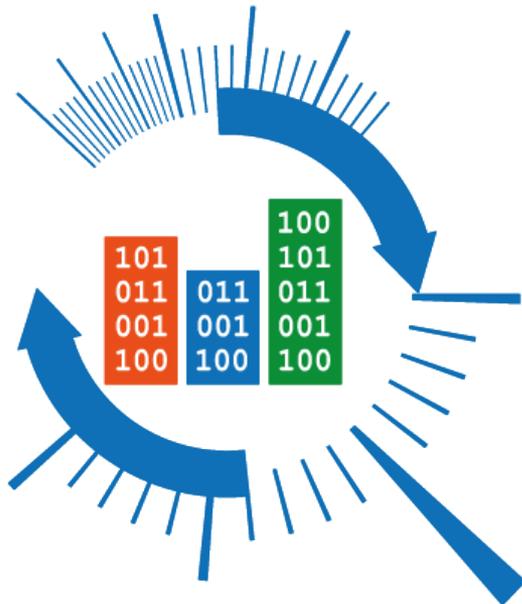


# Web Tutorial 1: Research data management general/motivation

Thomas Wiedenhoefer

Physikalisch-Technische Bundesanstalt (PTB)



## Research data management in European metrology





# Workshop Program and Schedule

<u>Time</u>	<u>Contents</u>
<b>2025-06-03, 14:00-15:30 CEST</b>	<b>"RDM in EPM projects" Module 1: Motivation, obligations and support</b>
5 min	Greeting
30+10 min	Plenary: Motivation and general information
10+10 min	<b>Breakout session 1:</b> RDM in JRP/CSP proposals
10+10 min	<b>Breakout session 2:</b> RDM in starting/running projects
10+10 min	Plenary: Obligations in EPM projects (EPM guides); Support
10 min	Questions and discussion
<b>2025-06-12, 14:00-15:30 CEST</b>	<b>"RDM in EPM projects" Module 2: Planning and data publication</b>
5 min	Greeting Plenary: Obligations in EPM projects
20+10 min	<b>Breakout session 1:</b> Describing RDM in JRP/CSP proposals
20+10 min	<b>Breakout session 2:</b> Preparing a data management plan + RDMO
20+10 min	Plenary: Data publication + sensitive data + versioning
10 min	Questions and discussion
<b>2025-06-20, 14:00-15:30 CEST</b>	<b>"RDM in EPM projects" Module 3: Metadata and good practices</b>
5 min	Greeting
20+10 min	Metadata: how to select them and how to record them
20+10 min	<b>Breakout session:</b> Exploration of the SI dig.fr. + API query
20+10 min	<b>Breakout session:</b> Case study about lab notebooks
10 min	Questions and discussion



# What are Research Data?

---

- Research data are factual records used as primary sources for scientific research.
- They can result from observation, experiment, simulation, survey, or derivation.
- Data exist in many forms: numerical values, images, texts, software code, audio, etc.
- Research data can be digital or non-digital (e.g., lab notebooks, physical samples).





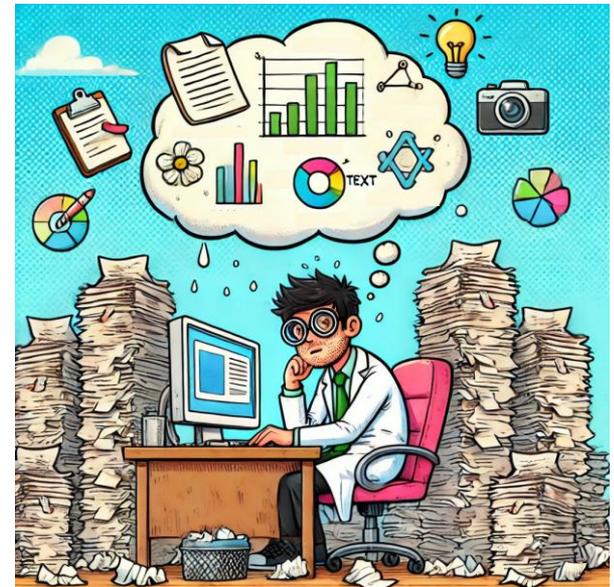
# The Role of Research Data in Science

---

- Research data are the foundation of all scientific work – they document observations, measurements, and results.
- They may include numerical values, text, images, software code, audio recordings, and more.
- Proper handling of data ensures reproducibility, transparency, reuseability, findability, and scientific credibility.

