FAIR principles

Obrad Vučkovac
University of Belgrade
Vinča Institute of Nuclear Sciences – Library

EIFL Open Science bootcamp

Training challenges

- Complex topic
- Terminology
  - metadata standards, communication protocols, ontologies, etc.
- Misconseptions
- Steep learning curve
- If it's not mandatory, can be difficult to persuade researchers to adhere to FAIR principles
- Lack of incentives for extra work

Photo by Luke Pamer on Unsplash

FAIR principles

The FAIR principles describe how research outputs should be organised so they can be more easily accessed, understood, exchanged and reused, by both humans and machines.
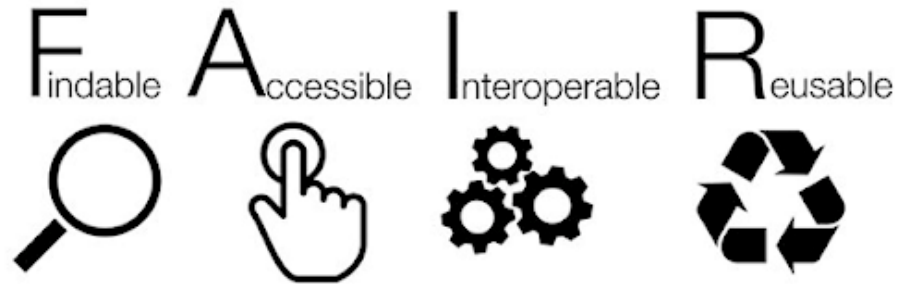
Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

Funder requirements

- ## Horizon Europe
  - *"responsible management of research data in line with the FAIR principles of 'Findability', 'Accessibility', 'Interoperability' and 'Reusability', notably through the generalised use of data management plans, and open access to research data under the principle 'as open as possible, as closed as necessary', under the conditions required by the grant agreement";*

- ## Bill & Melinda Gates Foundation
  - *"Grantees are encouraged to adhere to the FAIR principles to improve the findability, accessibility, interoperability, and reuse of digital assets".*
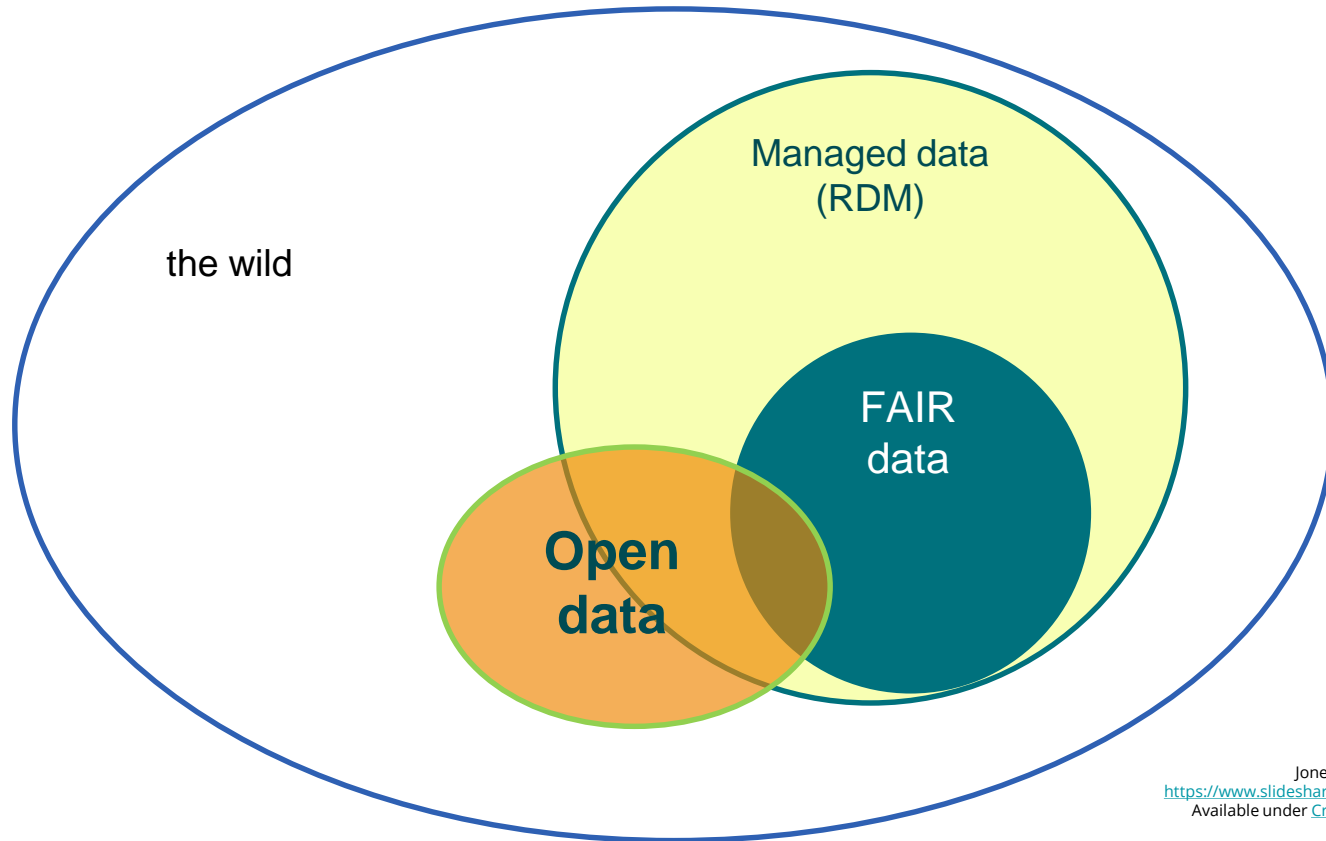
Why are FAIR principles needed?

