

Web Tutorial 1: Introduction to Research Data Management

TC IM 1449: Research Data Management and
the European Open Science Cloud

Dr.-Ing. Federico Grasso Toro



This work is licensed under a Creative Commons Attribution 4.0 International (CC-BY 4.0) license.

Content

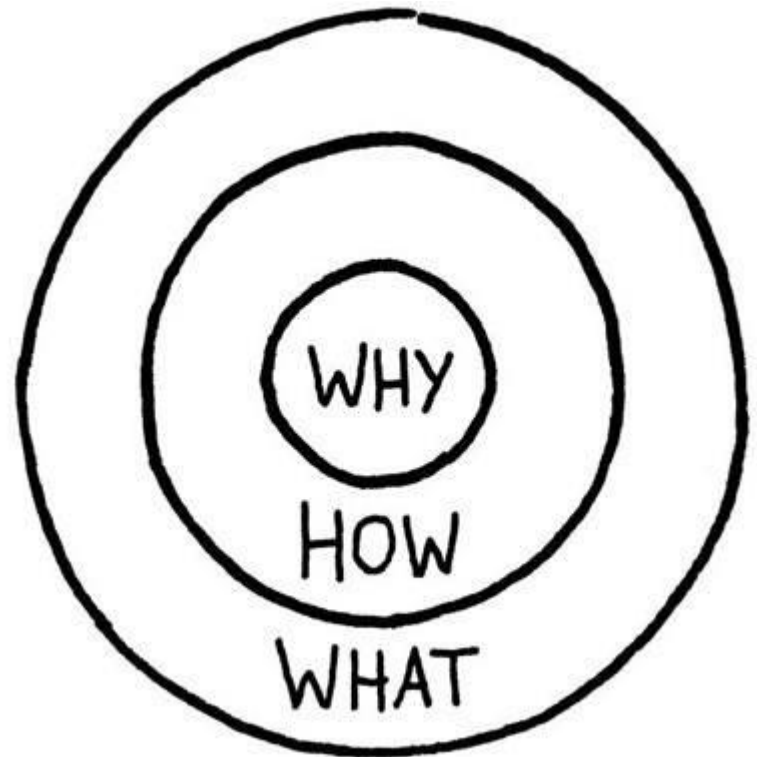
- Part 1: Intro to Research Data Management (RDM)
- Part 2: Intro to RDM in metrology (i.e. EMPIR projects)
- Part 3: Intro to Data Management Plans (DMPs)
- Part 4: Intro to Metadata
- Part 5: Intro to RDM in EURAMET projects

Content

- Part 1: Introduction to Research Data Management
 - Why RDM?
 - Whereby of RDM
 - RDM related definition(s)

TC IM 1449 Tutorials

- Guidance for researchers
- Motivation for adopting FAIR
- Aim at EOSC-ready



Part 1: Introduction to Research Data Management

Why RDM?

- **Digital technologies** now are used very widely in research, and this is enabling new research and scientific paradigms.
- Research funders and publishers know that digital research data can be expensive to produce but inexpensive to share, making **reusability** more feasible and desirable.
- The challenge is to **ensure digital research findings** can be reproduced and cited.



Why = The Purpose

What is your cause? What do you believe?

How = The Process

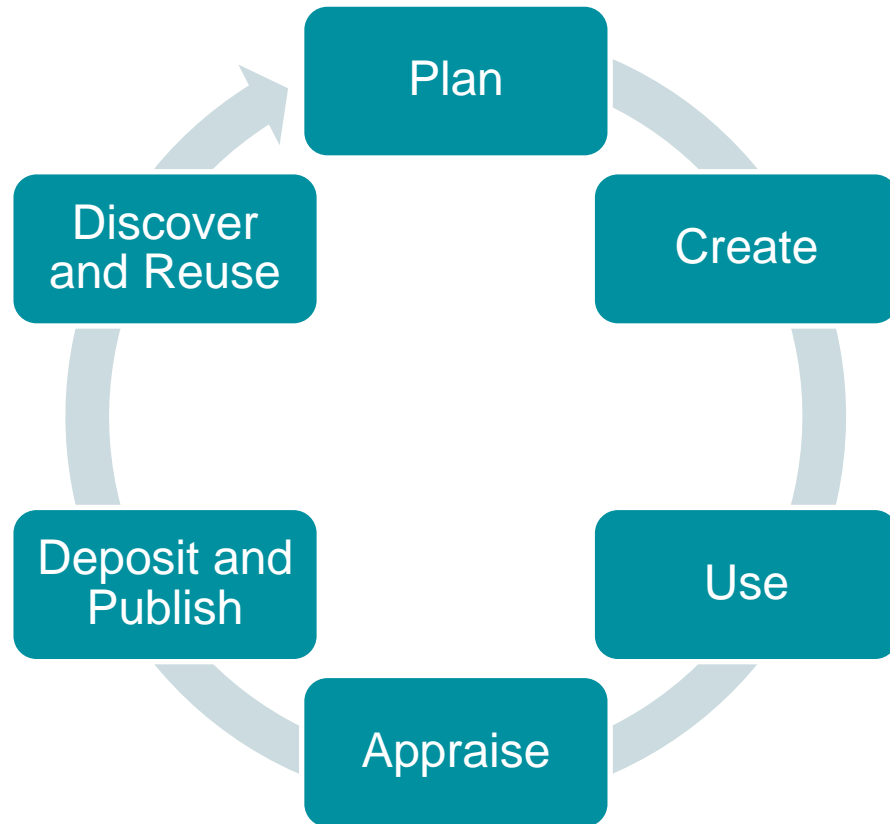
Specific actions taken to realize the Why.

What = The Result

What do you do? The result of Why. Proof.

