Action title:

FAIR NEtwork of micrometeorological measurements

Action No.: **CA20108**

Acronym: **FAIRNESS**

Action chair: Branislava Lalić

WG No: **1**

WG leader: Stevan Savić

Deliverable writers: Lalić, B., Koči, I.

Deliverable: **D1.4 – Guidelines for FAIR principles and methodological improvements (M39)**

Date: **4.2.2025.**

Revision 1: **29.8.2025**

FAIR Data Beginner's Workbook

This workbook provides step-by-step instructions and beginner-friendly checklists to help you implement the FAIR data principles (Findable, Accessible, Interoperable, Reusable). Each section includes practical guidance and a compliance checklist so you can verify your progress.

Remember that data are not just numbers — they can also be figures, photos, schemes, algorithms, protocols, videos, audio recordings, or any other form of information that can be stored and shared.

Before you start, we suggest deciding where your data will be stored, whether FAIR/open or not, and familiarising yourself with that repository. If you choose Zenodo, here is the short instruction on how to use it (link).

Before we begin with the step-by-step guidance, here's a quick reminder of what **FAIR** means.
FAIR stands for **Findable, Accessible, Interoperable, and Reusable** — four key principles that ensure data can be easily discovered, understood, and used by both people and machines. The checklist below shows the essential elements for each principle, helping you assess how FAIR your data currently are and where improvements can be made.

Put simply: **FAIR data is data that anyone can find and understand**, but access to and use of the data always follow the rules set by the data owner. These rules may allow open use, place certain restrictions, or limit access entirely — FAIR ensures that, whatever the policy, the data can still be discovered and interpreted correctly.

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

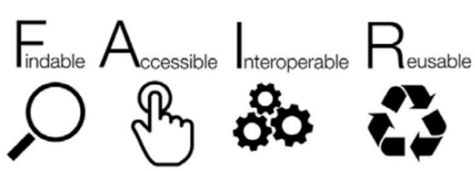
Data and metadata should conform to recognised formats and standards to allow them to be combined and exchanged.

- ☐ 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



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

1 – Make Your Data Findable

Step 1 – Assign a persistent **identifier**[1]

- Use a public **repository**[2] that provides a DOI (e.g., Zenodo, Figshare, Dryad, PANGAEA) or a Handle.
- If a DOI is not possible, use a stable URL from your institutional repository or project web page.

Step 2 – Provide clear metadata

- List metadata that best describe your dataset (title, description, creator, keywords, license, date, geographic coverage, methodology).
- Use standard abbreviations and variable names from your field (e.g., Tmax = maximum daily temperature).
- Save metadata in a simple, readable format (CSV, TXT, JSON, XML).

Step 3 – Make your data and metadata searchable

- Upload them to a searchable repository or catalogue.
- Complete repository fields such as title, keywords, and abstract.

Step 4 – Ensure discoverability on the web

- Make the metadata record public and link it to the dataset.
- Include contact details for questions.
- Test whether your record appears in repository search and Google Dataset Search.

[1]**Persistent identifier** is like a **permanent address for your data** on the internet (such as a **DOI – Digital Object Identifier**) that always takes you to the same dataset or resource. Even if the data moves to a different server or website, the persistent identifier will still work — it's updated behind the scenes so you don't have to hunt for the new location.

[2]Repository is like a library for your data, which stores, organises and keeps it long-term.

FAIR Compliance Checklist – Findable

☐ Chosen a searchable repository or catalogue
☐ Dataset has a persistent identifier (DOI, Handle, or stable URL)
☐ Metadata includes title, description, creator(s), keywords, license, and dates
☐ Metadata uses standard abbreviations/units relevant to my field
☐ Metadata is in a readable format (CSV, TXT, JSON, XML, etc.)
☐ Dataset and/or metadata are uploaded to the chosen repository
☐ Repository fields (title, abstract, keywords) are filled in completely
☐ Metadata record is public
☐ Contact details are included
☐ I can find my record via repository search and Google Dataset Search

Step 2 – Make Your Data Accessible

Step 1 – Store your data in a stable, long-term repository

- Choose a repository committed to long-term preservation.
- Avoid temporary storage like personal drives or short-term cloud links.

Step 2 – Define access conditions

- Decide if your data will be **open**[3], **restricted**[4], or **embargoed**[5].
- Clearly state your **access policy**[6] in the metadata.

Step 3 – Use standard access protocols

- Ensure data can be retrieved via standard, widely used protocols (HTTP, FTP, OAI-PMH).
- Avoid custom or obscure access methods.

