



## Trust and FAIR Digital Objects: awareness

Maaike Verburg (DANS-KNAW)

maaike.verburg@dans.knaw.nl
@MaaikeVerburg

#IASSIST2022 | #FAIRAwareTool





### **FAIR Digital Objects**

#### DIGITAL OBJECT

#### Data, code and other research outputs

At its most basic level, data or code is a bitstream or binary sequence. For this to have meaning and to be FAIR, it needs to be represented in standard formats and be accompanied by Persistent Identifiers (PIDs), metadata and documentation. These layers of meaning enrich the object and enable reuse.

### IDENTIFIERS

#### Persistent and unique (PIDs)

Digital Objects should be assigned a unique and persistent identifier such as a DOI or URN. This enables stable links to the object and supports citation and reuse to be tracked. Identifiers should also be applied to other related concepts such as the data authors (ORCIDs), projects (RAIDs), funders and associated research resources (RRIDs).

### STANDARDS & CODE

#### Open, documented formats

Digital Objects should be represented in common and ideally open file formats. This enables others to reuse them as the format is in widespread use and software is available to read the files. Open and well-documented formats are easier to preserve. Data also need to be accompanied by the code use to process and analyse the data.

#### METADATA

#### **Contextual documentation**

In order for Digital Objects to be assessable and reusable, they should be accompanied by sufficient metadata and documentation. Basic metadata will enable data discovery, but much richer information and provenance is required to understand how, why, when and by whom the objects were created. To enable the broadest reuse, they should be accompanied by a plurality of relevant attributes and a clear and accessible usage license.

Image from: European Commission, Directorate-General for Research and Innovation, *Turning FAIR into reality: final report and action plan from the European Commission expert group on FAIR data*, Publications Office, 2018, <a href="https://data.europa.eu/doi/10.2777/1524">https://data.europa.eu/doi/10.2777/1524</a>

- The units which can actually be FAIR
- Can only exist in a FAIR ecosystem
- Creates the expectation of services that are FAIR-enabling



### **FAIR-enabling services**

- Services that influence the FAIRness of the objects in their holdings
  - Assigning persistent identifiers, facilitating metadata, connecting to related objects, allow objects to be found and reused by others, etc.

- Expectations:
  - Make objects FAIR
  - Keep objects FAIR over time

FAIR-Enabling trustworthy digital repositories

FAIR+Time: Preservation for a Designated Community SSHOC SEARSEAR



L'Hours, Hervé, Kleemola, Mari, von Stein, Ilona, van Horik, René, Herterich, Patricia, Davidson, Joy, Rouchon, Olivier, Mokrane, Mustapha, & Huber, Robert. (2021). FAIR + Time: Preservation for a Designated Community (01.00). Zenodo. <u>https://doi.org/10.5281/zenodo.4783115</u>



### **FAIR tools**

Expectations  $\rightarrow$  Monitoring and evaluation  $\rightarrow$  Metrics and tools

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





### Background



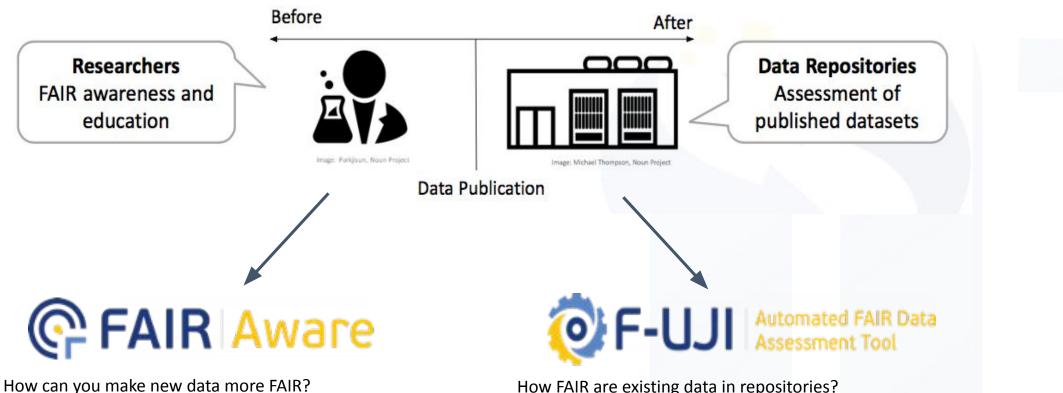


- March 2019 February 2022
- Goal: Practical solutions for the use of FAIR principles throughout the research data life cycle