"The increasing availability of online resources means that data need to be created with longevity in mind. Providing other researchers with access to your data facilitates knowledge discovery and improves research transparency."
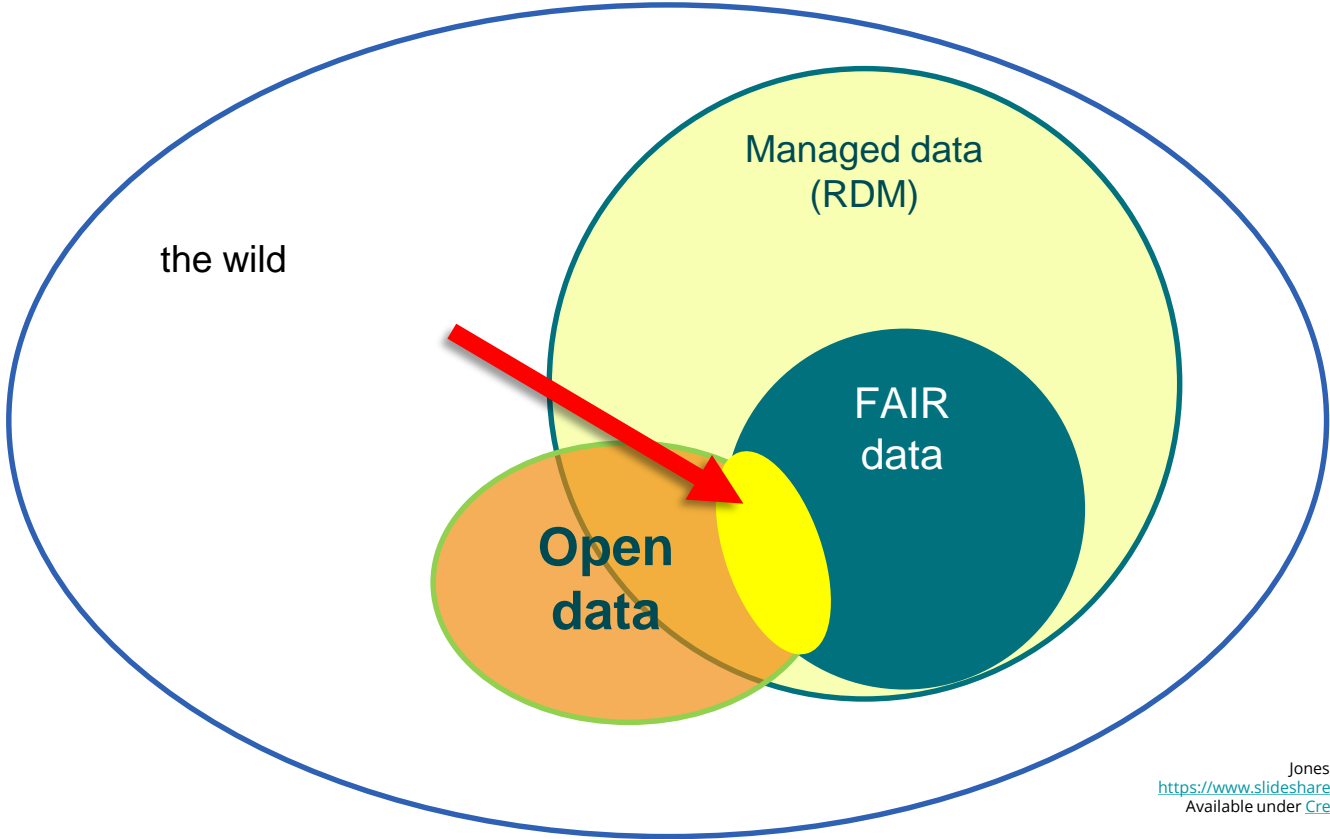
How to make your data FAIR, OpenAIRE