Step 4 – Keep metadata open

- Make metadata freely accessible, even if the data itself is restricted.
- **Share metadata**[7] via repositories, catalogues, or project websites.

[3]**Open access means that a**nyone can download, use, and share your data

[4]Restricted access means that only approved users can access data.

[5]**Embargoed access means that t**he data is stored and visible in the repository but can't be downloaded until the embargo period ends.

[6]Access policy is your **rulebook for who can see and use your data, and under what conditions**.

[7]Metadata is your **data's "shop window"** — it tells people what's inside before they open the box. Sharing metadata makes your data and your work visible.

FAIR Compliance Checklist – Accessible

☐ Data is stored in a stable, long-term repository

☐ Repository is known to preserve data for many years

☐ Chosen and documented an access policy (open, restricted, embargoed)

☐ Access policy is clearly stated in the metadata

☐ Data can be retrieved using standard protocols (HTTP, FTP, OAI-PMH)

☐ No special or proprietary software is required to access the data

☐ Metadata is freely accessible even if the data is restricted

☐ Metadata is published in at least one searchable platform

Step 3 – Make Your Data Interoperable

Step 1 – Use common, open file formats

- For numbers: CSV, TXT, XLSX
- For images: TIFF, PNG, JPEG
- For geospatial data: GeoTIFF, NetCDF, Shapefile

Step 2 – Follow metadata **standards**[8]

- Select metadata fields based on existing community or domain guidelines.
- Include methodology, equipment, design details, and relevant contextual information.

Step 3 – Use standard vocabularies and units

- Follow established standards for symbols, units, and parameter names. For example, for maximum air temperature use symbol T_{max}, unit – K, and name it as it is – "maximum air temperature" (not maximum temp., maximal temp. etc.
- Use controlled **vocabularies**[9], **thesauri**[10], or **ontologies**[11] if available.

Step 4 – Link to related resources

- Provide links or references to related datasets, publications, or documentation not contained in your metadata file.

[8]An example of metadata file content and detail explanation you could find in "Metadata file– how to make it?" (link)

[9]A controlled vocabulary is simply a fixed list of standard words everyone agrees to use. This avoids confusion from using different names for the same thing. For example, instead of one dataset saying *"temperature"* and another saying *"temp_air"*, both would use the agreed term *"air temperature"*.

[10]A thesaurus is not only lists words, but also shows how they are related. For example, the word *"precipitation"* could have narrower terms like *"rainfall"*, *"snowfall"*, or *"hail"*.

[11]An ontology is a kind of map that shows not just the words, but how the concepts behind them connect to each other. For example, an ontology could explain that *"relative humidity"* depends on both *"air temperature"* and *"specific humidity"*. Ontologies make it possible for computers to understand and combine weather data with other types of information.

FAIR Compliance Checklist – Interoperable

☐ Data is in a common, open file format appropriate for its type

☐ File formats are widely supported and machine-readable

☐ Metadata fields follow relevant community or domain guidelines

☐ Methodology, equipment, and context are clearly described

☐ Units, symbols, and variable names follow accepted standards

☐ Controlled vocabularies, thesauri, or ontologies are used where available

☐ Metadata contains links or references to related datasets or publications

☐ Connections to external resources are clear and functional

Step 4 – Make Your Data Reusable

Step 1 – Check data quality and completeness

• Compare your data against reference datasets to verify accuracy.

• Ensure all relevant variables and metadata fields are present.

Step 2 – Choose and state a **license**[12]

• Pick an appropriate data usage license (CC0, CC-BY).

• Include license details in your metadata.

Step 3 – Document creation and processing

• Clearly state how, why, and by whom the data was created and processed.

• If too detailed for metadata, create a separate documentation file and link to it.

Step 4 – Follow domain standards

• Format and structure data so it can be used by both humans and machines.

• Structure files to include as much relevant information as possible.

[12] A license tells you **who owns the data** and **how you may use it. More about licenses and how to choose it you can find here https://creativecommons.org/share-your-work/cclicenses/**

FAIR Compliance Checklist – Reusable

☐ Data has been checked for accuracy against reference datasets

☐ All relevant variables and metadata fields are complete

☐ Chosen an appropriate usage license (e.g., CC0, CC-BY)

☐ License is included in the metadata

☐ How, why, and by whom the data was created is documented

☐ Processing steps are clearly explained, with extra files linked if needed

☐ Data is formatted and structured for both human and machine use

☐ Data tables include relevant details (e.g., coordinates, units) where possible