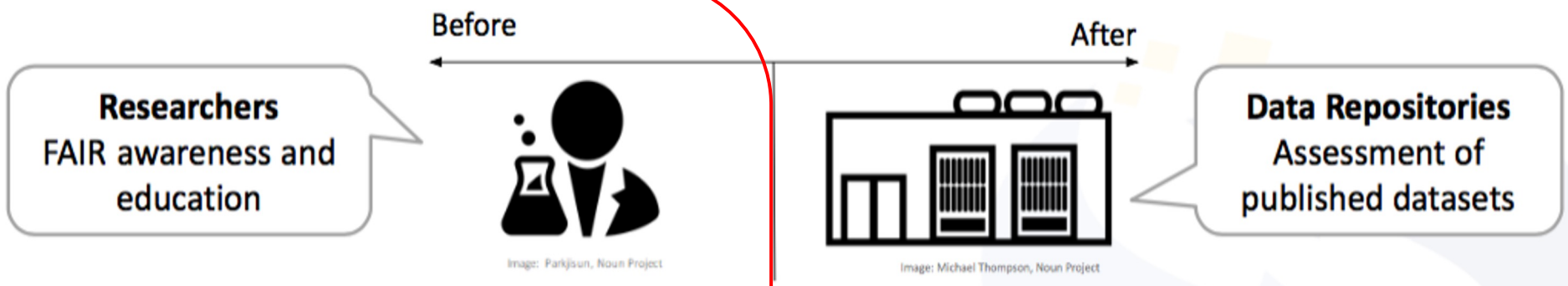
How can you make new data more FAIR?
https://www.fairsfair.eu/fair-aware

F-UJI Automated FAIR Data Assessment Tool

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

Projects' FAIR assessment tools

Awareness



FAIR Aware

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

F-UJI Automated FAIR Data Assessment Tool

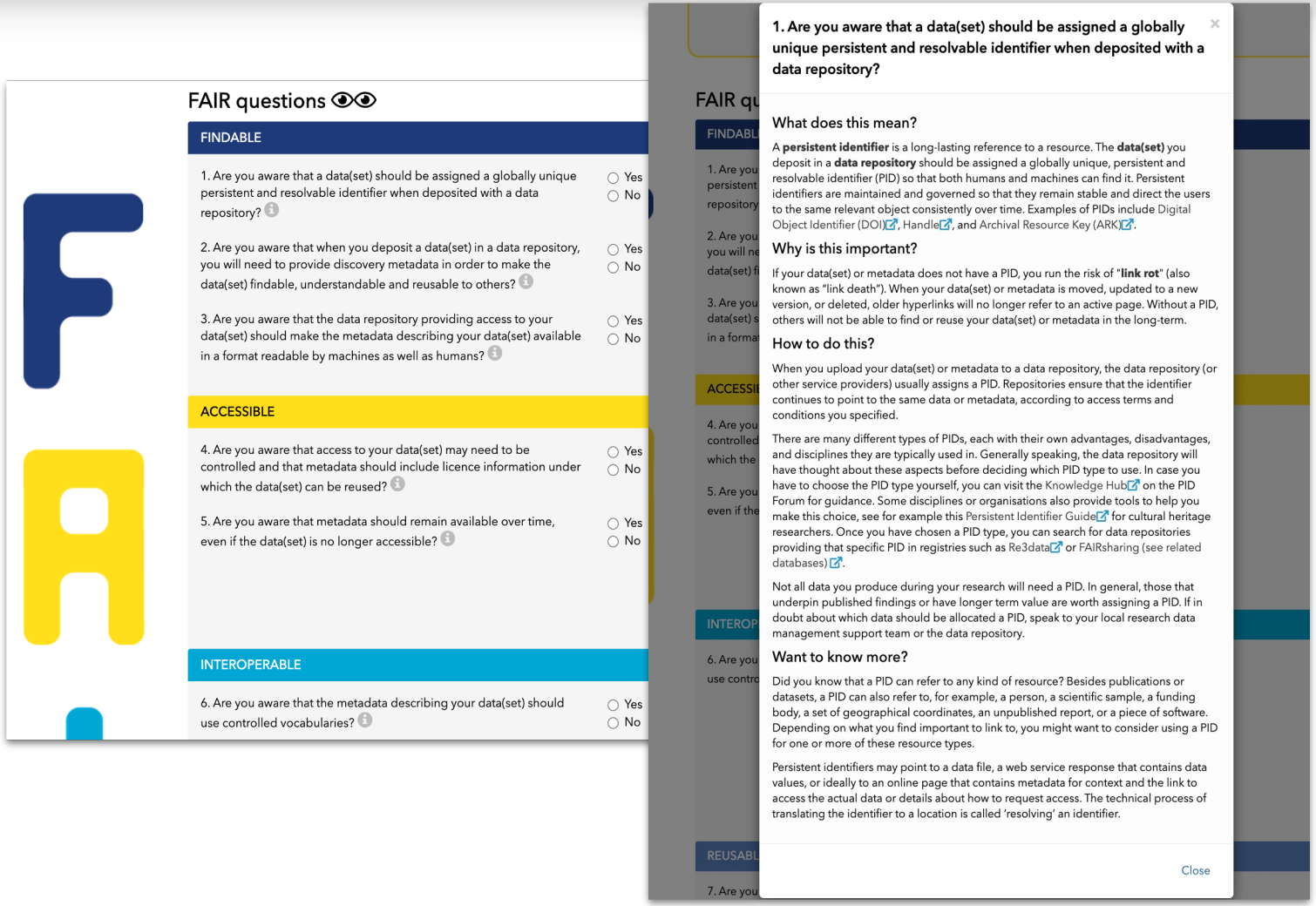
How FAIR are existing data in repositories?
https://www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool

FAIR Aware

Your first step towards your FAIR data(set)

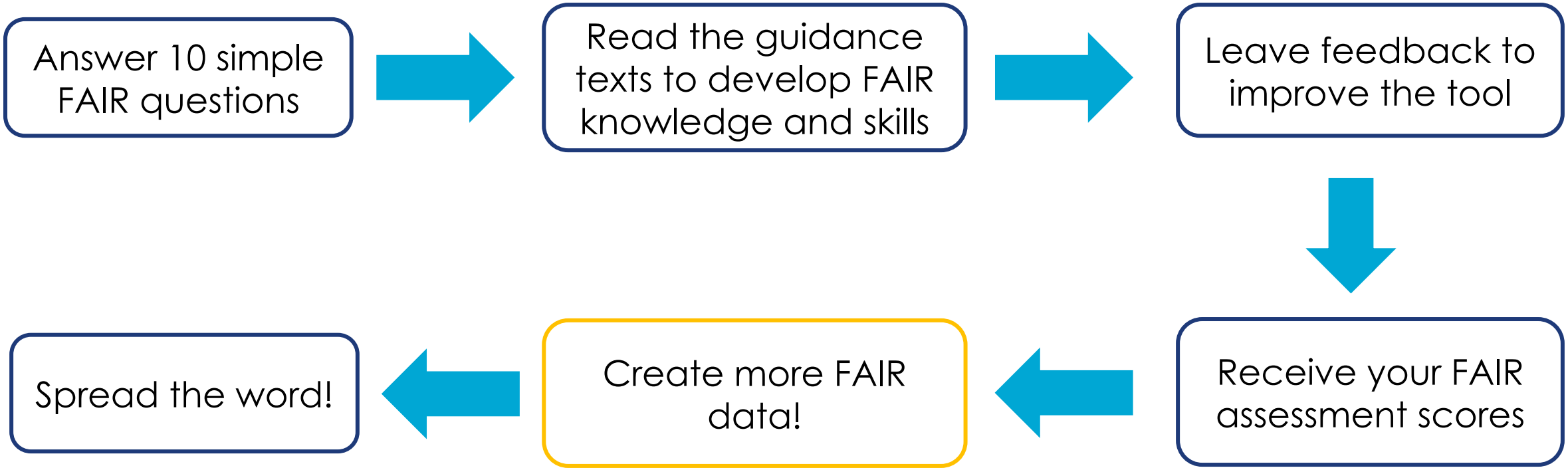
Aim: help researchers and data professionals create more FAIR data before depositing in a repository