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



### **FAIR assessment tools**

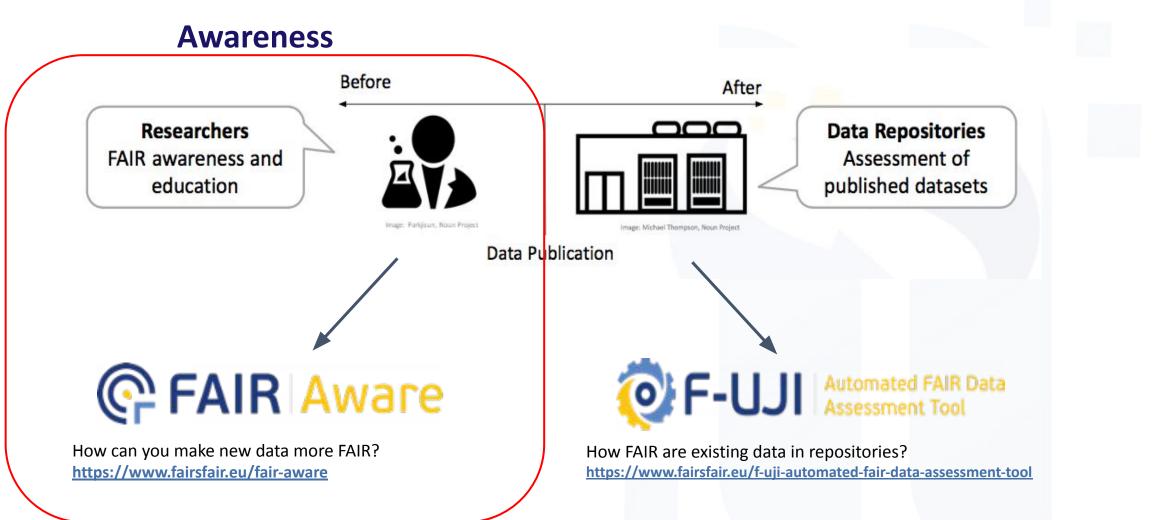


https://www.fairsfair.eu/fair-aware

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



### **FAIR** assessment tools





## **FAIR** Aware

#### Your first step towards your FAIR data(set)

- **10 FAIR practices**
- Simple questions
- **Extensive guidance texts**
- Aim: help researchers and data professionals create more FAIR data before deposit

FAIR questions 👁 👁		unique persistent and resolvable identifier when deposited with a data repository?	
FINDABLE		AIR qu INDARI What does this mean?	
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	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 (DO)(27, Handle(27, and Archival Resource Key (ARK)(27.	
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?	<ul><li>○ Yes</li><li>○ No</li></ul>	Are you ata(set) I If your data(set) or metadata does not have a PID, you run the risk of " <b>link rot</b> " (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, ata(set) or others will not be able to find or reuse your data(set) or metadata in the long-term.	
3. Are you aware that the data repository providing access to your	⊖ Yes	a formal How to do this?	
data(set) should make the metadata describing your data(set) available in a format readable by machines as well as humans?	⊖ No	CCESSI 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.	
ACCESSIBLE		There are many different types of PIDs, each with their own advantages, disadvantages, hinch the 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	
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?	<ul><li>○ Yes</li><li>○ No</li></ul>	Are you have to choose the PID type yourself, you can visit the Knowledge Huble <sup>2</sup> 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 <sup>2</sup> 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 <sup>2</sup> or FAIRsharing (see related databases) <sup>2</sup> .	
5. Are you aware that metadata should remain available over time, even if the data(set) is no longer accessible?	<ul><li>○ Yes</li><li>○ No</li></ul>	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.	
		Are you Want to know more?	
		se contro 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.	
INTEROPERABLE		Persistent identifiers may point to a data file, a web service response that contains data	
6. Are you aware that the metadata describing your data(set) should use controlled vocabularies? 🕄	<ul><li>○ Yes</li><li>○ No</li></ul>	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.	

. Are you aware that a data(set) should be assigned a globally



### **FAIR-Aware**

- Assesses knowledge, not objects
- Manual self-assessment



Klaus Riede @KlausRiede

I just used **#FAIRAwareTool** to assess and increase my knowledge on the **#FAIR** data principles! Try it out for yourself here: fairaware.dans.knaw.nl

...

@DANS\_knaw\_nwo | @FAIRsFAIR\_eu | #FAIRAware Tweet vertalen

1:21 p.m. · 10 mrt. 2022 · Twitter Web App

- When implemented leads to more FAIR data being deposited
- Details responsibilities of different stakeholders and what expectations to have of repositories

 $\rightarrow$  Choose trustworthy repositories that show FAIR-enabling qualities



# **FAIR** Aware

Your first step towards your FAIR data(set)



A French translation of the tool is available. Other translations are being worked on to bring FAIR closer to different communities.





Users can ask questions or come together to discuss issues and ideas on the dedicated space at the FAIR Data Forum

The trainer functionality allows others to quickly access the results of a group of people. FAIR-Aware can be employed as a collective learning tool.



In FAIR-IMPACT, the tool will be further developed and focus will be put on bringing the tool closer to disciplinary-specific contexts

https://doranum.fr/appli-fair-aware-pleine-page-vf/







@fairimpact\_eu /company/fair-impact-eu-project

