

# Introduction to FAIR Data

Michal Růžička <ruzicka@ics.muni.cz> & EOSC CZ Secretariat  
Masaryk University | CERIT-SC

Institute of Physics of the Czech Academy of Sciences  
2025-11-13



Co-funded by  
the European Union



# About Me

- Leader of EOSC CZ open **working group Core Services**.
  - <https://www.eosc.cz/en/working-groups/core-services>
- Leader of **end-user-services** oriented **work packages** in the **NRP** project.
  - <https://www.eosc.cz/en/projects/national-repository-platform-for-research-data-os-i-nrp/national-repository-platform>
- **Open Science support team** member at CERIT-SC | ICS MU
  - Focus on **data management** and **FAIR data** support.
  - Including **sensitive data** (SensitiveCloud).
- Good **coffee** and **tea lover**.

13 November 2025

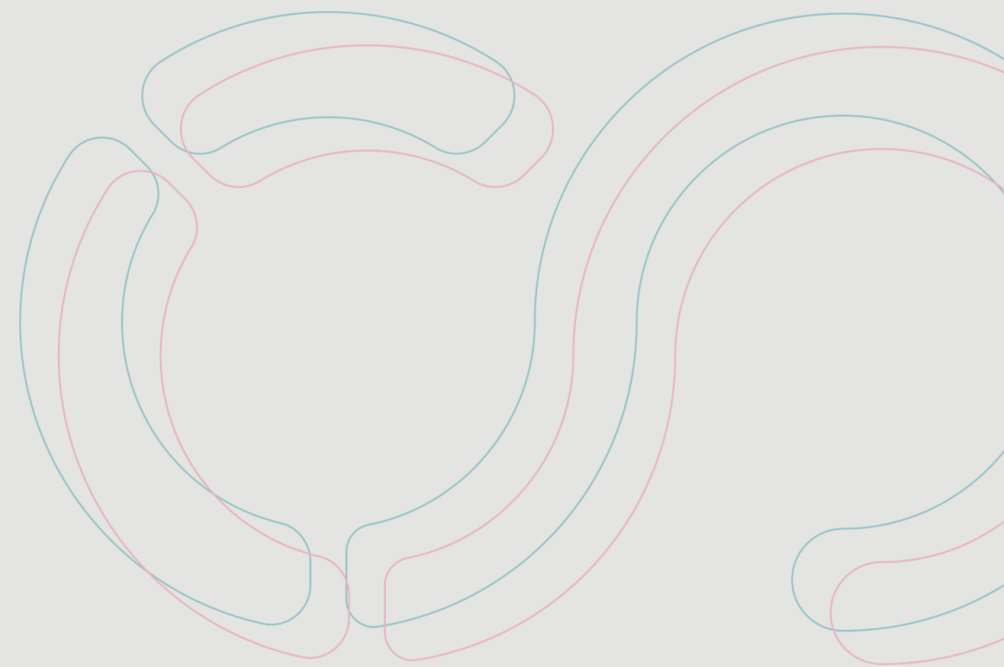
Introduction to FAIR Data

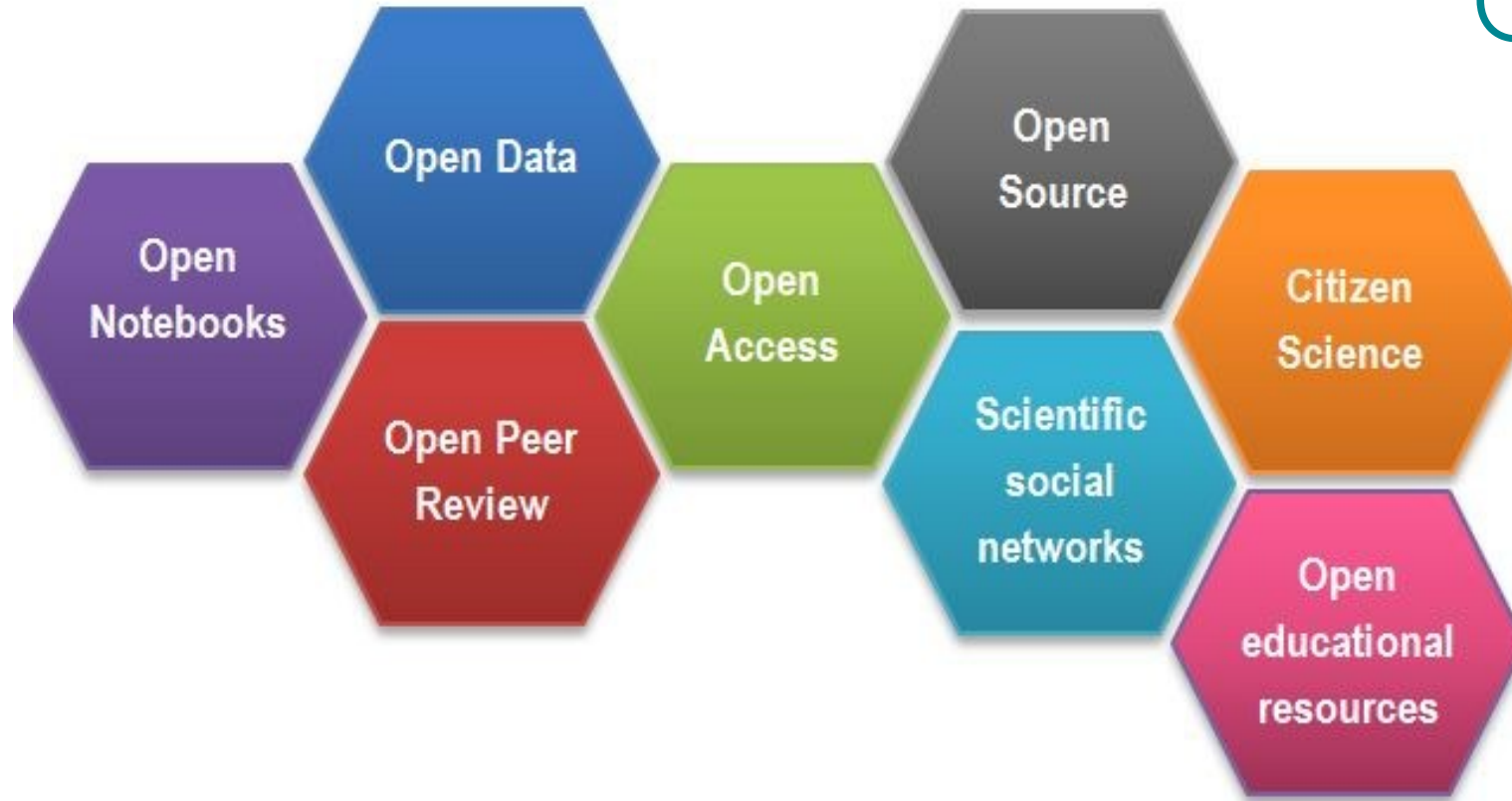


# Outline

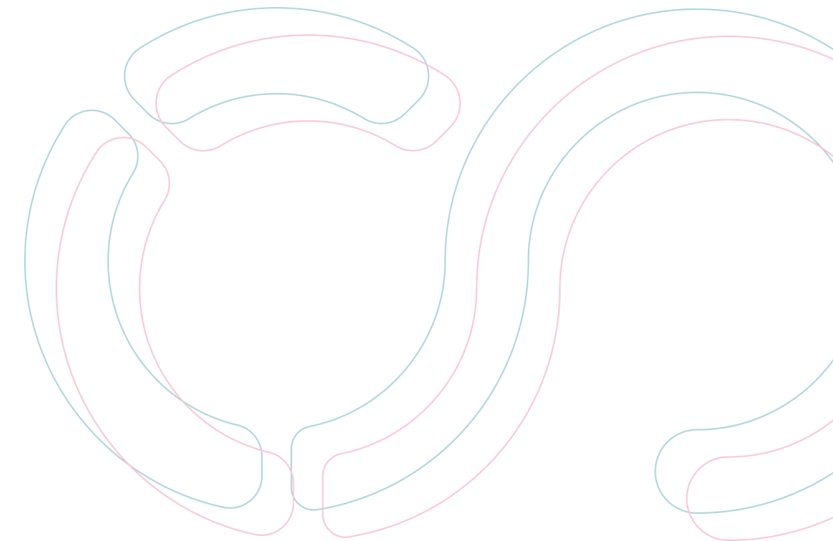
1. Context
2. Research Data Life-Cycle
3. Data, Managed Data, FAIR Data, Open Data...
4. Data Management Plan (DMP) in Research Project Calls
5. Where to Store Research Data
  - Practical Examples
6. Tools and Resources to Support FAIR Principles
7. Persistent Identifiers (PIDs)
8. Repositories for Research Data
9. European Open Science Cloud (EOSC)
10. National Repository Platform (NRP)
  - Focusing