Part 1: Introduction to Research Data Management

Whereby of RDM



Part 1: Introduction to Research Data Management

RDM definition

Research data are (digital) information which is created during scientific activity (e.g. measurements, observations, surveys, literature search, etc.).

They constitute the fundament of any scientific work and collect their results.



Source: <https://www.forschungsdaten.info/praxis-kompakt/glossar/#c269824>

Part 1: Introduction to Research Data Management

RDM related definition(s)

RDM planning: support and services for planning activities typically performed before research data is collected / created.

Active data infrastructure: facilities to store data being actively used in current research activities, to provide access to that storage and tools to assist in working with the data.

Data stewardship: tools and services to aid in the description, deposit, and continuity of access to completed research data outputs.

Data management support: awareness raising and advocacy, data management guidance and training.



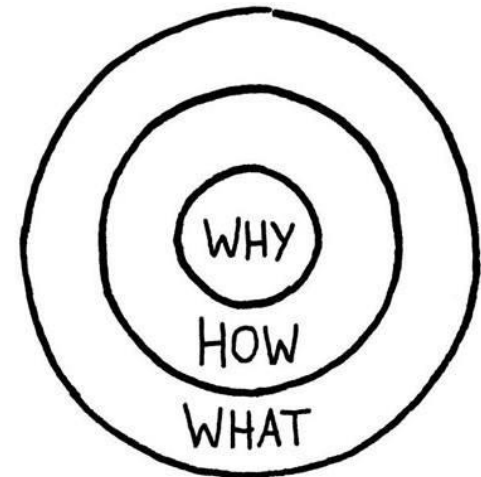
Source: <https://www.dcc.ac.uk/resources/developing-rdm-services/rdm-strategy-moving-plans-action#Definitions>

Part 1: Introduction to Research Data Management

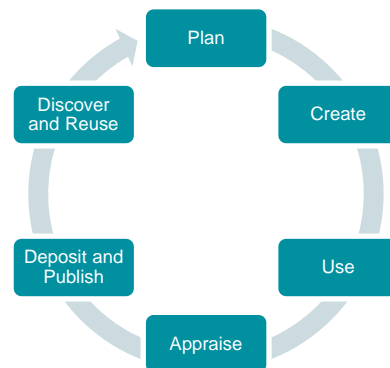
Summary Part 1: **Data management is part of good research practice**

Why do we need to manage research data?

- To make research easier!
- To stop yourself drowning in irrelevant stuff
- In case you need the data later
- To avoid accusations of fraud or bad science
- To share data so others can use and learn from it
- To get credit for producing the data



Whereby of RDM:



Summary of reason:

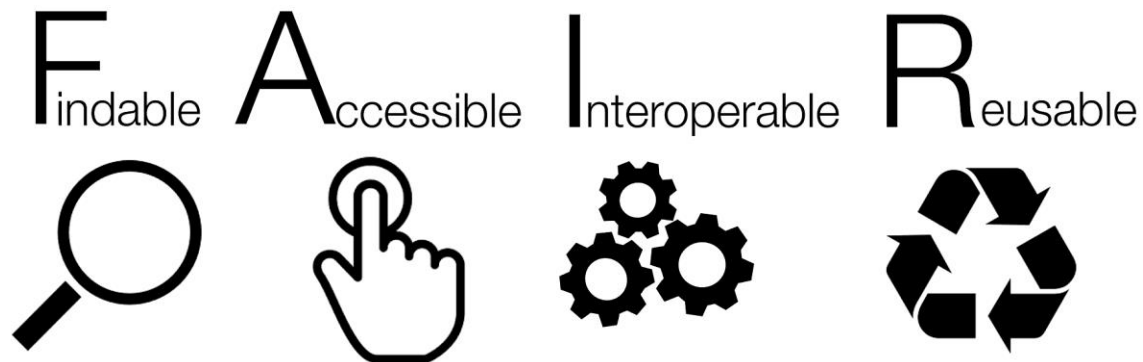
- To increase efficiency
- To ease sharing / re-use
- To reduce data losses



Content

- Part 2: Intro to RDM in metrology (i.e. EMPIR)
 - Grant Agreement*

ARTICLE 29 — DISSEMINATION OF RESULTS — **OPEN ACCESS** — VISIBILITY OF EMPIR FUNDING
 - FAIR principles



Part 2: Intro to RDM in metrology

Data Policy (Common principles on research data)

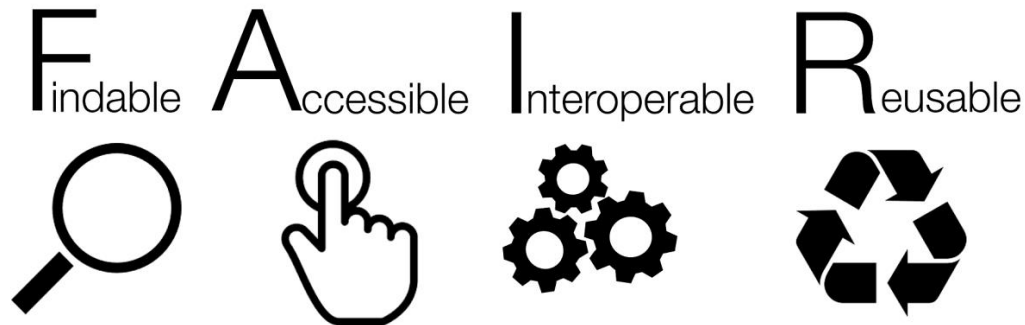
Common Principles:

<https://www.ukri.org/funding/information-for-award-holders/data-policy/common-principles-on-data-policy>

Guidance on best practice in the management of research data:

