National Initiatives for Open Science in Europe

How to make your data FAIR?

- Practical Steps and Recommendations for implementing FAIR principles in your data set-

National Capacity Building Training for North Macedonia

Vladimir Trajkovik, FINKI (FCSE), UKIM

15.10.2021

What is FAIR?

'Findable' as discoverable with metadata, identifiable and locatable by means of a standard identification mechanism

'Accessible' as always available and obtainable; even if the data is restricted, the metadata is open

'Interoperable' as both syntactically parse-able and semantically understandable, allowing data exchange and reuse between researchers, institutions, organizations or countries

'Reusable' as sufficiently described and shared with the least restrictive licenses, allowing the widest reuse possible and the least cumbersome integration with other data sources

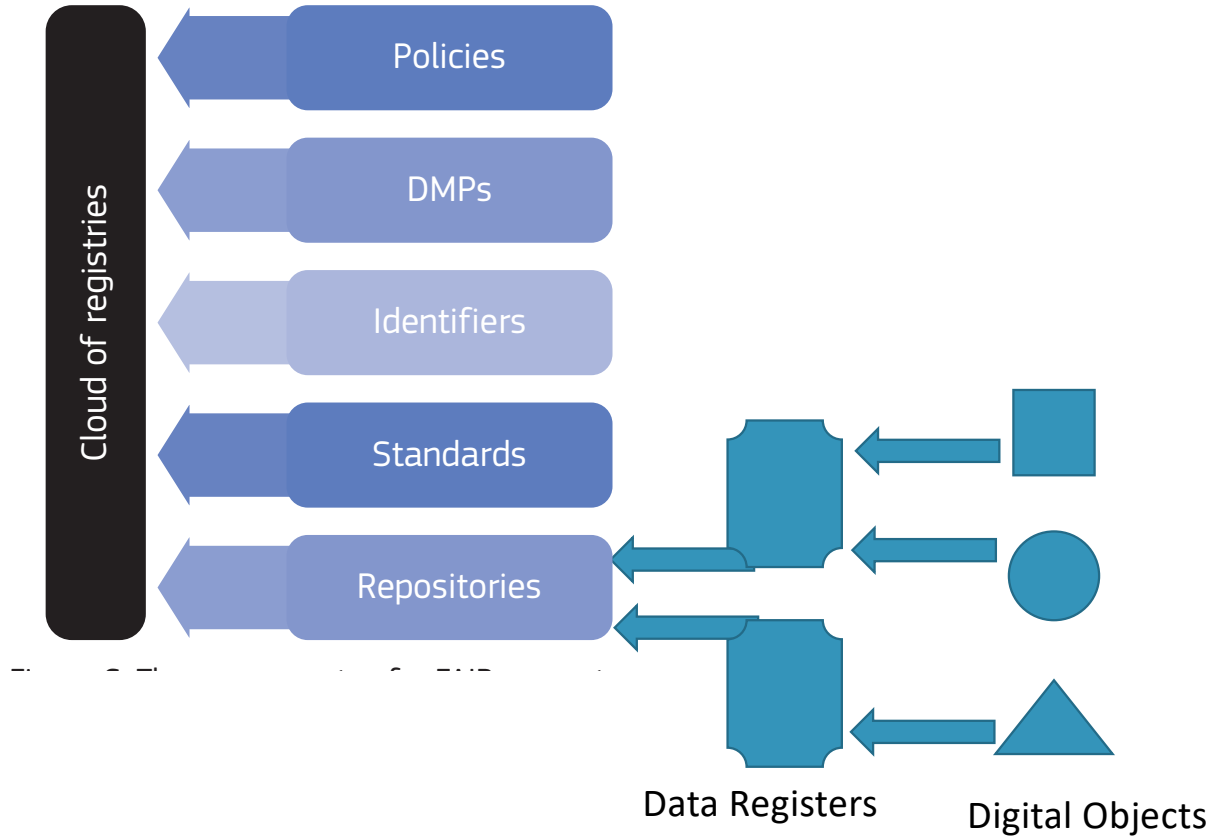
FAIR data is about principles not standards!