# Context





# Open Science



Source: <https://www.fosteropenscience.eu/node/1420>

# Research Data Life-Cycle



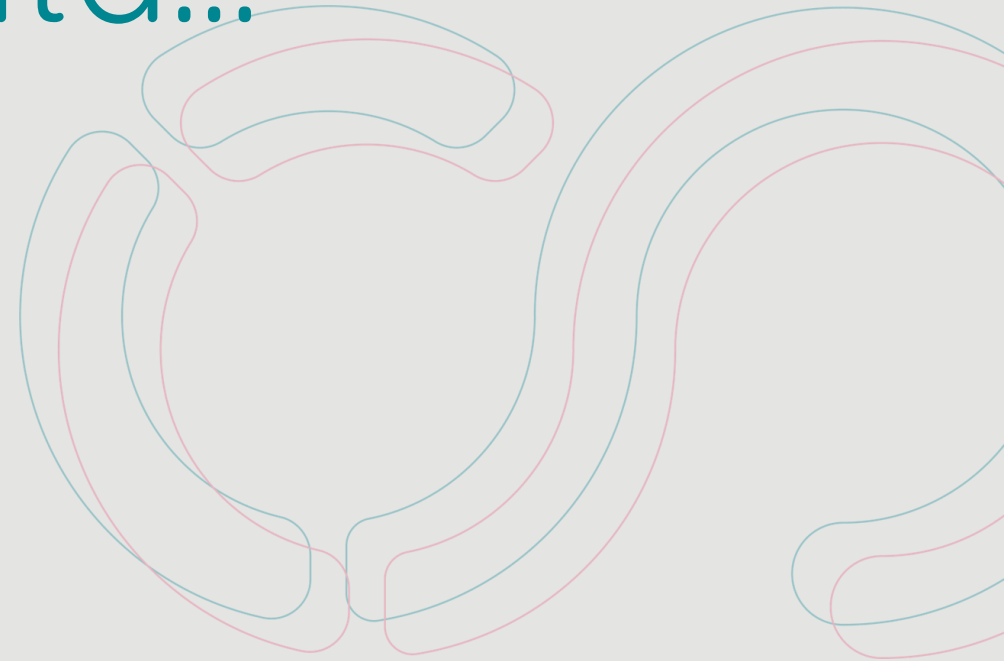
# Research Data Life-Cycle



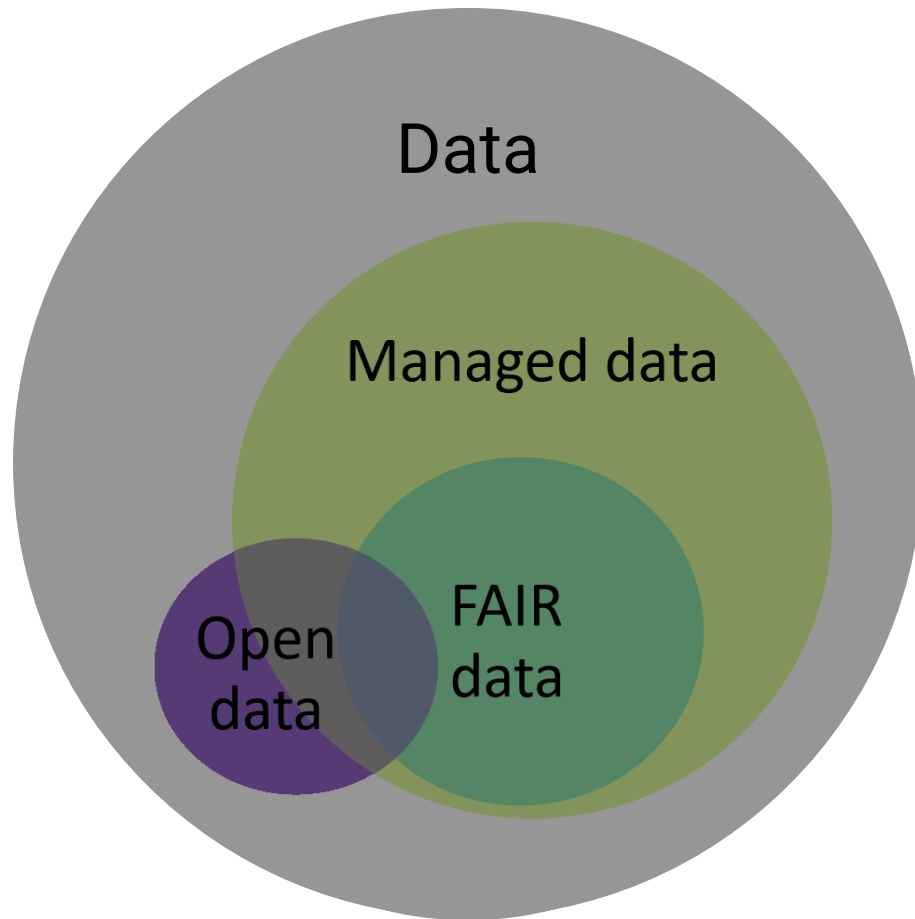
Source: ELIXIR RDMkit, <https://rdmkit.elixir-europe.org/>

- What **data do you (re)use**?
  - Including **licenses** that allow you to do so.
- What **data do you generate** and how?
- **Where** do you **store, backup, and long-term archive** data?
- How do you **permanently and uniquely identify** it?
- How do you **process** it?
- How do you **analyze** it?
- Where do you **publish and share** it?
- Who **pays** for all of this?
- **What are the data really about**?
- **What are the data suitable (or unsuitable) for**?
- **Who can reuse** the data?
- What specific data support your results?
- How can they be used to **repeat your experiments, etc.**?

# Data, Managed Data, FAIR Data, Open Data...



# Research Data Levels



Source: FAIR příručka pro data steward komunitu v ČR, <https://doi.org/10.71495/hxfc-6f57>

- FAIR principles:

- Findable
- Accessible
- Interoperable
- Reusable

# 5-Stars Deployment Scheme for Open Data

★ Make your stuff available on the Web (**whatever format**) under an **open license**.

★★ Make it available as **structured data** (e.g., Excel instead of image scan of a table).

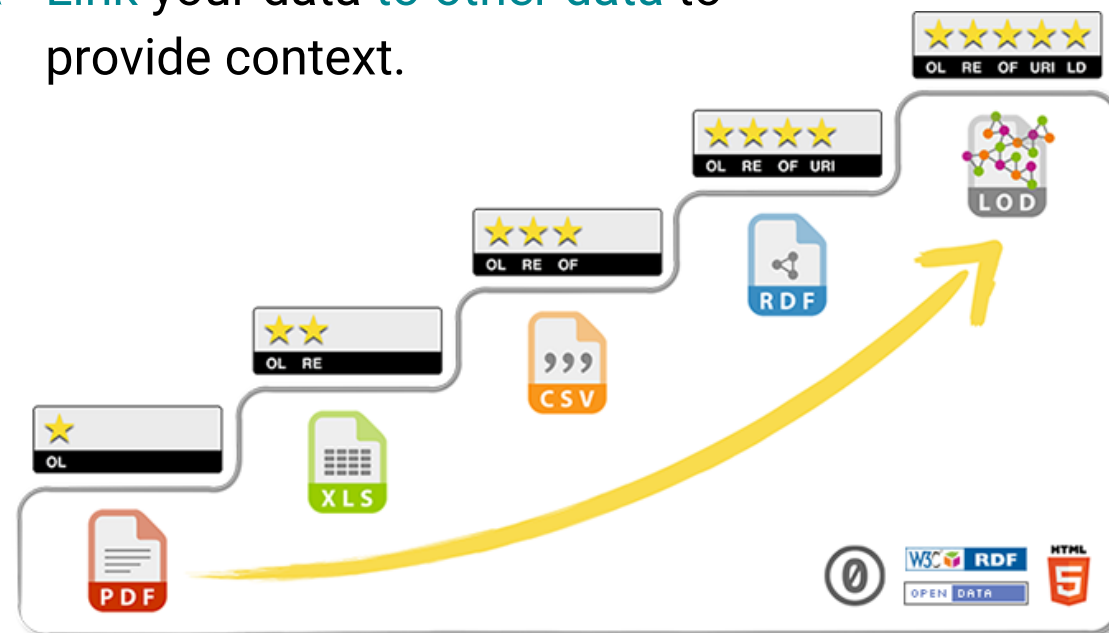
The dataset is provided in a **machine-readable format** that allows **automated machine processing**.

★★★ Make it available in a **non-proprietary open format** (e.g., CSV instead of Excel).

★★★★ Use **URIs** to denote things, so that people can point at your stuff.

★★★★★ Link your data **to other data** to provide context.

In 2010, Tim Berners-Lee published a system for assessing the openness of linked data, the so-called **5-Star Linked Open Data**.