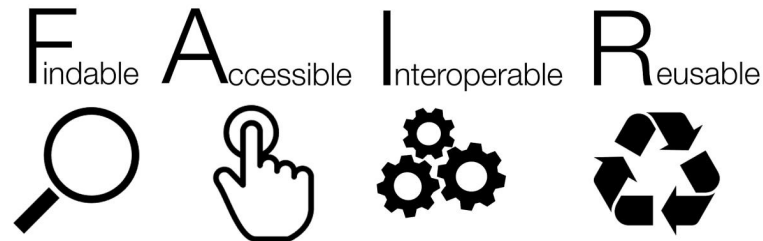
<https://www.ukri.org/wp-content/uploads/2020/10/UKRI-221020-guidance-on-best-practice-management-of-research-data.pdf>

FAIR principles



Part 2: Intro to RDM in metrology (i.e. EMPIR)

The Go FAIR initiative



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

Source:

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

Extra sources:

GO-FAIR Initiative: <https://www.go-fair.org/go-fair-initiative/>

HOW TO GO FAIR: <https://www.go-fair.org/how-to-go-fair/>

International FAIR Symposium: <https://www.go-fair.org/events/international-fair-convergence-symposium/>

Part 2: Intro to RDM in metrology (EMPIR)

General approach to research

Box 2 | The FAIR Guiding Principles

To be Findable:

- 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 it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- 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

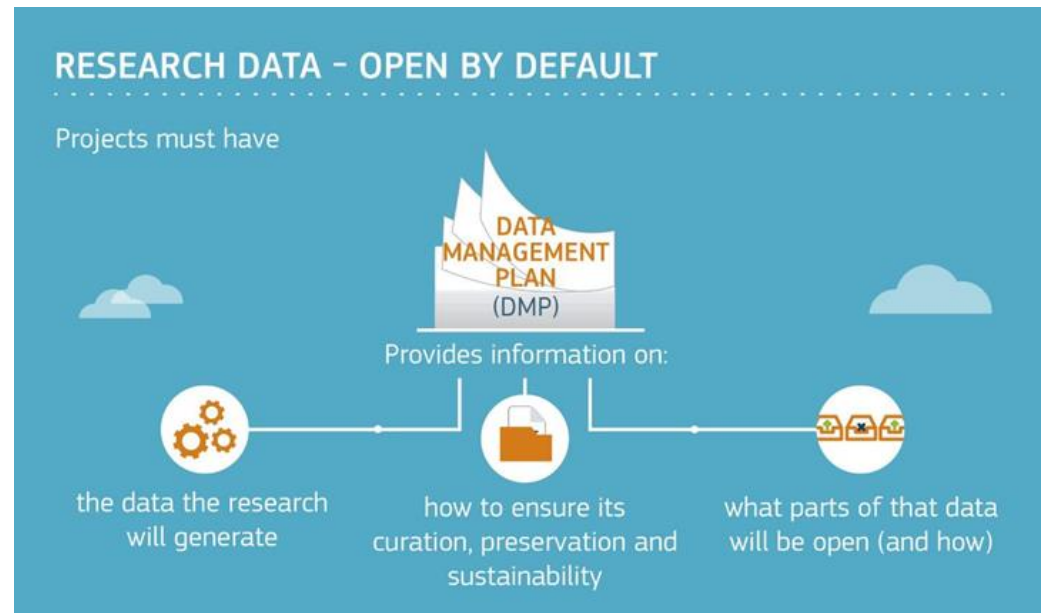
To be Reusable:

- 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

Content

- Part 3: Intro to DMP (DCC's funder requirements)

- Brief Intro
- Existing tools
- Common themes
- Initial tips
- Tutorial 2 Information**



Source: <https://www.openaire.eu/what-is-a-data-management-plan>

Part 3: Intro to Data Management Plans (DMPs)

Intro to 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.

Part 3: Intro to Data Management Plans (DMPs)


Existing tools (1)

[Home](#) [Public DMPs](#) [Funder requirements](#) [Help](#)  Language ▾


Welcome

DMPonline helps you to create, review, and share data management plans that meet institutional and funder requirements. It is provided by the Digital Curation Centre (DCC).


Join the growing international community that have adopted DMPonline:




17,622 Users



203 Organisations



23,083 Plans



89 Countries

Some funders mandate the use of DMPonline, while others point to it as a useful option. You can [download funder templates](#) without logging in, but the tool provides tailored guidance and example answers from the DCC and many research organisations. Why not sign up for an account and try it out?

Sign in

Create account

* Email

* Password


[Forgot password?](#)

☐ Remember email

Sign in

- or -

Sign in with your institutional credentials

© 2010 - 2020 Digital Curation Center • [About](#) [Contact us](#) [Terms of use](#) [Privacy statement](#) [Accessibility statement](#) [Github](#) 

Tool description: <https://www.dcc.ac.uk/dmponline>
Tool (online): <https://dmponline.dcc.ac.uk/>

Part 3: Intro to Data Management Plans (DMPs)

Existing tools (2)

[RDMO](#) [News](#) [About ▾](#) [Events ▾](#) [Information ▾](#) [Community ▾](#) [Deutsch](#) [English](#)

The Research Data Management Organiser (RDMO) enables institutions as well as researchers to plan and carry out their management of research data. RDMO can assemble all relevant planning information and data management tasks across the whole life cycle of the research data.

RDMO is ready for application in smaller or bigger projects. In the next projectphase, which started November 2017, the RDMO tool will be extended and the project partners AIP, FHP, and KIT Library will collaborate with the RDMO users to improve its usage. The tool will be extended by enhancing its implementation of roles and interfaces to institutional infrastructure, e.g. repositories, ticketing systems, and the infrastructure for authentication and authorization. Tutorials, documentation and other material are planned for dissemination, and workshops for users and developers.

On these pages, we provide information about the [project](#), our [workpackages](#), the [software](#) and its [documentation](#), [news](#) and [cooperations](#).