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

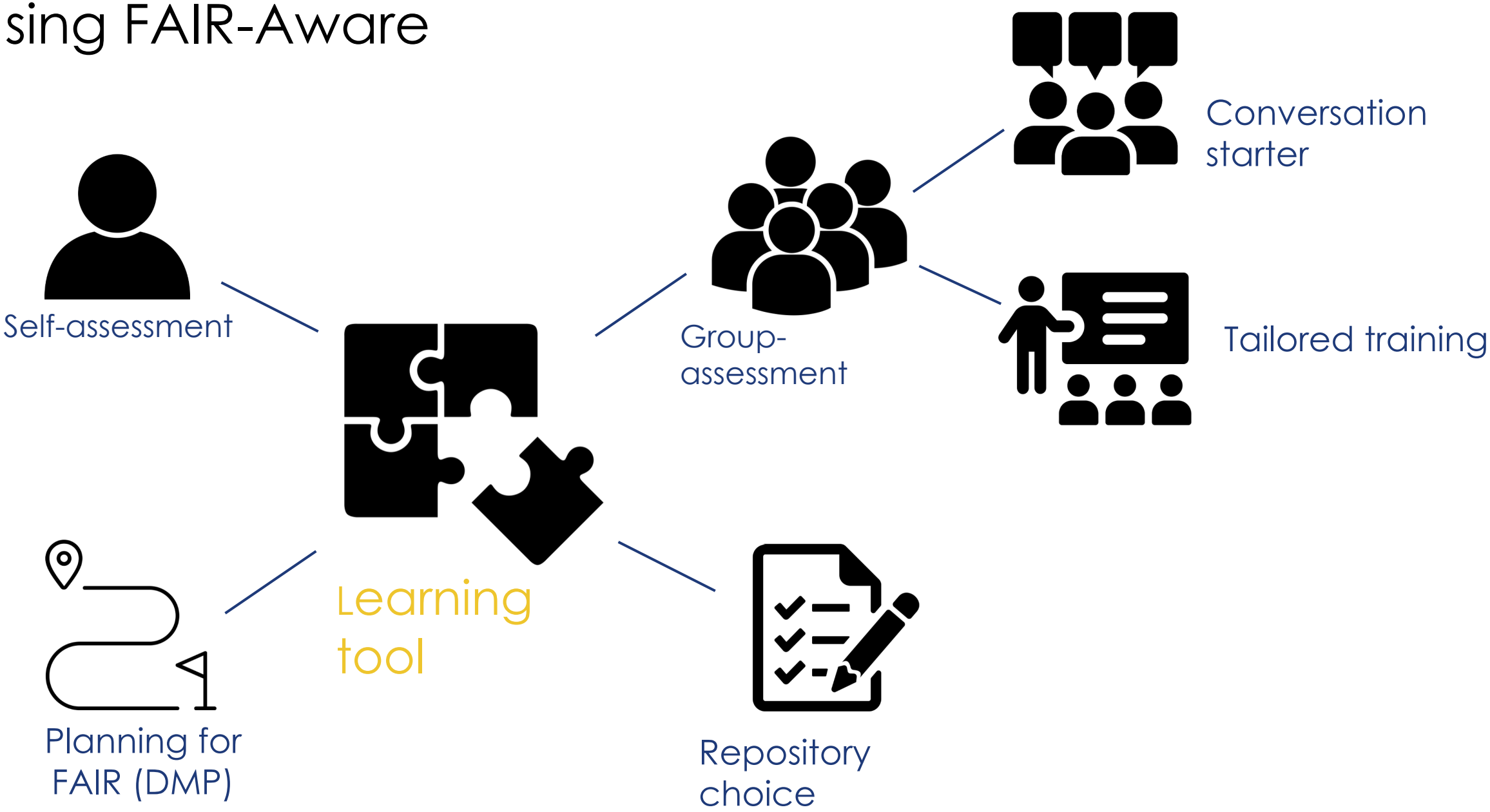


FAIR Aware

Your first step towards your FAIR data(set)