# Why Research Data Management(RDM) 1/2?

- Digital research data are expensive to generate but easy to lose or misuse.
- **You**, Funders and journals require data to be FAIR: Findable, Accessible, Interoperable, Reusable.
- Good RDM supports efficient, ethical, and impactful research workflows.
- Support research data management in Partnership projects in line with EC requirements.





# Why Research Data Management(RDM) 2/2?

---

- Support research data management in Partnership projects in line with EC and FAIR requirements.
  - Follow the obligations outlined in the Grant Agreement
  - Submit a Data Management Plan (DMP) if required
    - First version: within 6 months of project start
    - Updates: in each reporting period and upon major changes
- **Preparation for AI-driven research:** Only machine-readable, well-structured data can be reused in automated systems and knowledge graphs.



# Paradigm Shift: From Data Sharing to Data Visiting

---

- Traditional model: Data is sent to others – loss of control
- New approach: Algorithms "visit" the data
- Ensures privacy, sovereignty, and ethical access  
Use case: Sensitive data (e.g. health, biodiversity)
- "You want our data? Come visit us." — African Union



# FAIR Principles – Summary

---

- Findable – Metadata and identifiers (e.g. DOI) make data easy to locate.
- Accessible – Data can be retrieved using standard protocols (even with authentication).
- Interoperable – Use of standard vocabularies and formats enables integration.
- Reusable – Clear licenses and rich metadata support use by others.

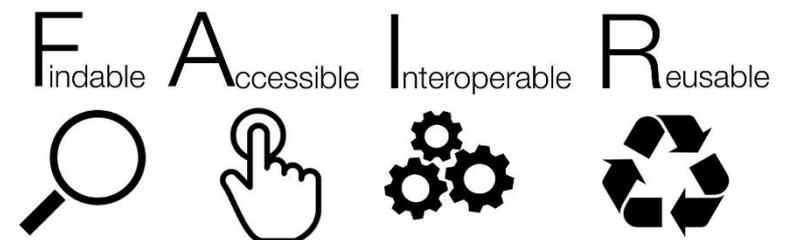
**Machine-actionable** – Data structures must allow automated interpretation by software systems.



# FAIR + AI = Fully AI-ready Data

---

- FAIR evolved: Data must be machine-actionable
- Beware of "Fake FAIR" or "FAIR-washing"
- FAIR Digital Objects: identifiable, structured, and semantically rich
- Goal: "The machine knows what I mean"





# Examples of “Fake FAIR” or “FAIR-washing”

---

- **Findable**



DOI assigned



but no or poor metadata

- **Accessible**



Access possible via login or request



but not openly accessible or lacks clear access conditions

- **Interoperable**



Uses XML or JSON



but without controlled vocabularies or community standards

- **Reusable**



License provided



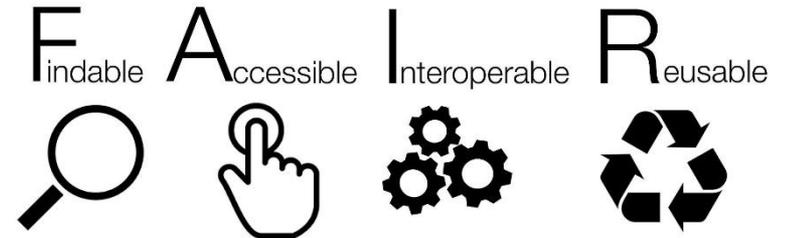
but unclear, contradictory, or overly restrictive license



# Fair Principles in practice

---

## *The Go FAIR initiative* <sup>1)</sup>



1. Make data openly available where possible
2. Have policies & plans. Preserve data of long-term value
3. Metadata for discovery / reuse. Link to data from publications
4. Be mindful of legal, ethical and commercial constraints
5. Allow limited embargoes to protect the effort of creators
6. Acknowledge sources to recognise IP and abide by T&Cs
7. Ensure cost-effective use of public funds for RDM

**FAIR Implementation Profiles (FIPs)** are a way to define discipline-specific FAIR criteria in a machine-readable way.

1) international initiative that supports the practical implementation of the FAIR principles for research data management.