More information about the project and our software can be found on the following websites:

The free software:
github.com/rdmorganiser

Demo instance:
rdmo.aip.de

Register for our public mailing list:
rdmo@listserv.dfn.de

Follow us on Twitter:
twitter.com/rdmorganiser

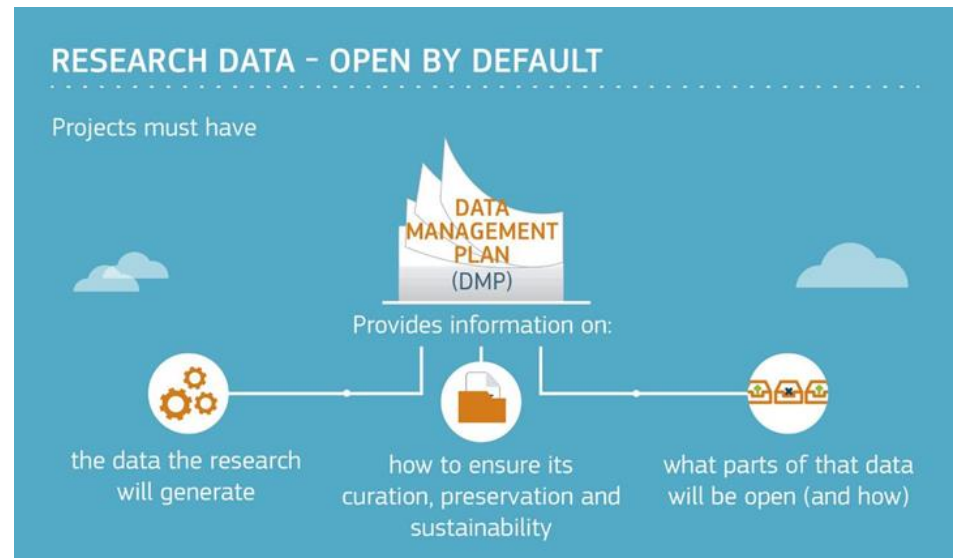
Discuss with us on Slack:
rdmo.slack.com (Send us a e-mail: rdmo-team@listserv.dfn.de and we will send you an invitation.)

Tool description: <https://rdmorganiser.github.io/en/>
Tool Documentation: <https://rdmorganiser.github.io/en/documentation/>

Part 3: Intro to Data Management Plans (DMPs)

Common Themes for DMPs

1. Description of data to be collected / created (i.e. content, type, format, volume...)
2. Standards / methodologies for data collection & management
3. Ethics and Intellectual Property (highlight any restrictions on data sharing e.g. embargoes, confidentiality)
4. Plans for data sharing and access (i.e. how, when, to whom)
5. Strategy for long-term preservation.



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

Part 3: Intro to Data Management Plans (DMPs)

Tips on writing DMPs:

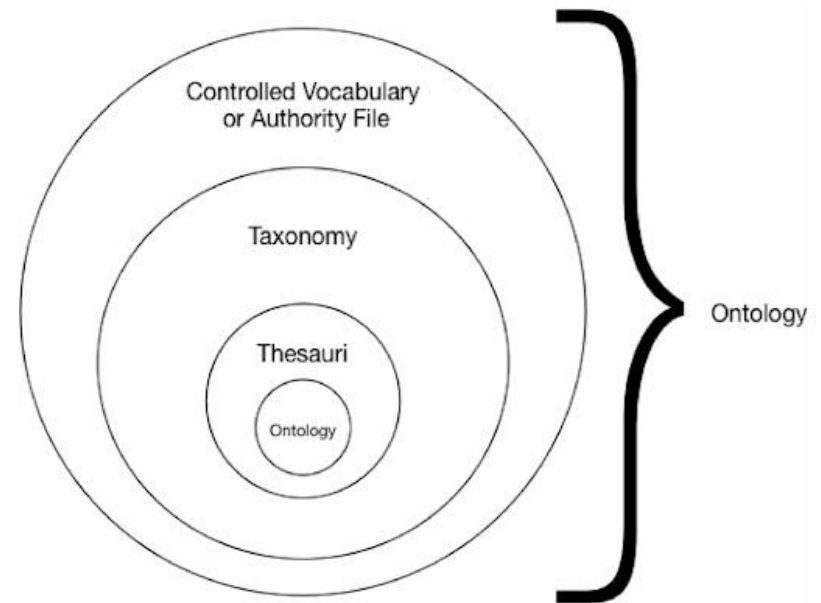
- Keep it simple, short and specific
- Seek advice - consult and collaborate
- Base plans on available skills and support
- Make sure implementation is feasible
- Justify any resources or restrictions needed
- If possible, use a tool saving the information in machine-readable format for better reusability

TUTORIAL 2: Preparation and implementation of data management plans (DMPs) for EMPIR projects