Humans increasingly rely on computational support from machines. FAIR data enable to find, access, interoperate, and reuse of data with no or minimal human intervention.
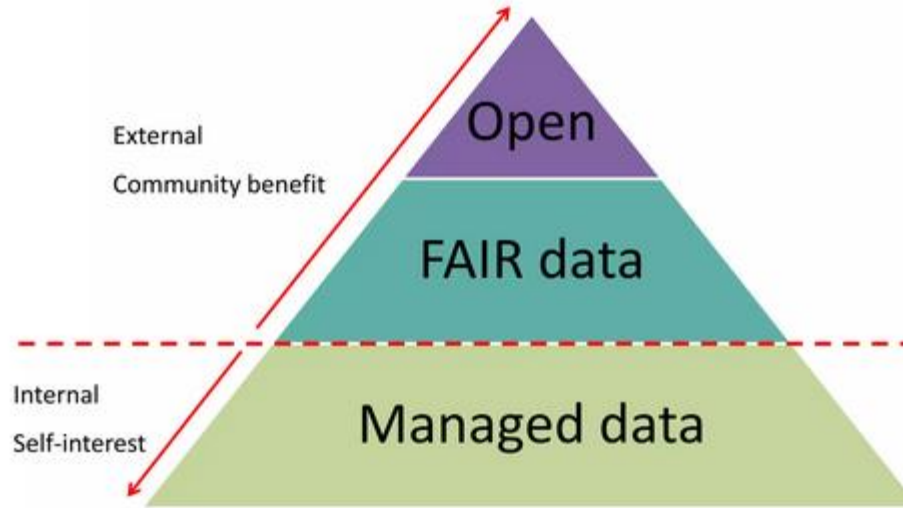
RDM, FAIR & Open data

RDM & FAIR & Open Data

the wild

Managed data
(RDM)

FAIR
data

**Open
data**

RDM & FAIR & Open Data

RDM & FAIR & Open Data

External
Community benefit

Internal
Self-interest

Open

FAIR data

Managed data

Training concept: Explaining the FAIR principles

Findable

The first step in (re)using data is to find them.