Providing other researchers with access to your data facilitates knowledge discovery and improves research transparency.

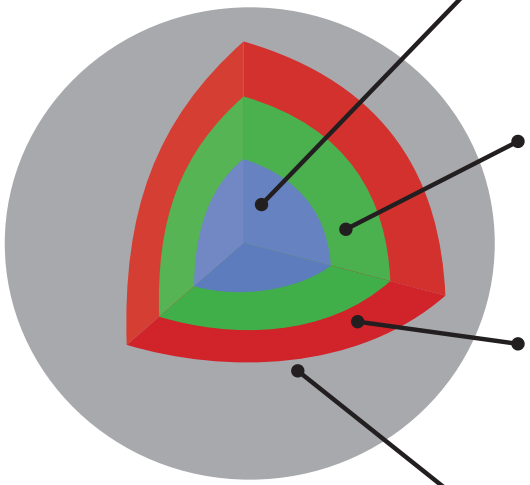
Data can be FAIR but not open.

Open data may not be FAIR.

FAIR Digital Objects & FAIR Eco-System



FAIR Digital objects (DO)



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

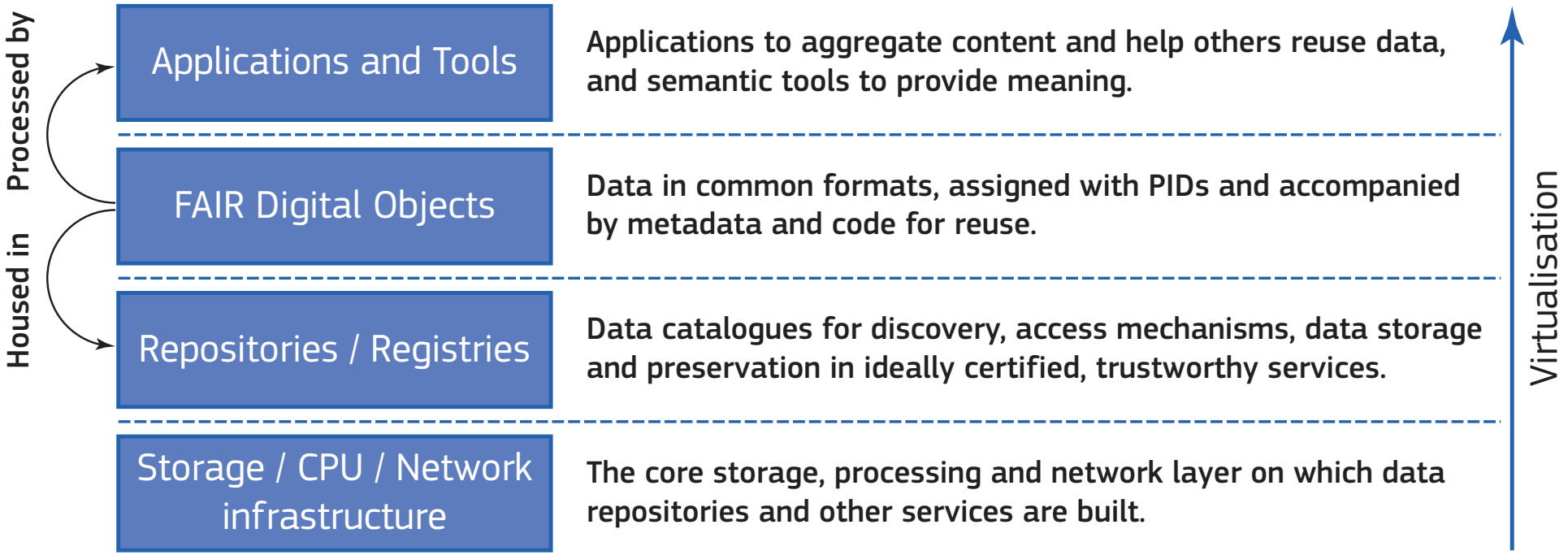
Digital Objects should be represented in common and ideally open file formats. This enables others to reuse them as the format is in widespread use and software is available to read the files. Open and well-documented formats are easier to preserve. Data also need to be accompanied by the code use to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