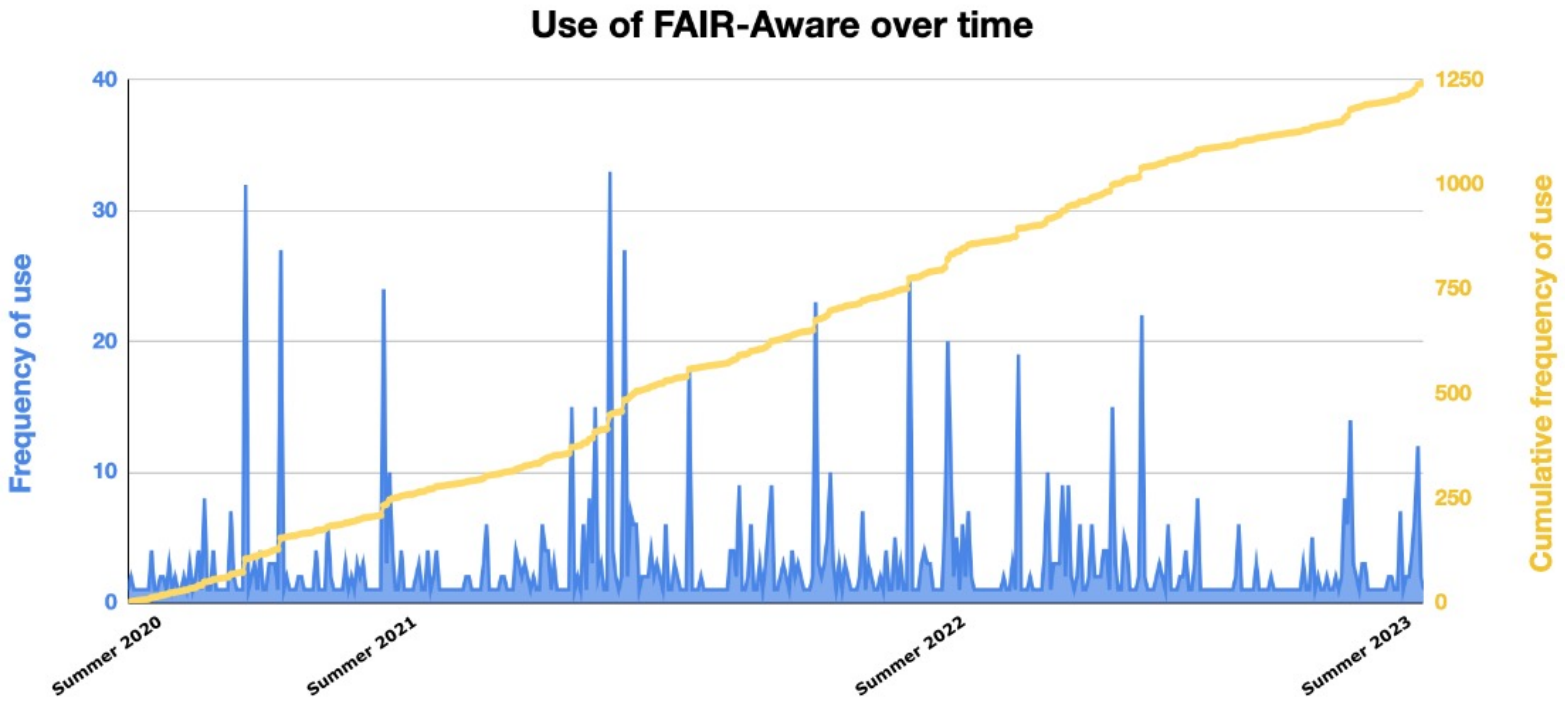


Using FAIR-Aware



Images from the Noun Project

Results so far



Over 1200 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

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)

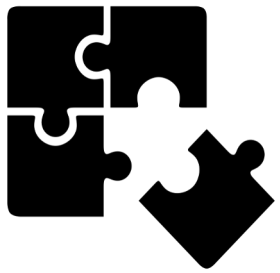
Share your answer at https://partici.fi/63422455 (Part 2: FAIR-Aware)

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...