F1. (Meta)data are assigned a globally unique and persistent identifier

F2. Data are described with rich metadata (defined by R1 below)

F3. Metadata clearly and explicitly include the identifier of the data they describe

F4. (Meta)data are registered or indexed in a searchable resource
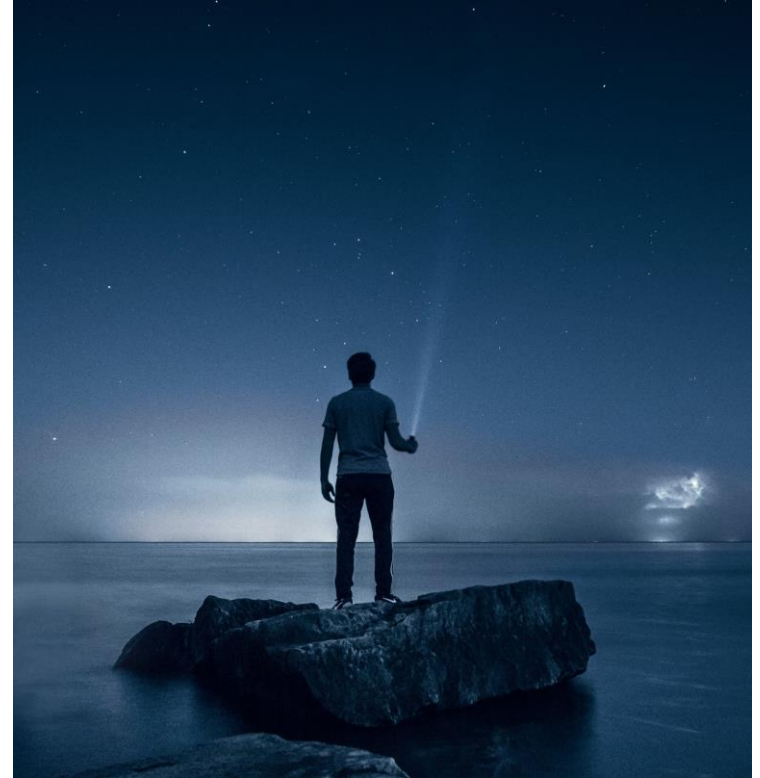

Photo by Warren Wong on Unsplash

Accessable

Users need to know how data can be accessed, possibly including authentication and authorisation.

A1. (Meta)data are retrievable by their identifier using a standardised communications protocol

A1.1 The protocol is open, free, and universally implementable

A1.2 The protocol allows for an authentication and authorisation procedure, where necessary

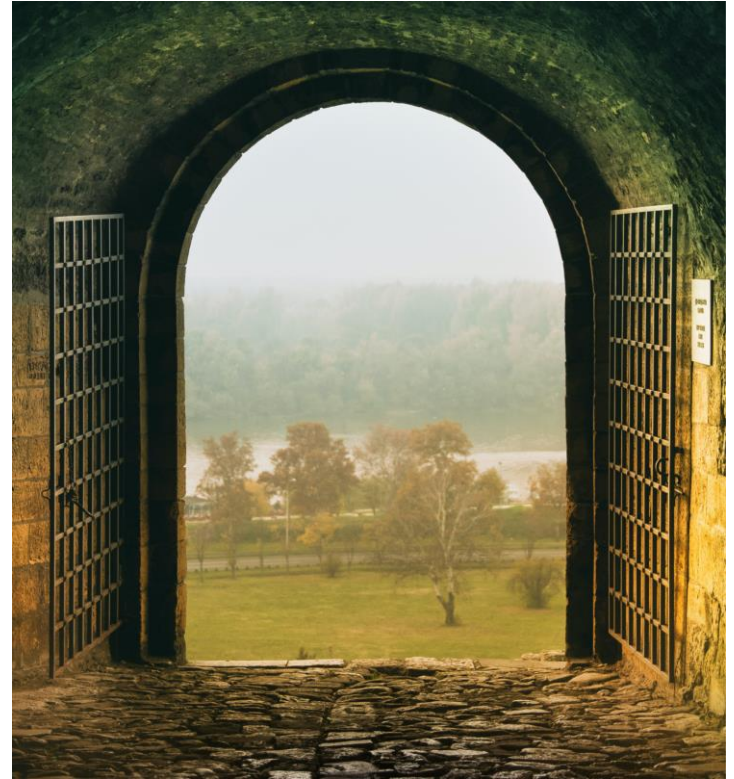A2. Metadata are accessible, even when the data are no longer available

Interoperable

The data usually need to be integrated with other data, and to be able to interoperate with applications or workflows for analysis, storage, and processing.