Source: <https://www.go-fair.org/fair-principles/>



# FAIR by Design: FAIR Implementation Profiles (FIPs)

---

- Disciplines have specific FAIR requirements
- FIPs describe machine-readable FAIR compliance
- Over 480 global profiles (incl. environment, metrology, history)
- Goal: automated FAIR verification and cross-domain integration



# Why Most Data Is Not Reusable

---

- 80% of data becomes unusable after project end
- Broken links in scientific papers = lost data
- Missing metadata = no integration or reuse
- Fix: sustainable, machine-readable metadata structures



# DMPs – Help

---

- Template based on Horizon 2020 (42 questions); adapted for EURAMET by TC-IM 1449
- Template + sample texts from EURAMET's MSU
- Personal advice
- DMP platform: **MetRDMO**
  - Hosting: TU Darmstadt
  - Content: TC-IM 1449
  - Coordination: PTB





# Advantages of a DMP platform

---

- Structured record of project- and data-related information
- Support for answering questions
  - Smaller, clear questions
  - Sample answer texts
  - Help texts
- Re-use of information:
  - Textual output - DMP generation and versioning
  - Project documentation; Notation and initiation of tasks (“machine actionable”)
  - Structured research data management
  - Transmission of information from/to repositories, vocabulary servers, ...
- Application-specific question catalogues & export templates
- Large community; material exchange via GitHub



# Consequences of Poor Data Management

---

- Data loss or corruption can make results unverifiable or invalid.
- Unclear ownership or permissions can block publication or sharing.
- Lack of documentation reduces data reuse and collaboration potential.