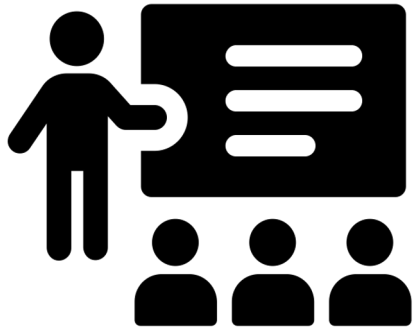
- Become familiar with the tool
- Learning outcomes/objectives
  - Essentially, what do you want your trainees to take away from the training?
  - Bloom's Taxonomy of Educational Objectives[1] can be helpful if you're uncertain
- Example datasets for target audiences
  - Domain specific?
  - Archive/repository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

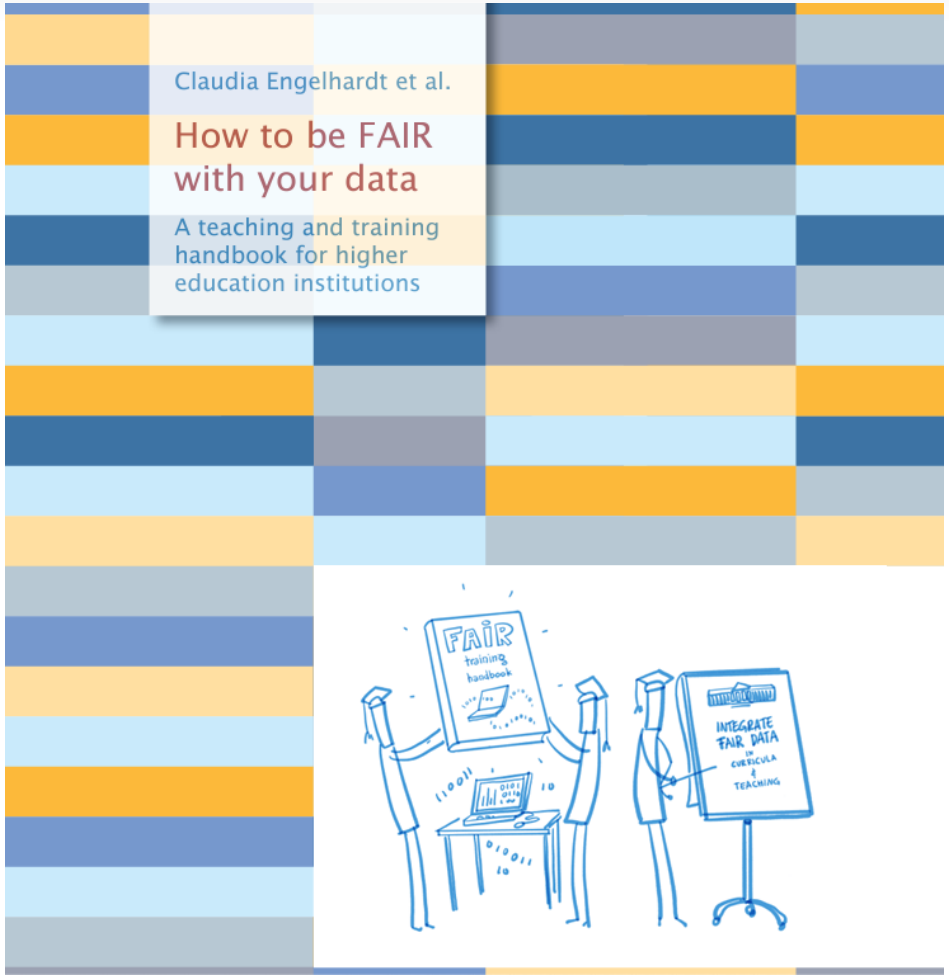
- Template of activity sheet (for your own (re-)use): https://doi.org/10.5281/zenodo.8089632
- Training exercise with worked examples: https://doi.org/10.5281/zenodo.8089501