Source: FAIR příručka pro data steward komunitu v ČR, <https://doi.org/10.71495/hxfc-6f57>

# FAIR Principles

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

- FAIR is an abbreviation of the four English adjectives denoting these principles.
- **Findable**
  - For **people and machines**.
  - We need (**machine-readable**) **metadata** that is available to **search engines/users**.
  - We need **unique persistent identifiers** (PIDs).
- **Accessible**
  - Technical solution for **access to data and metadata** – use **standard open protocols** (e.g. **https://**) common in the field.
  - The data does not have to be freely accessible – then the **authentication mechanisms** should be **common and standard** (e.g. **OpenID Connect**) in the field.
  - If the data itself is inaccessible (not public or has been deleted), **at least the metadata should be freely available** (for Findable and audit-record of the deleted data).
- **Interoperable**
  - **Similar to Accessible**, but not about access protocols, **but about the data itself**.
  - Data and metadata in **machine-readable open formats** common in the field.
  - **Other** data/metadata should be **referenced** using standard persistent identifiers.
- **Reusable**
  - Rather the goal we are aiming at in the previous points, rather than a separate principle.
  - **Clear declaration of license**, use of open licenses (e.g. CC0, CC-BY).
  - Data and metadata in **industry-standard and usable formats**.
  - **Provenance metadata** – a relatively new area dealing with the metadata record of all operations with data from their inception to the present.

# FIP Mini-Questionnaire

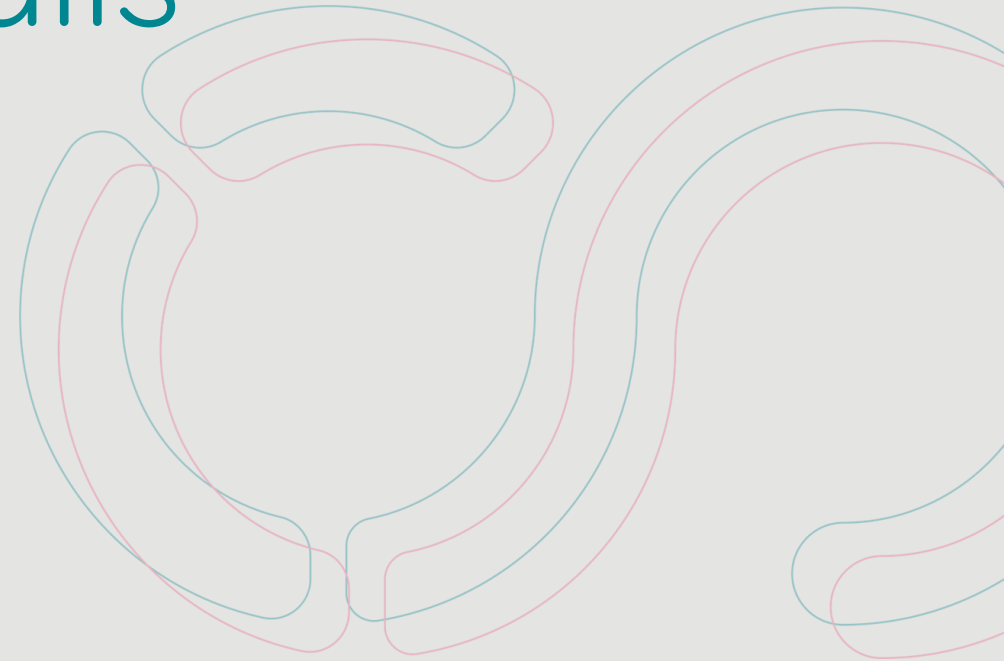
Building your FAIR implementation profile.

<https://www.go-fair.org/how-to-go-fair/fair-implementation-profile/>

Community description	
Name of Community	e.g. ENVRI
Description of Community	
Supporting Links	
Research Domain	e.g. Environmental Sciences
Data Steward	e.g. ORCID #
Date of FIP creation	

FAIR principle	Question	FAIR enabling resource types	Your answers
F1	What globally unique, persistent, resolvable identifiers do you use for metadata records?	Identifier type	e.g. PURL, DOI
F1	What globally unique, persistent, resolvable identifiers do you use for datasets?	Identifier type	
F2	Which metadata schemas do you use for findability?	Metadata schema	
F3	What is the technology that links the persistent identifiers of your data to the metadata description?	Metadata-Data linking mechanism	
F4	In which search engines are your metadata records indexed?	Search engines	
F4	In which search engines are your datasets indexed?	Search engines	
A1.1	Which standardized communication protocol do you use for metadata records?	Communication protocol	
A1.1	Which standardized communication protocol do you use for datasets?	Communication protocol	
A1.2	Which authentication & authorisation technique do you use for metadata records?	Authentication & authorisation technique	
A1.2	Which authentication & authorisation technique do you use for datasets?	Authentication & authorisation technique	
A2	Which metadata longevity plan do you use?	Metadata longevity	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for metadata records?	Knowledge representation language	
I1	Which knowledge representation languages (allowing machine interoperation) do you use for datasets?	Knowledge representation language	
I2	Which structured vocabularies do you use to annotate your metadata records?	Structured vocabularies	
I2	Which structured vocabularies do you use to encode your datasets?	Structured vocabularies	
I3	Which models, schema(s) do you use for your metadata records?	Metadata schema	
I3	Which models, schema(s) do you use for your datasets?	Data schema	
R1.1	Which usage license do you use for your metadata records?	Data usage license	
R1.1	Which usage license do you use for your datasets?	Data usage license	
R1.2	Which metadata schemas do you use for describing the provenance of your metadata records?	Provenance model	
R1.2	Which metadata schemas do you use for describing the provenance of your datasets?	Provenance model	

# Data Management Plan (DMP) in Research Project Calls



# Open Science and Projects

- Requirements for **compliance with Open Science principles** have already been **established as a standard** part of **research project calls**.
- Often mandatory and optional principles:
  - **Mandatory**: Open Access, Data Management (DMP)
  - **Optional**: Citizen Science, Open Source, Preprints, ...
- **DMP** is a **living** document: Regular **updates** are expected.