I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.

I2. (Meta)data use vocabularies that follow FAIR principles

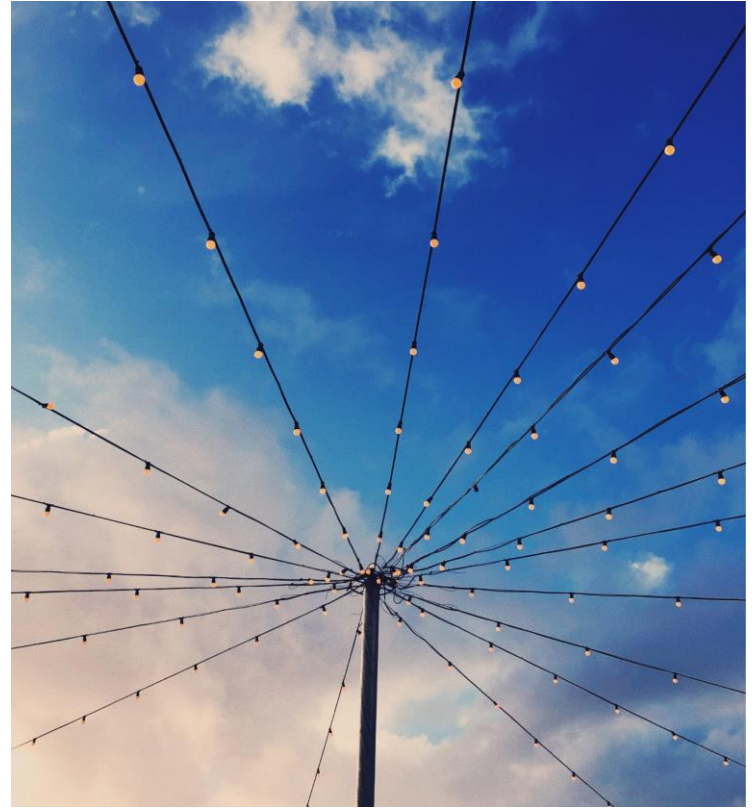I3. (Meta)data include qualified references to other (meta)data

Photo by Bruno Figueiredo on Unsplash

Reusable

To reuse data, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

R1. (Meta)data are richly described with a plurality of accurate and relevant attributes

R1.1. (Meta)data are released with a clear and accessible data usage license

R1.2. (Meta)data are associated with detailed provenance

R1.3. (Meta)data meet domain-relevant community standards

Training concept: FAIR assessment

Ask researchers to bring their own data, or use existing.

FAIR assessment tools:

F-UJI by FAIRsFAIR

ARDC FAIR Assessment tool

SATYFID by DANS

(Self-Assessment Tool to Improve the FAIRness of Your Dataset)

How FAIR are your data? EUDAT (Zenodo)

RDA FAIR data maturity model

The RDA FAIR Data Maturity Model Working Group develops as an RDA Recommendation a common set of core assessment criteria for FAIRness and a generic and expandable self-assessment model for measuring the maturity level of a dataset. The aim is not to develop yet another FAIR assessment approach but to build on existing initiatives, looking at common elements and allowing the group to identify core elements for the evaluation of FAIRness.

How FAIR are your data?

Findable
It should be possible for others to discover your data. Rich metadata should be available online in a searchable resource, and the data should be assigned a persistent identifier.