Monday, February 15th, 10 AM CET

Content

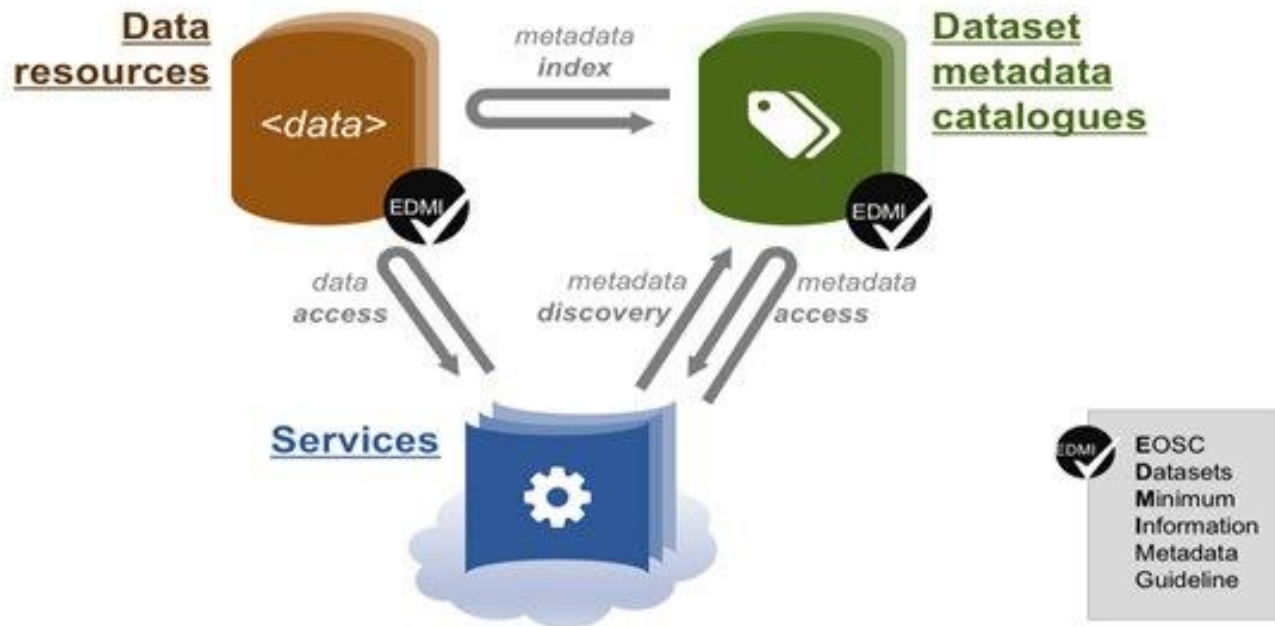
- Part 4: Intro to metadata
 - Intro to metadata concept
 - Related to FAIR
 - Metadata Retrieval API
 - **Information about Tutorial 3**



Part 4: Intro to metadata

Metadata (“data about data”)

- metadata in research



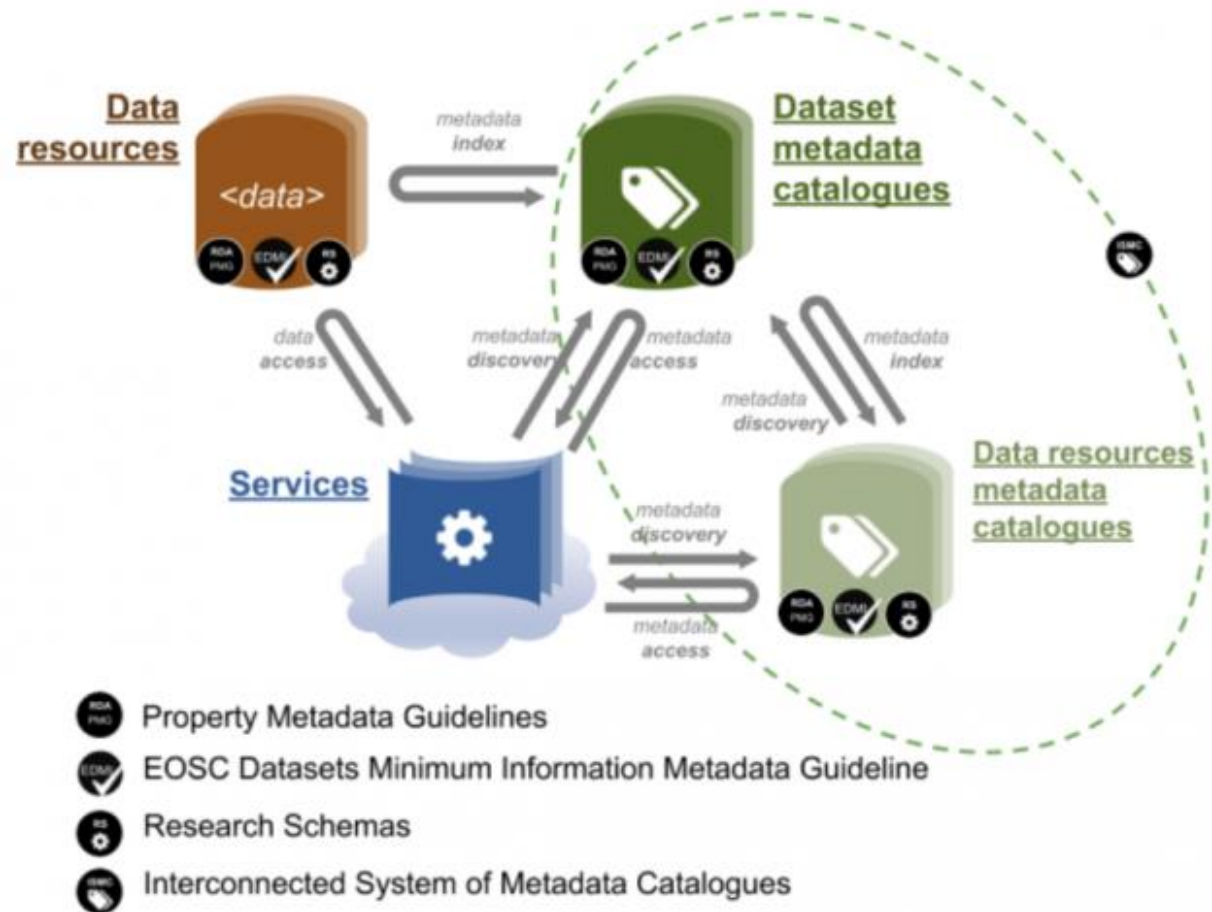
Source: <https://eoscipilot.eu/metadata-catalogues-strategy>

EDMI project: <https://eosc-edmi.github.io/>

Part 4: Intro to metadata

Metadata (“data about data”)