# DMP and Horizon Europe

- DMP is mandatory.
- DMP is a living document with regular updates:
  - Initial DMP.
  - Updates during the project.
  - Final DMP at the end of the project.
- *Horizon Europe Open Science Requirements in Practice*
  - OpenAIRE webinar, 2023-03-13
  - Slides available at:  
<https://doi.org/10.5281/zenodo.7324363>

# DMP and GAČR

- DMP is mandatory according to this year's project agreements.
  - Beware, including projects already started from 2024-01-01.
  - The requirement is included in all GAČR project calls announced for following years.
- GAČR, *STANDARD PROJECTS – 2025, Terms and Conditions for the Public Tender and Grant Projects in Research, Experimental Development, and Innovation*, available at: [https://gacr.cz/wp-content/uploads/2024/02/TD\\_STD\\_2025.pdf](https://gacr.cz/wp-content/uploads/2024/02/TD_STD_2025.pdf).

Page 64:

[...] The Beneficiary agrees to submit a Data Management Plan (DMP) on or before the date of the submission of the first Interim Report and to update the DMP regularly, as needed, and to submit it regularly as part of the Interim and Final Reports, including information on the availability and dissemination of the research results and research data. The updated research data management plan shall be included in the Interim and Final Reports. The research data management plan shall include, as a minimum, information on: (i) what kind of research data will be generated, processed or collected by the Beneficiary within the project; (ii) what methods and principles of data management will be used; (iii) whether and how the data will be shared, published and/or disclosed, and, where appropriate, an explanation of why the research data cannot be disseminated or disclosed (the Beneficiary shall not disclose information on research data the disclosure of which would unduly interfere with the intellectual property rights, trade secrets, national security and/or legitimate commercial interests of the Beneficiary or those of a third party); (iv) and how the data will be stored during the project and preserved after the project. [...]

# DMP and GAČR

- Research **data management** has to be **described** in the scientific research **project proposal**.
- DMP is a life document with regular updates.
  - Each Interim Report.
  - With Final Report.
- GAČR, *STANDARD PROJECTS – 2025, Terms and Conditions for the Public Tender and Grant Projects in Research, Experimental Development, and Innovation*, available at:  
[https://gacr.cz/wp-content/uploads/2024/02/TD\\_STD\\_2025.pdf](https://gacr.cz/wp-content/uploads/2024/02/TD_STD_2025.pdf).

Page 19:

*[...] a brief description of the research data to be generated, used, and stored in the course of the Project, and how these data will be handled; in particular, information on the availability and dissemination of the research results and research data, in accordance with the principle that research results and research data are not made public only where justified (an update shall be provided with each Interim Report and Final Report); the Beneficiary agrees to submit a Data Management Plan (DMP) no later than the date of delivery of the first Interim Report, and to update the DMP periodically as needed; [...]*

# DMP and Applied Health Research Projects by the Ministry of Health of the Czech Republic

- DMP is mandatory for projects started from 2025-05-01.
  - Continuous maintenance and regular updates are required.
- Compliance with FAIR principles is expected.
- Ministry of Health of the Czech Republic,  
*Tender Documentation for Public Tender no. 2 in Experimental Development and Innovation in the Applied Health Research Support program for the Years 2024–2030*, available at: <https://mzd.gov.cz/wp-content/uploads/2024/05/ZD-VES-2025-EN-1.pdf>.

Page 24:

[...] A brief description of the **research data** that will be *used, collected or generated* during the Project and how it will be handled; the Beneficiary is *required* to have a *Data Management Plan (DMP)* in place at the latest *at the time of submission* of the *first sub-report* and to *update* the DMP *on a regular basis* if necessary *and to submit it on request*; the DMP should *include*, among others, information on what *methods and principles of data management* will be used with *respect to the FAIR principles* (findability, accessibility, interoperability and reusability); [...]

# DMP and INTER-EXCELLENCE II by the Ministry of Education, Youth and Sports of the Czech Republic

- DMP is mandatory.
  - Continuous maintenance and regular updates are required.
- Ministry of Education, Youth and Sports,  
*Model contract for granting support, INTER-ACTION sub-programme, INTER-EXCELLENCE II programme*, available at:  
[https://www.msmt.cz/file/62614\\_1\\_1/](https://www.msmt.cz/file/62614_1_1/).

Page 6:

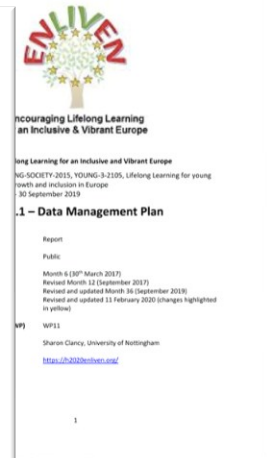
*[...] develop and submit a Data Management Plan to the Provider in the Interim Project Report; furthermore, update it regularly during the course of the Project and submit an updated version as part of each subsequent Project Report. [...]*

# ENLIVEN ('Encouraging Lifelong Learning for an Inclusive and Vibrant Europe') Data Management Plan

- <https://hdl.handle.net/11353/10.1139743>
- DMP from the area of **Social Sciences**.
  - Audio and video data, transcription of interviews.
  - **Restricted access** to the data.
- Use of **existing** data + collection of **own** data.
- **Continuous updating** of the DMP.
  - Changes are summarized in a table.
  - Migration of some of the data to the UK.
  - Data saved encrypted in MS 0365 Teams.
- **Ethic** aspects are described in a **standalone document** referenced from the DMP.
- Joined project of multiple institutions → explicit **definition of responsibilities**.
- Nice **description of data protection**.
  - Information on anonymization procedures in the DMP appendix.
- Documentation and produced **publications** (including project website) are **covered in the DMP**.
- Description of the used hardware and software could be more detailed.
  - Nevertheless, the backup process and strategy are described.
- Described **intellectual properties** and **QA process**.