METADATA

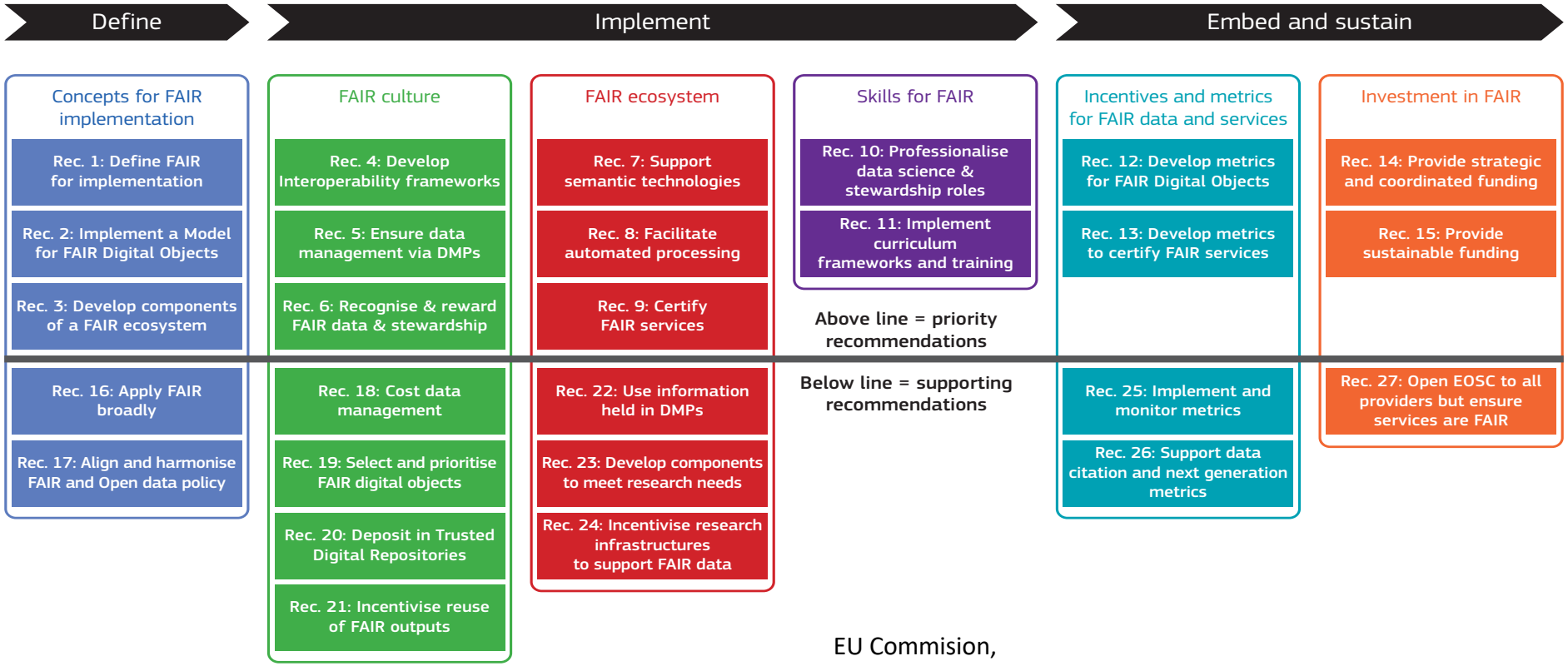
Contextual documentation

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.