- Related to FAIR



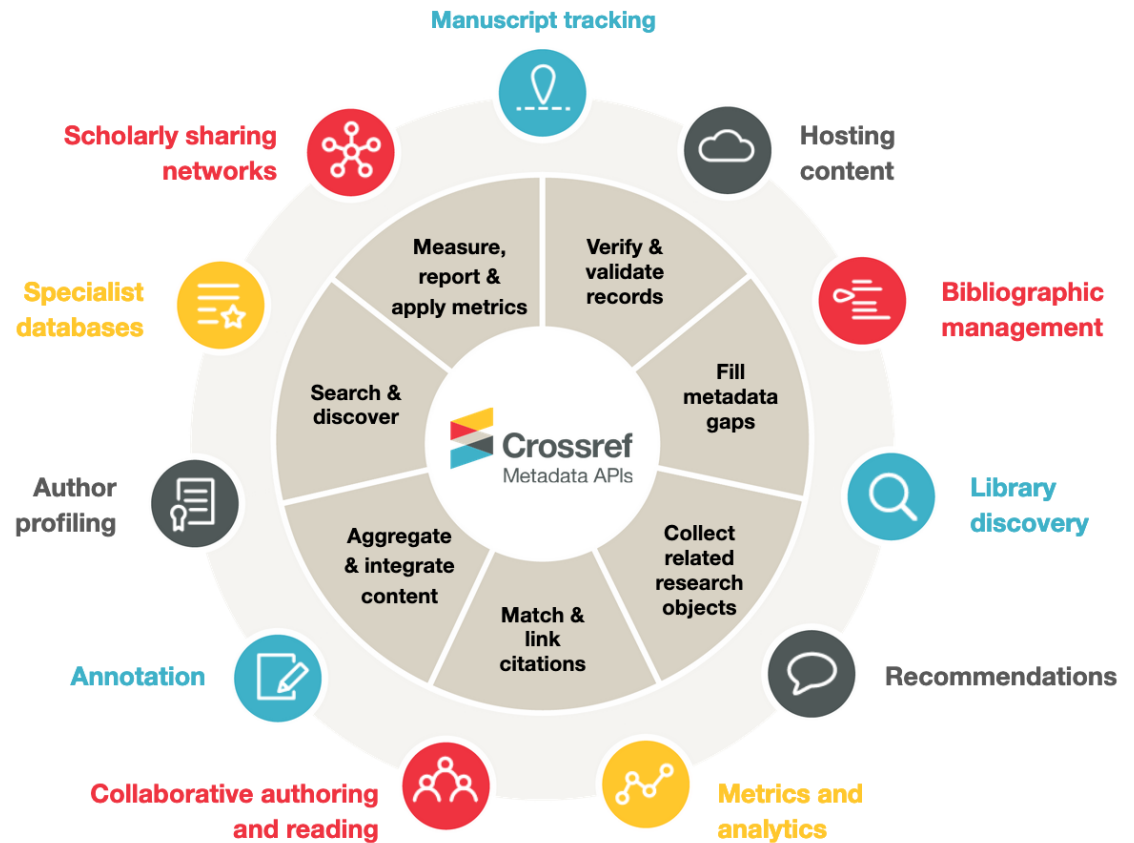
Source: <https://eoscpiot.eu/news/eosc-data-interoperability-ensure-availability-scientific-data>

EOSC Metadata Guidelines: <https://eoscpiot.eu/edmi-metadata-guidelines>

Part 4: Intro to metadata

Metadata (“data about data”)