## Table of Contents

1	Changes to Data Management Plan (D11.1) – September 2019.....	3
2	Introduction .....	6
2.1	Project Summary.....	6
2.2	Types of Data .....	6
2.3	Organisation of the ENLIVEN project.....	7
3	Data Management.....	8
3.1	Principles .....	8
3.2	Organisation .....	9
3.3	Types of Data .....	9
3.3.a	Aggregated secondary statistical data .....	9
3.3.b	Secondary data in form of anonymised micro data sets: .....	9
3.3.c	Primary data collection .....	9
3.4	Data Use and Protection.....	10
3.4.a	Secondary analysis of EUROSTAT and related scientific-use micro data files .....	10
3.4.b	Interview Transcription and Data Protection .....	11
3.5	Documentation .....	12
3.5.a	Documentation Reports.....	12
3.5.b	Project Management Documents .....	13
3.5.c	Website .....	13
3.6	Hardware and Software.....	13
3.6.a	Intelligent Decision Support System .....	13
3.6.b	Data Backup and Recovery .....	13
3.7	Intellectual Property and Ownership .....	13
3.7.a	Intellectual Property .....	13
3.7.b	Joint ownership .....	14
3.8	Open Access.....	14
3.9	Quality Assurance .....	14
4	Work-Package-specific Data Issues.....	15
4.1	WP1: Mapping European and national policies and programmes, and their contribution to economic and social inclusion .....	15
4.2	WP2: Constraints and facilitators of access and participation .....	15
4.3	WP3: The role of European governance in adult education & learning policy .....	15
4.4	WP4: Improving our understanding of the effect of system characteristics by building stronger data and adding a longitudinal, regional & sectoral focus .....	16
4.5	WPs5-7: Studying the role of workplace learning and patterns of work organisations for early career structuration; qualitative interviews on learning biographies .....	16
4.6	WP8: Knowledge discovery on evidence-based policy making in participating countries; & WP9: Establishment of Intelligent Decision Support System for evidence-based policy making .....	16
4.7	WP10-11: Dissemination and Project Management & Integration.....	17
5	Appendices.....	18
5.1	Table 1: Key elements of the framework to ensure anonymization within the ENLIVEN research process (for storage/use within the project): (to be refined within the research project) .....	18
5.2	Table 2: Processing of data in the qualitative research implemented by the ENLIVEN project.....	19



# Where to Store Research Data

## Practical Examples



# Example: Storages at MU and Recommendations for their Use

- Centrally managed by Institute of Computer Science
  - <https://it.muni.cz/en/categories/data-storage>
- Recommendations their use:
  - Data type categorization.
  - Storage technology categorization.
  - Matrix of suitability of different storages for different types of data.

# Example: Storages at MU – Suitability of storage for different types of data

STORAGE TYPE	USAGE			
	GREEN: PUBLIC DATA	BLUE: INTERNAL DATA	ORANGE: DISCRETE DATA	RED: SENSITIVE DATA
PORTABLE MEDIA (FLASH DRIVES, EXTERNAL HDD, CD, DVD, ...)	Appropriate	Possible Encryption recommended	Inappropriate Possible when using encryption	Inappropriate
LOCAL STORAGE				
IN COMPUTERS (DESKTOP, LAPTOPS)	Appropriate	Appropriate	Appropriate Encryption recommended	Inappropriate possible in well-justified cases, when performing an individual analysis, using encryption and applying other security measures resulting from the analysis
IN MOBILE DEVICES (MOBILE PHONES, TABLETS, ...)	Appropriate	Appropriate Screen lock required (pattern, fingerprint reader, PIN, password)	Possible Encryption required Strong screen lock required (fingerprint reader, PIN, password)	Inappropriate possible in well-justified cases, when performing an individual analysis, using encryption and applying other security measures resulting from the analysis
ICS NETWORK AND CLOUD STORAGE (SO-CALLED STANDARD AND MEDIUM NETWORK STORAGE, SEE IT CATALOGUE, CERIT-SC STORAGE)	Appropriate	Appropriate	Appropriate	Appropriate It is recommended to perform an individual analysis, use encryption and apply other security measures that result from the analysis

# Example: Storages at MU – Suitability of storage for different types of data

STORAGE TYPE	USAGE			
	GREEN: PUBLIC DATA	BLUE: INTERNAL DATA	ORANGE: DISCRETE DATA	RED: SENSITIVE DATA
IS MUNI REPOSITORY (E.G. DOCUMENT SERVER, FILE DEPOSITORY, ETC.)	Appropriate	Appropriate	Appropriate	Appropriate It is recommended to perform an individual analysis, use encryption and apply other security measures that result from the analysis
CESNET STORAGE (E.G. CESNET ARCHIVE STORAGE, OWNCLOUD, FILESENDER, ..., SEE CESNET DATA STORAGE DEPARTMENT)	Appropriate	Appropriate	Appropriate	Appropriate It is recommended to perform an individual analysis, use encryption and apply other security measures that result from the analysis
EXTERNAL STORAGE				
WITH A CONTRACT WITH MUNI				
MUNI MICROSOFT O365 (MUNI O365 ONEDRIVE, SHAREPOINT, ..., VIZ MUNI O365)	Appropriate	Appropriate	Appropriate Encryption recommended	Possible only with adequate procedural coverage of the situation based on an individual analysis and the application of security measures that result from the analysis
MUNI GOOGLE G SUITE FOR EDUCATION (SEE MUNI GOOGLE APPS)	Appropriate	Appropriate	Inappropriate Possible when using encryption	Inappropriate
GRAMMARLY	Appropriate	Appropriate	Inappropriate	Inappropriate
WITHOUT A CONTRACT WITH MUNI				
PUBLIC GOOGLE, MICROSOFT, DROPBOX, ... STORAGES	Appropriate	Inappropriate	Inappropriate	Inappropriate