The placement of FAIR DO in the technical infrastructure layers



FAIR Action Plan



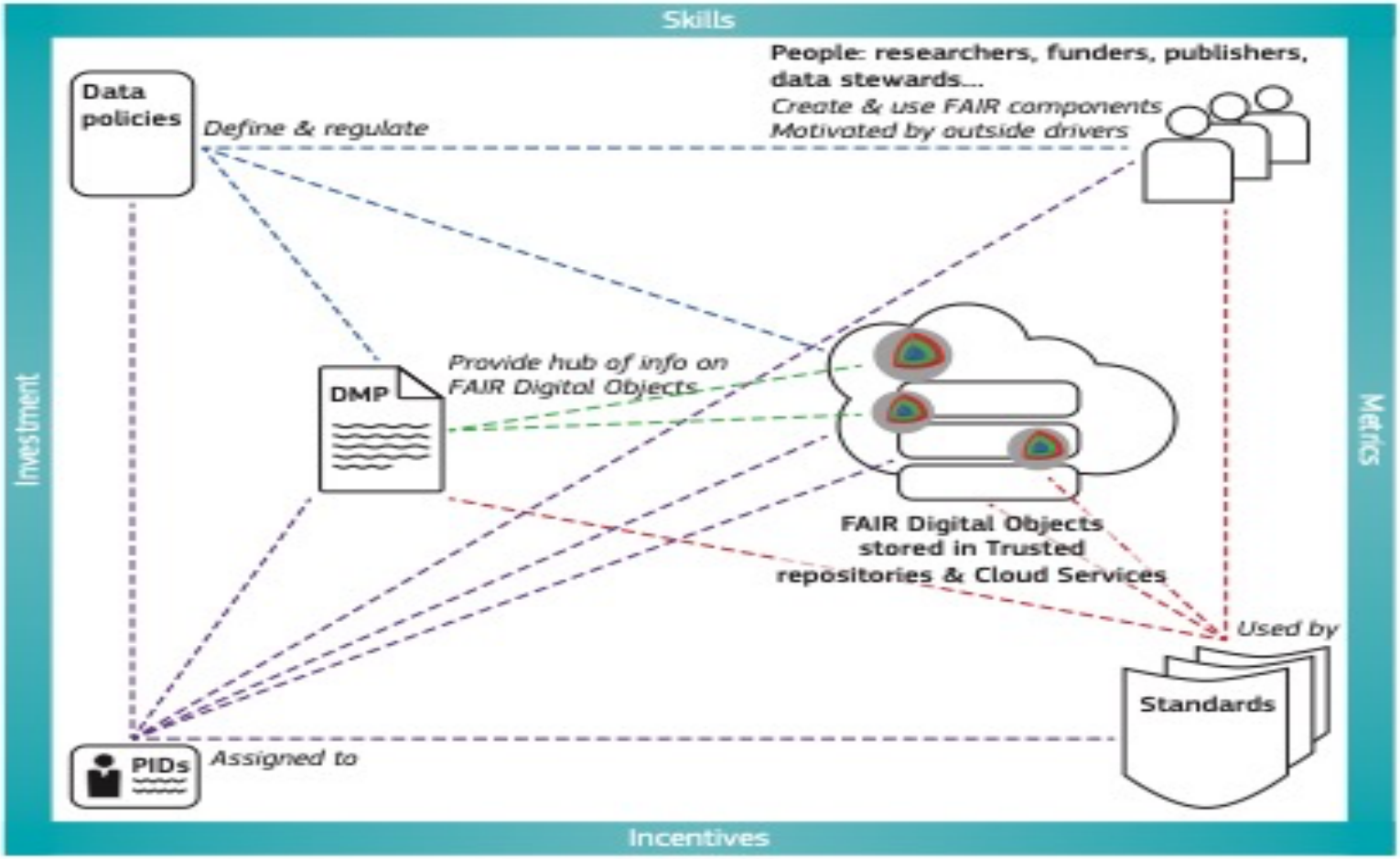
Index to FAIR Action Plan recommendations

EU Commision,
Final Report and Action Plan from the European Commission Expert Group on FAIR Data, 2018

Data Management Plan (DMP)

- DMPs hold valuable information on the data and related outputs, which should be structured in a machine- actionable way to enhance reuse. Investment should be made in DMP standards and tools that adopt common standards and support ‘active’ DMPs to enable information exchange across the FAIR data ecosystem.
- Initial versions of a DMP should be produced early in the research workflow, providing an opportunity to reflect on decisions that will affect the FAIRness of the data. While they may seem an administrative burden at first, the process of creating - and updating - DMPs can provide important insights and lessons on how to gather, curate and disseminate data, building a common understanding across the project from an early stage and reducing administrative burdens over the project lifecycle.
- In order for data to be fully understood, reproducible and reusable to the greatest extent possible, associated outputs such as software, workflows and protocols should also be shared. DMPs should be applied broadly to the full range of outputs needed for FAIR.