- Metadata retrieval API



Source: <https://www.crossref.org/services/metadata-retrieval/>

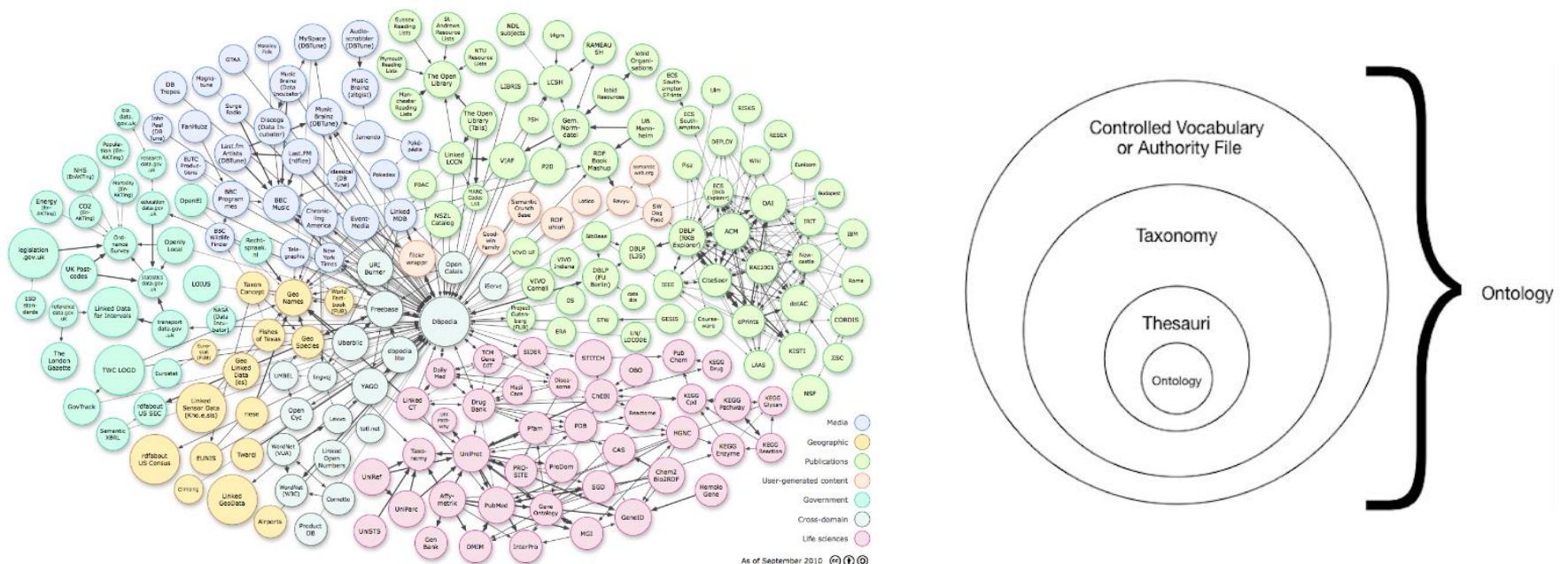
Part 4: Intro to metadata

The role of METADATA and FAIR principles:

<http://aims.fao.org/activity/blog/fair-principles-digital-objects-role-metadata>

TUTORIAL 3: METADATA for RDM in EMPIR projects

Friday, March 11th, 2021



Content

- Part 5: Intro to RDM in EURAMET projects
 - Metadata required by Grant Agreement (Article 29).
 - Minimum Metadata for good practice: Compliance for FAIR.
 - Towards EOSC Minimum Viable Ecosystem.

Part 5: Intro to RDM in EURAMET projects

Metadata required by Grant Agreement

ARTICLE 29 — DISSEMINATION OF RESULTS — OPEN ACCESS — VISIBILITY OF EMPIR FUNDING

29.1 Obligation to disseminate results

Unless it goes against their legitimate interests, each beneficiary must — as soon as possible — ‘disseminate’ its results by disclosing them to the public by appropriate means (other than those resulting from protecting or exploiting the results), including in scientific publications (in any medium)

This does not change the obligation to protect results in Article 27, the confidentiality obligations in Article 36, the security obligations in Article 37 or the obligations to protect personal data in Article 39, all of which still apply.

A beneficiary that intends to disseminate its results must give advance notice to the other beneficiaries of — unless agreed otherwise — at least 45 days, together with sufficient information on the results it will disseminate.

Any other beneficiary may object within — unless agreed otherwise — 30 days of receiving notification, if it can show that its legitimate interests in relation to the results or background would be significantly harmed. In such cases, the dissemination may not take place unless appropriate steps are taken to safeguard these legitimate interests.

If a beneficiary intends not to protect its results, it may — under certain conditions (see Article 26.4.1) — need to formally notify EURAMET before dissemination takes place

Source: Grant Agreement (Art. 29)

Part 5: Intro to RDM in EURAMET projects

Metadata required by Grant Agreement

ARTICLE 29 — DISSEMINATION OF RESULTS — OPEN ACCESS — VISIBILITY OF EMPIR FUNDING

29.2: Open access to scientific publications

Each beneficiary must ensure open access (free of charge, online access for any user) to all peer-reviewed scientific publications relating to its results.

In particular, it must:

(a) as soon as possible and at the latest on publication, deposit a machine-readable electronic copy of the published version or final peer-reviewed manuscript accepted for publication in a repository for scientific publications;

Moreover, the beneficiary must aim to deposit at the same time the research data needed to validate the results presented in the deposited scientific publications.

(b) ensure open access to the deposited publication — via the repository — at the latest:

(i) on publication, if an electronic version is available for free via the publisher, or

(ii) within six months of publication (twelve months for publications in the social sciences and humanities) in any other case.

(c) ensure open access — via the repository — to the bibliographic metadata that identify the deposited publication."

The bibliographic metadata must be in a standard format and must include all of the following:

- the terms "EMPIR" "European Union (EU)" and "Horizon 2020"
- the name of the action, acronym and grant number;
- the publication date, and length of embargo period if applicable, and
- a persistent identifier.

Source: Grant Agreement (Art. 29)

Part 5: Intro to RDM in EURAMET projects

Minimum Metadata for good practice: Compliance for FAIR



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

Part 5: Intro to RDM in EURAMET projects

Towards “EOSC Minimum Viable Ecosystem”

EURAMET wants EMPIR (and EMP) to be EOSC-ready

SOFTWARE DEVELOPERS/SERVICE PROVIDERS

Interoperable services and open data rely on the principles of software openness. The software used in EOSC services should guarantee interoperability and comply with standards, be they de facto or by right (de Jure). Data produced and handled with EOSC software services should respect the FAIR principles; services within EOSC should be secure and comply with the European authorisation and authentication policies; as a general policy, the software elements are provided upstream to open source projects, to guarantee the required level of sustainability; to provide persistent identifiers, identification scheme and machine-readable metadata about the resources.