# Pattern Recognition $\neq$ Understanding

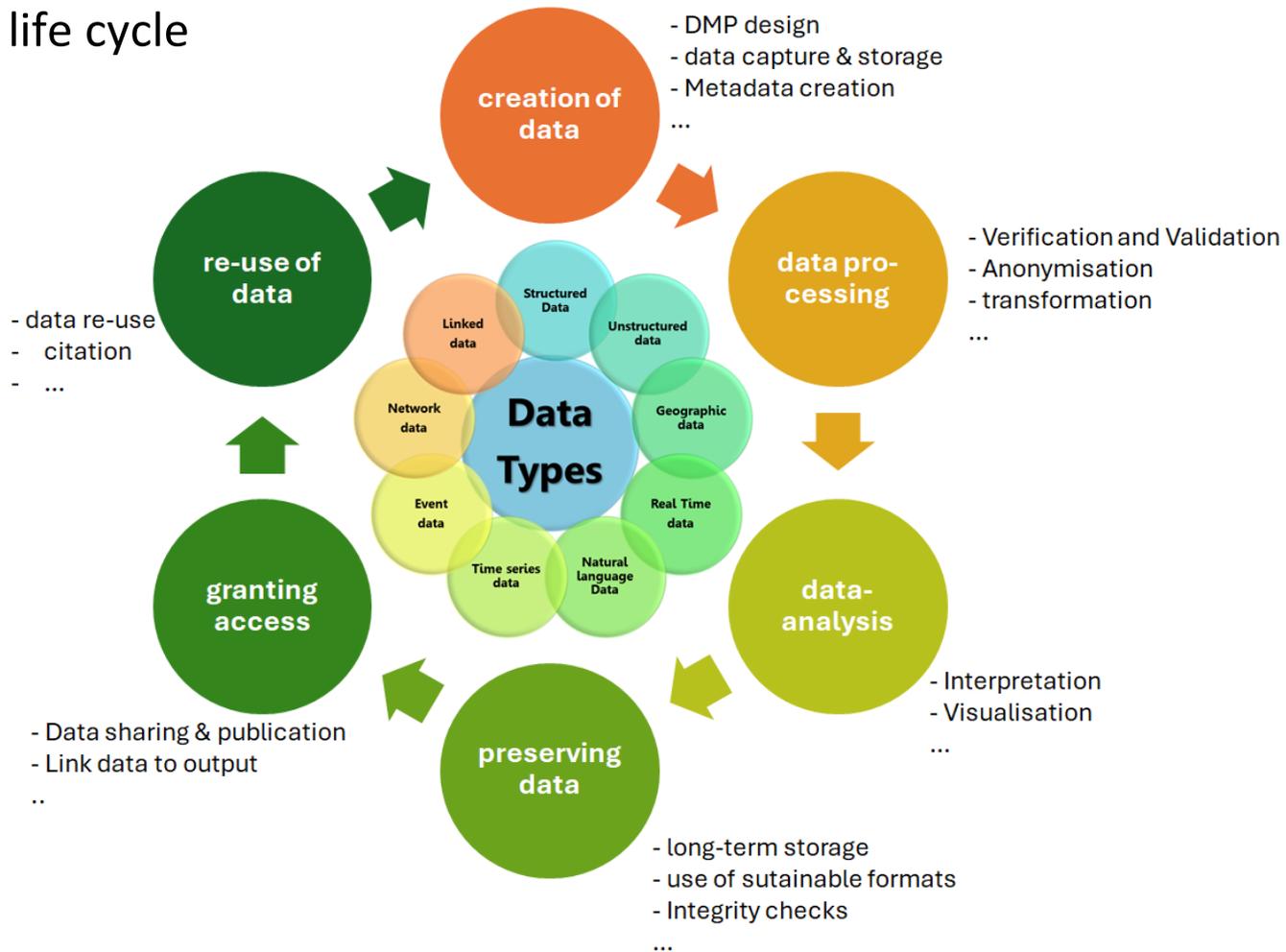
---

- AI finds patterns, but lacks semantic comprehension
- There are many examples where AI failed to find the correct data or misinterpreted it.
- Needed: Hybrid Intelligence (human insight + AI scale)
- Output must be understandable and verifiable by humans



# Research data management

- Data life cycle





# Roles in Research Data Management

---

- Researchers: Generate and document data responsibly.
- Data stewards: Ensure quality, metadata, and long-term access.
- IT support: Provide secure infrastructure and backup.
- Project leads: Coordinate policy, training, and compliance.



# Intro to Data Management Plans (DMPs)

---

A brief plan written at the start of your project to define:

- how will your data be created?
- how will it be documented?
- who will access it?
- where will it be stored?
- who will back it up?
- whether (and how) will it be shared & preserved?

DMPs are often submitted as part of grant applications, **but are useful whenever researchers are creating data.**

DMPs can **help working in a more structured way**, remind about tasks, collect together metadata / information to be reused later in the data lifecycle.

Source: Data Management Plans: <https://www.dcc.ac.uk/dmps>



# RDM and Metrology

---

- Metrological research requires traceability, comparability, and reproducibility of data.
- Standardized data handling ensures confidence in results across laboratories and projects.
- RDM supports digitalization efforts and structured knowledge transfer in metrology.



# The DMP – Your RDM Strategy

---

- A Data Management Plan (DMP) describes how data will be handled during the project.
- It covers types, formats, responsibilities, storage, sharing, and preservation.
- DMPs are living documents, updated as the project evolves.



# Initial tips on writing DMPs

---

## *Tips on writing DMPs:*

- Keep it simple, short and specific
- Seek advice - consult and collaborate
- Base plans on available skills and support
- Make sure implementation is feasible
- Justify any resources or restrictions needed
- If possible, use a tool saving the information in machine-readable format for better reusability

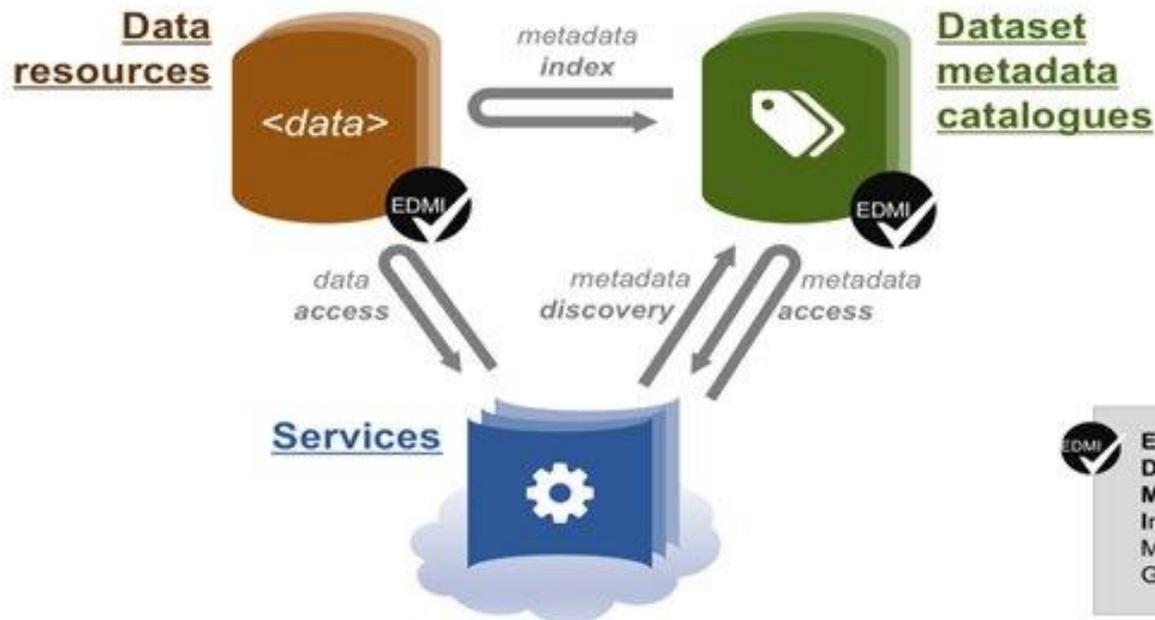
Source (from DCC): [www.youtube.com/watch?v=7OJtiA53-Fk](https://www.youtube.com/watch?v=7OJtiA53-Fk)