FAIR Eco-System in action



Findable

Has a persistent identifier (PID); DOI is example of PID.

When depositing your data in a repository, make sure you select a repository that assigns a persistent identifier (for example Zenodo).

Has rich metadata.

Follow standard metadata schemes, general ones such as Dublin Core, or discipline specific (Consult the DCC metadata directory, the RDA Metadata Directory and a portal of data standards at FAIRsharing) .

Is searchable and discoverable online.

Check the how discoverable (indexed) is the repository you plan to use.

Accessible - as open as possible, as closed as necessary

if access is allowed, data should be retrievable without the need for specialized protocols

even if the full content is not made openly available, the data must be as findable as possible

Findable + Accessible = Data Repository

Stores the data safely

Make sure the data is findable

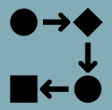
Describes the data appropriately (metadata)

Adds license information

You can deposit data to a general repository (e.g. Zenodo, Harvard Dataverse) or a subject-specific repository (e.g. Dryad).

What about your discipline? Search www.re3data.org for more suitable data repositories. A demonstration of searching for research data repositories using the re3data directory will be given after this presentation.

Interoperable = Common (open) formats and standards + Controlled vocabularies



So it can be integrated with other data, applications and workflows.

Try NOT to create data with proprietary software

Try to use community agreed schemas, keywords, thesauri or ontologies where possible

Reusable = well-documented + clear license information

README file to help ensure that your data can be correctly interpreted and reanalyzed:

- for each filename, a short description of what data it includes, optionally describing the relationship to the tables, figures, or sections within the accompanying publication;
- for tabular data: definitions of column headings and row labels; data codes (including missing data); and measurement units;
- any data processing steps, especially if not described in the publication, that may affect interpretation of results;
- a description of what associated datasets are stored elsewhere, if applicable;
- whom to contact with questions.

Reusable = well-documented + clear licence information

Data should have a **clear license** to govern the terms of its reuse.

The Open Access guidelines under Horizon 2020 recommend CC-0 or CC-BY as a straightforward and effective way to make it possible for others to mine, exploit and reproduce the data.

EUDAT provides a wizard to help you choose an appropriate license

So, how FAIR are your data?

Jones, S. & Grootveld, M. (2017, November).

How FAIR are your data? (CC-BY)

Zenodo. *http://doi.org/10.5281/zenodo.1065991*

How FAIR are your data? (1/2)

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

How FAIR are your data? (2/2)

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

OPEN AIRE Resources

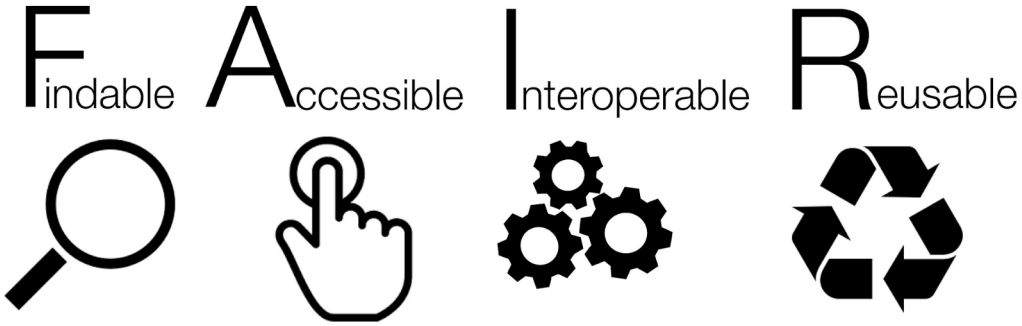
https://www.openaire.eu/how-to-make-your-data-fair

https://www.openaire.eu/how-to-create-a-data-management-plan

https://www.openaire.eu/data-formats-preservation-guide

https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs

Thank you!



https://ni4os.eu

@NI4OS_eu @NI4OS

Source: https://en.wikipedia.org/wiki/FAIR_data#/media/File:FAIR_data_principles.jpg