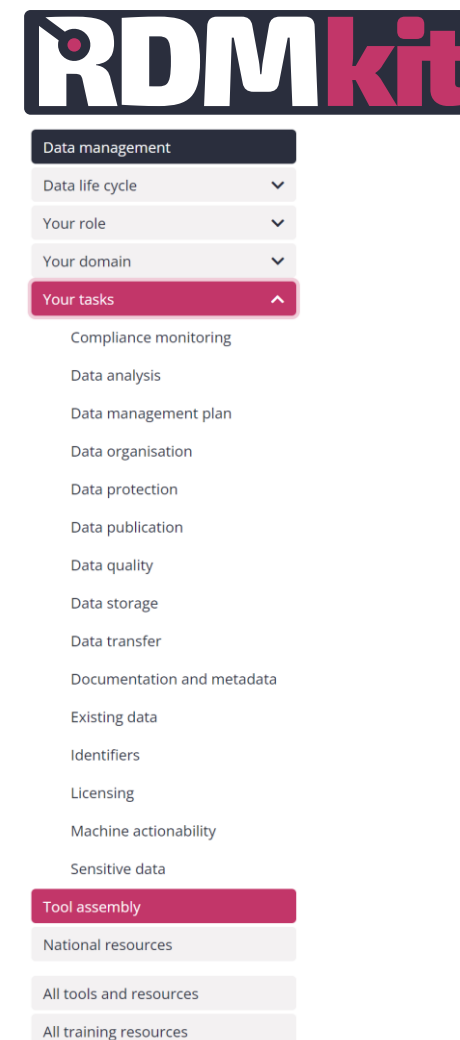
# Tools and Resources to Support FAIR Principles



# Tools and Resources to Support FAIR Principles

## The ELIXIR Research Data Management Kit (RDMkit)

- <https://rdmkit.elixir-europe.org/>
- Online guide to data management in the life sciences.
- Links to several external sources.



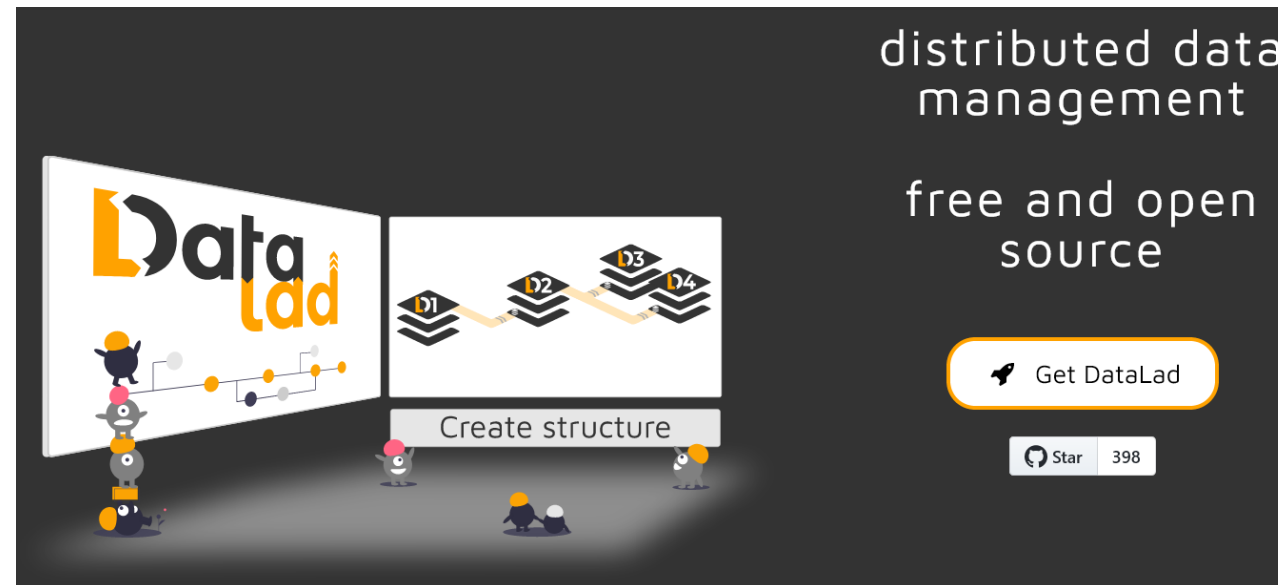
# JupyterLab

- # Processing to enable sharing.

# Tools and Resources to Support FAIR Principles

## DataLad

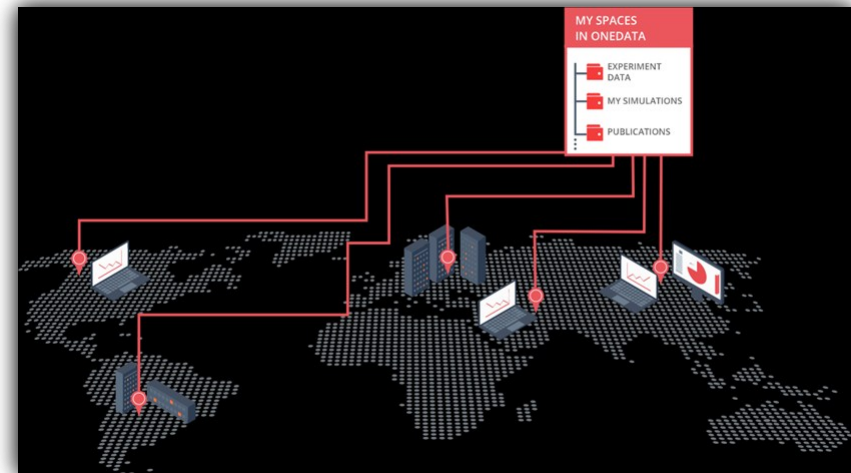
- <https://www.datalad.org/>
- *DataLad is a free and open-source distributed **data management** system that **keeps track** of your data, creates structure, **ensures reproducibility**, supports collaboration, and integrates with widely used data infrastructure.*



# Tools and Resources to Support FAIR Principles

## OneData

- <https://onedata.org/>
- Perform *heavy computations* on *huge datasets*. Access your data in a *dropbox-like fashion* *regardless of its location*. Publish and *share* your results with *public* or *closed communities*.