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[1]

- 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)

Data Archiving and Networked Services

DANS

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 | @kfergy

Maaike Verburg | maaike.verburg@dans.knaw.nl | @MaaikeVerburg

FAIR-Aware | fair-aware@dans.knaw.nl | fairaware.dans.knaw.nl

DANS | info@dans.knaw.nl | @DANS_knaw_nwo | 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?

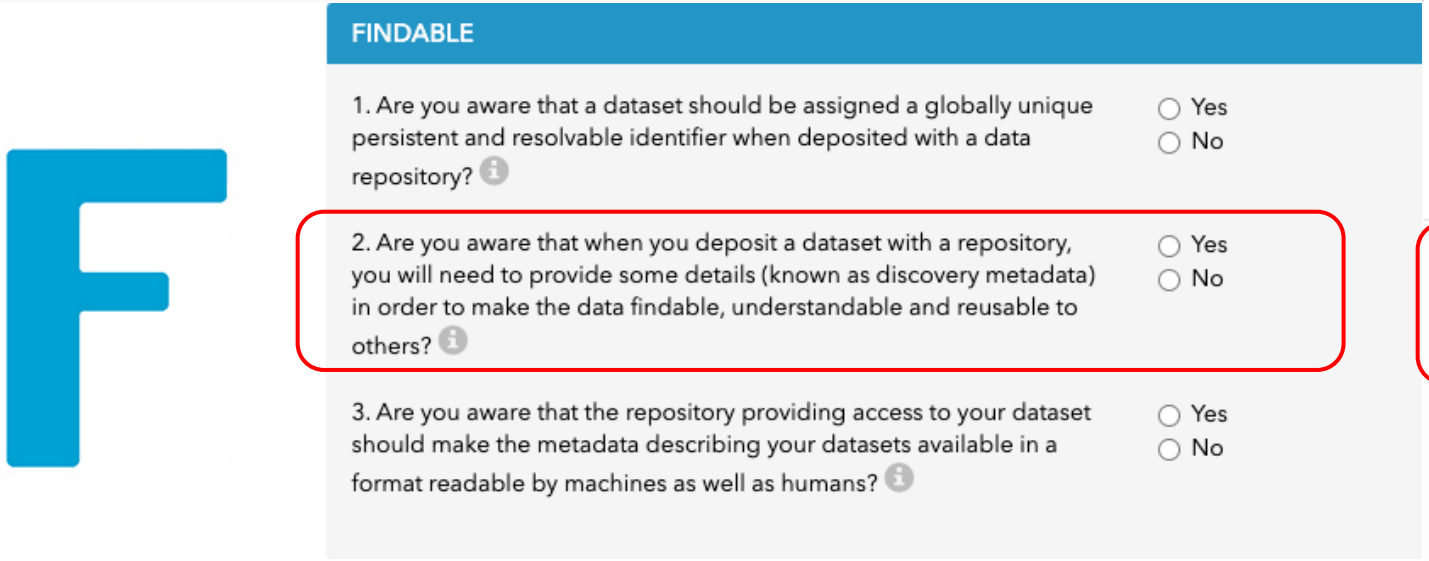
File formats refer to methods for encoding digital information. If there is no open format available, then you may also use a proprietary format. For example, CSV for tabular data, NetCDF for multidimensional data and GeoTIFF for raster imagery.

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?



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? ○ 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

10. Are you aware that maintaining your dataset FAIR over time requires professional data curation and preservation?

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