☐ A persistent identifier is assigned to your data
☐ There are rich metadata, describing your data
☐ The metadata are online in a searchable resource e.g. a catalogue or data repository
☐ The metadata record specifies the persistent identifier

Accessible
It should be possible for humans and machines to gain access to your data, under specific conditions or restrictions where appropriate. FAIR does not mean that data need to be open! There should be metadata, even if the data aren't accessible.

☐ Following the persistent ID will take you to the data or associated metadata
☐ The protocol by which data can be retrieved follows recognised standards e.g. http
☐ The access procedure includes authentication and authorisation steps, if necessary
☐ Metadata are accessible, wherever possible, even if the data aren't

Interoperable
Data and metadata should conform to recognised formats and standards to allow them to be combined and exchanged.

☐ Data is provided in commonly understood and preferably open formats
☐ The metadata provided follows relevant standards
☐ Controlled vocabularies, keywords, thesauri or ontologies are used where possible
☐ Qualified references and links are provided to other related data

Reusable
Lots of documentation is needed to support data interpretation and reuse. The data should conform to community norms and be clearly licensed so others know what kinds of reuse are permitted.

☐ The data are accurate and well described with many relevant attributes
☐ The data have a clear and accessible data usage license
☐ It is clear how, why and by whom the data have been created and processed
☐ The data and metadata meet relevant domain standards

Findable Accessible Interoperable Reusable

'How FAIR are your data?' checklist, CC-BY by Sarah Jones & Marjan Grootveld, EUDAT. Image CC-BY-SA by SangyaPundir

FAIR – data, metadata, infrastructures, and more

Barker, M., Chue Hong, N.P., Katz, D.S. *et al.* Introducing the FAIR Principles for research software. *Sci Data* **9**, 622 (2022). https://doi.org/10.1038/s41597-022-01710-x

Garcia, L., Batut, B., Burke, M. L., Kuzak, M., Psomopoulos, F., Arcila, R., Attwood, T. K., Beard, N., Carvalho-Silva, D., Dimopoulos, A. C., Del Angel, V. D., Dumontier, M., Gurwitz, K. T., Krause, R., McQuilton, P., Le Pera, L., Morgan, S. L., Rauste, P., Via, A., Kahlem, P., … Palagi, P. M. (2020). Ten simple rules for making training materials FAIR. *PLoS computational biology*, *16*(5), e1007854. https://doi.org/10.1371/journal.pcbi.1007854

RDA IG – FAIR Principles for Research Hardware
https://www.rd-alliance.org/groups/fair-principles-research-hardware

FAIR *upgrades*

"[...] to make data FAIR whilst preserving them over time requires trustworthy digital repositories (TDRs) with sustainable governance and organizational frameworks, reliable infrastructure, and comprehensive policies supporting community-agreed practices."

Lin, D., Crabtree, J., Dillo, I. *et al.* The TRUST Principles for digital repositories. *Sci Data* **7**, 144 (2020).
https://doi.org/10.1038/s41597-020-0486-7

Transparency, **R**esponsibility, **U**ser focus, **S**ustainability, **T**echnology

https://www.rd-alliance.org/rda-community-effort-trust-principles-digital-repositories

FAIR *upgrades*

CARE principles

Collective benefit; **A**uthority to control; **R**esponsibility; **E**thics

CARE principles complement the existing FAIR principles encouraging open and other data movements to consider both (Indigenous) people and purpose in their advocacy and pursuits.

more on YT: https://youtu.be/309QIZt9H74

Twitter hashtag: #BeFAIRandCARE

Training tips

- FAIR is very important! It is the way to preserve data and make it findable and accessable for future reuse, and to align with funders requirements

- Explain terminology with examples

- Use real FAIR data in excercises and demonstrations

- Demonstrate FAIRness assessment
  - ask researchers to assess their data

- Present use cases

Questions?

Obrad Vučkovac

University of Belgrade

Vinča Institute of Nuclear Sciences - Library

ORCID: 0000-0001-5616-2680