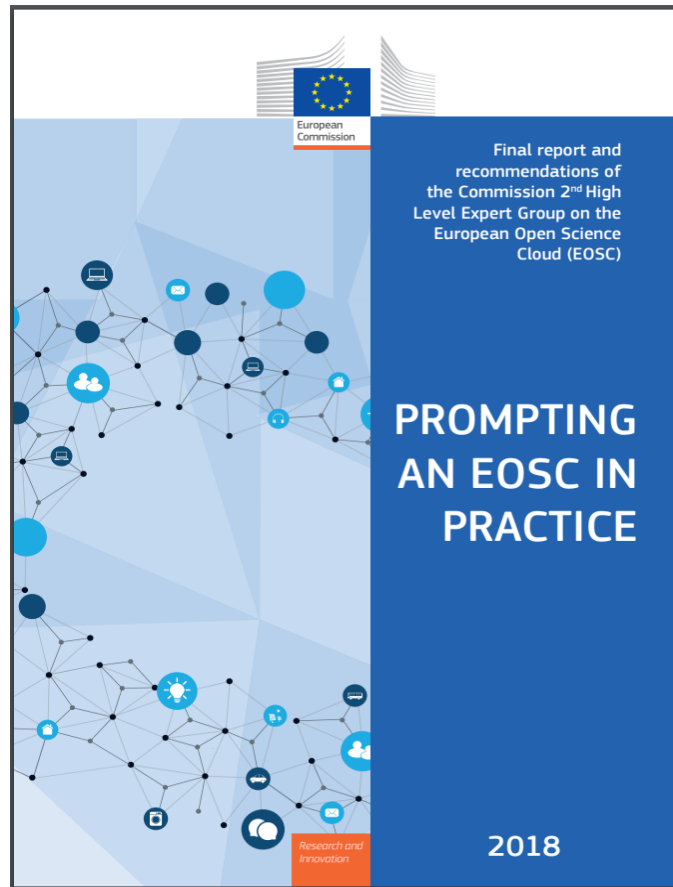
For this key activity to be successful in terms of engaging human talent, breakthrough ideas leading to innovation need to be awarded with the proper recognition. Putting in place transparent mechanisms to recognise successful software development, such as creating an **‘EOSC-Ready’ certification for software** products, would have a positive impact on the software development ecosystem in Europe. The successful development of an ‘EOSC-Ready’ branded software product, would improve the reputation of researchers and technologists and dynamically harness the potential of European developers, across academia and industry.

Source: Prompting an EOSC in practice

<https://op.europa.eu/en/publication-detail/-/publication/5253a1af-ee10-11e8-b690-01aa75ed71a1>

Part 5: Intro to RDM in EURAMET projects

Towards “**EOSC Minimum Viable Ecosystem**”



Source: Prompting an EOSC in practice

<https://op.europa.eu/en/publication-detail/-/publication/5253a1af-ee10-11e8-b690-01aa75ed71a1>

Thank you for your attention!

NEXT TUTORIAL:

**Preparation and implementation of
data management plans (DMPs) for
EMPIR projects**

Monday, February 15th, 10 AM CET

Speaker: Giacomo Lanza (PTB)

This work is licensed under a Creative Commons Attribution 4.0 International (CC-BY 4.0) license, which allows a free reuse and share for any purpose, as long as appropriate credit to the original source is provided. Please see

<https://creativecommons.org/licenses/by/4.0/> for more details.



This work is licensed under a Creative Commons Attribution 4.0 International (CC-BY 4.0) license.

Example from EMPIR projects

BiRD Project



Website: <https://www.birdproject.eu/>

F: <https://www.birdproject.eu/papers-and-presentations/>

AIR: <https://www.birdproject.eu/measurement-data/>

<https://www.birdproject.eu/references/>

Results: <https://www.birdproject.eu/project-results/>

Example from EMPIR projects

AeroTox

The screenshot shows the Zenodo website interface. At the top, the Zenodo logo is on the left, followed by a search bar containing "Search 18HLT02 AeroTox". To the right of the search bar are buttons for "Upload" and "Communities". Further right are buttons for "Log in" and "Sign up". Below the header, the main title of the community is "18HLT02 AeroTox - Measurements for mitigating adverse health effects from atmospheric particulate pollutants". On the left side, there are two filter panels: "All versions" with an "Access Right" filter showing "Open (1)", and "File Type" filter showing "Xlsx (1)". The main content area displays "Found 1 result." with a pagination bar showing "1" of 1 results. To the right of the results, there is a "Sort by:" dropdown menu set to "Most recent" and "asc.". Below the search bar, there is a blue button that says "January 11, 2021 (2. Excel sheets are easier to understand and completely in English)", a "Dataset" button, and an "Open Access" button. The title of the result is "Homogeneity Measurements in an Aerosol Mixing Chamber" by Giordano Andrea and Horender Stefan. The description states: "The published excel files contain the experimental data presented in the master thesis of Andrea Giordano entitled 'CFD Simulation of an Aerosol Mixing Chamber' available at https://infoscience.epfl.ch/record/280815?ln=en The data are presented in the thesis figs. 33 to 36 The nomenclat". It also says "Uploaded on January 13, 2021" and "1 more version(s) exist for this record". A "View" button is located to the right of the result.

zenodo

Search 18HLT02 AeroTox

Upload Communities

Log in Sign up

18HLT02 AeroTox - Measurements for mitigating adverse health effects from atmospheric particulate pollutants

All versions

Access Right

☐ Open (1)

File Type

☐ Xlsx (1)

Found 1 result.

< 1 >

Sort by: Most recent asc.

View

January 11, 2021 (2. Excel sheets are easier to understand and completely in English) Dataset Open Access

Homogeneity Measurements in an Aerosol Mixing Chamber

Giordano Andrea; Horender Stefan;

The published excel files contain the experimental data presented in the master thesis of Andrea Giordano entitled "CFD Simulation of an Aerosol Mixing Chamber" available at <https://infoscience.epfl.ch/record/280815?ln=en> The data are presented in the thesis figs. 33 to 36 The nomenclat

Uploaded on January 13, 2021

1 more version(s) exist for this record


Community in Zenodo

<https://zenodo.org/communities/aerotox/search?page=1&size=20>

Example from EMPIR projects

AeroTox

Indexed in



Publication date:
January 11, 2021


DOI:
DOI 10.5281/zenodo.4436486

Keyword(s):
Aerosol Homogenization CFD Validation

Awarding University:
EPFL

Related identifiers:
Previous versions
10.5281/zenodo.4432572

Communities:
18HLT02 AeroTox - Measurements for mitigating adverse health effects from atmospheric particulate pollutants

License (for files):
 Creative Commons Attribution 4.0 International

<https://www.openaire.eu/>

Community in Zenodo:

<https://zenodo.org/communities/aerotox/search?page=1&size=20>