# Short Intro to metadata

*Metadata (“data about data”)*

- metadata in research



Source: <https://eoscpilot.eu/metadata-catalogues-strategy>  
EDMI project: <https://eosc-edmi.github.io/>



# TC-IM 1449: Facts

<https://www.metrology-rdm.eu/>

## Research data management in European metrology

- 14 active NMIs + 7 in ML  
+ 4 overseas observers (NZ, TH, MX, AR)
- 20 active members + 35 in ML
- 4 presence workshops; 6 plenary meetings/year;  
WP meetings by necessity

**New participants are welcome anytime!**

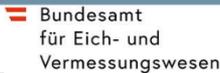
### COORDINATING INSTITUTE

PTB (Germany)



### PARTICIPATING PARTNERS

BEV (Austria)



CMI (Czechia)



DTI (Denmark)



IMBiH (Bosnia and Herzegovina)



INRIM (Italy)



IPQ (Portugal)



JV (Norway)



METAS (Switzerland)



MIKES (Finland)



NPL (United Kingdom)



RISE (Sweden)



SMD (Belgium)



VSL (Netherlands)



TECHNICAL COMMITTEE  
“Interdisciplinary Metrology”

WG “Metrology for Digital  
Transformation”

TC-IM  
1448

TC-IM  
1551

TC-IM  
1449



# Thanks for your attention!

**A special thanks to:** Alen Bošniaković, Diego Nahuel Coppa, Federico Grasso Toro, Robert Gunn, Mehrija Hasičić, Jean-Laurent Hippolyte, Daniel Hutzschenreuter, Holger Israel, Maitane Iturrate-García, Martin Koval, Michaela Küpferling, Keith Lines, Åge Andreas Falnes Olsen, Olivier Pellegrino, Anne-Sophie Piette, Marina Romanchikova, Thomas Wiedenhöfer, Louise Wright, Zoltán Zelenka, Narin Chantawong, Romain Coulon, Francisco Javier Flamenco Sandoval, Anjali Sharma, Hugo Gasca Aragón, Blair Hall, Gertjan Kok, Miruna Dobre, Sascha Eichstädt, Wolfgang Schmid, Joachim Erich Meier, Erkan Danaci, Damir Ilić, Luca Callegaro, Francesca Romana Pennecci

## Research data management in European metrology



### EMPIR 2021 data outputs:



### Ambitions:



### Measures/means:



### BENEFITS:

- enables faster open science development
- boosts technical competences
- helps to regulate the research ecosystem
- fosters cross-disciplinary interactions

The TC-IM 1449 project supports the implementation of good research data management practices by the EURAMET community





# Workshop Program and Schedule

<u>Time</u>	<u>Contents</u>
<b>2025-06-03, 14:00-15:30 CEST</b>	<b>"RDM in EPM projects" Module 1: Motivation, obligations and support</b>
5 min	Greeting
30+10 min	Plenary: Motivation and general information
10+10 min	Breakout session 1: RDM in JRP/CSP proposals
10+10 min	Breakout session 2: RDM in starting/running projects
10+10 min	Plenary: Obligations in EPM projects (EPM guides); Support
10 min	Questions and discussion
<b>2025-06-12, 14:00-15:30 CEST</b>	<b>"RDM in EPM projects" Module 2: Planning and data publication</b>
5 min	Greeting Plenary: Obligations in EPM projects
20+10 min	Breakout session 1: Describing RDM in JRP/CSP proposals
20+10 min	Breakout session 2: Preparing a data management plan + RDMO
20+10 min	Plenary: Data publication + sensitive data + versioning
10 min	Questions and discussion
<b>2025-06-20, 14:00-15:30 CEST</b>	<b>"RDM in EPM projects" Module 3: Metadata and good practices</b>
5 min	Greeting
20+10 min	Metadata: how to select them and how to record them
20+10 min	Breakout session: Exploration of the SI dig.fr. + API query
20+10 min	Breakout session: Case study about lab notebooks
10 min	Questions and discussion