# Persistent Identifiers (PIDs)



# Persistent Identifiers (PIDs)

- They are supposed to ensure the separation of the identification of **the object as such**,
  - person,
  - institution,
  - publication,
  - dataset,
- **from** its current **physical location**.
- **Example** – dataset *https-set*
  - Dataset **identifier**: <https://doi.org/10.48791/4mxp-r725>
  - Current physical **location**:  
<https://ucnmuni.sharepoint.com/teams/mu-UVT-https-set/Shared%20Documents/Forms/AllItems.aspx?id=%2Fteams%2Fmu%2DUVT%2Dhttps%2Dset%2FShared%20Documents%2Fhttps%2Dset%2Dv1%2E0%2E0&p=true&ga=1>
  - The physical location **is likely to change in the future** – a move to the National catch-all data repository, is being considered.
  - Changes don't matter – **users** are *always* presented with a link to the data being [DOI 10.48791/4mxp-r725](https://doi.org/10.48791/4mxp-r725), which always **takes them** to the **current location**.

# Persistent Identifiers (PIDs)

- They are intended to ensure **clarity**.
- **Example** – names of natural persons:
  - Multiple forms of writing the name of one physical person.
  - Multiple different individuals with the same name.
- They are intended to ensure **persistence**.
  - Metadata physically located with a third party, independent of the physical location of the referenced entity.
- The identifier's owner takes care of updating the metadata and updating the route to the current location.
- The third party is responsible for preserving the latest version and history of the metadata and the existence of the identifier itself, even if the owner of the identifier stops caring for it. And even if the identified entity is not retained as such.

<input type="checkbox"/>	<a href="#">NovakD (1)</a>	Novák, David (1)	<a href="#">Join</a>	<a href="#">Delete</a>
<input type="checkbox"/>	<a href="#">NovakE (5)</a>	Nováková, Eva (5) Nováková, E. (0)	<a href="#">Join</a>	<a href="#">Delete</a>
<input type="checkbox"/>	<a href="#">NovakJ (52)</a>	Novák, Josef (38) Novák, Jos. (1) Novák, J. (13)	<a href="#">Join</a>	<a href="#">Delete</a>
<input type="checkbox"/>	<a href="#">NovakJ2 (7)</a>	Novák, Jiří (7) Novak, Jiri (0)	<a href="#">Join</a>	<a href="#">Delete</a>
<input type="checkbox"/>	<a href="#">NovakJ7 (19)</a>	Novák, Josef (16) Novák, J. (3)	<a href="#">Join</a>	<a href="#">Delete</a>
<input type="checkbox"/>	<a href="#">NovakK (2)</a>	Novák, Karel (2)	<a href="#">Join</a>	<a href="#">Delete</a>
<input type="checkbox"/>	<a href="#">NovakM (2)</a>	Novák, Mirko (2) Novak, Miroslav M. (0) Novak, M. M. (0) Novák, M. (0)	<a href="#">Join</a>	<a href="#">Delete</a>
<input type="checkbox"/>	<a href="#">NovakM2 (2)</a>	Nováková, Markéta (2)	<a href="#">Join</a>	<a href="#">Delete</a>
<input type="checkbox"/>	<a href="#">NovakM3 (1)</a>	Novák, Miroslav (1)	<a href="#">Join</a>	<a href="#">Delete</a>
<input type="checkbox"/>	<a href="#">NovakO (2)</a>	Novák, Ondřej (2)	<a href="#">Join</a>	<a href="#">Delete</a>
<input type="checkbox"/>	<a href="#">NovakP (1)</a>	Novák, Petr (1)	<a href="#">Join</a>	<a href="#">Delete</a>
<input type="checkbox"/>	<a href="#">NovakS (1)</a>	Novák, Stanislav (1)	<a href="#">Join</a>	<a href="#">Delete</a>
<input type="checkbox"/>	<a href="#">NovakV (57)</a>	Novák, Vítězslav (55) Novák, V. (2)	<a href="#">Join</a>	<a href="#">Delete</a>
<input type="checkbox"/>	<a href="#">NovakV2 (7)</a>	Novák, Vilém (7) Novák, V. (0)	<a href="#">Join</a>	<a href="#">Delete</a>
<input type="checkbox"/>	<a href="#">NovakV3 (53)</a>	Novák, Vladimír (53)	<a href="#">Join</a>	<a href="#">Delete</a>
<input type="checkbox"/>	<a href="#">NovakZ (1)</a>	Novák, Zdeněk (1)	<a href="#">Join</a>	<a href="#">Delete</a>

Source: Authors' database of the project [DML-CZ](#)

# Popular PIDs

- People

- **ORCID:** <https://orcid.org/>
  - Example: [0000-0001-6399-5453](https://orcid.org/0000-0001-6399-5453)

- Institutions

- **ROR:** <https://ror.org/>
  - Example: [02j46qs45](https://ror.org/02j46qs45)

- Publications

- **DOI:** <https://www.crossref.org/>
  - Example: [10.5817/CP2022-3-1](https://doi.org/10.5817/CP2022-3-1)

- Datasets

- **DOI:** <https://datacite.org/>
  - Example: [10.48791/4mxp-r725](https://doi.org/10.48791/4mxp-r725)
- **Handle:** <https://handle.net/>
  - Example: [11222.digilib/130328](https://handle.net/11222.digilib/130328)

- Books

- **ISBN:** <https://www.isbn-international.org/>
  - Example: [978-3-16-148410-0](https://www.isbn-international.org/978-3-16-148410-0)

- Journals

- **ISSN:** <http://portal.issn.org/>
  - Example: [0378-5955](http://portal.issn.org/0378-5955)

- Business articles

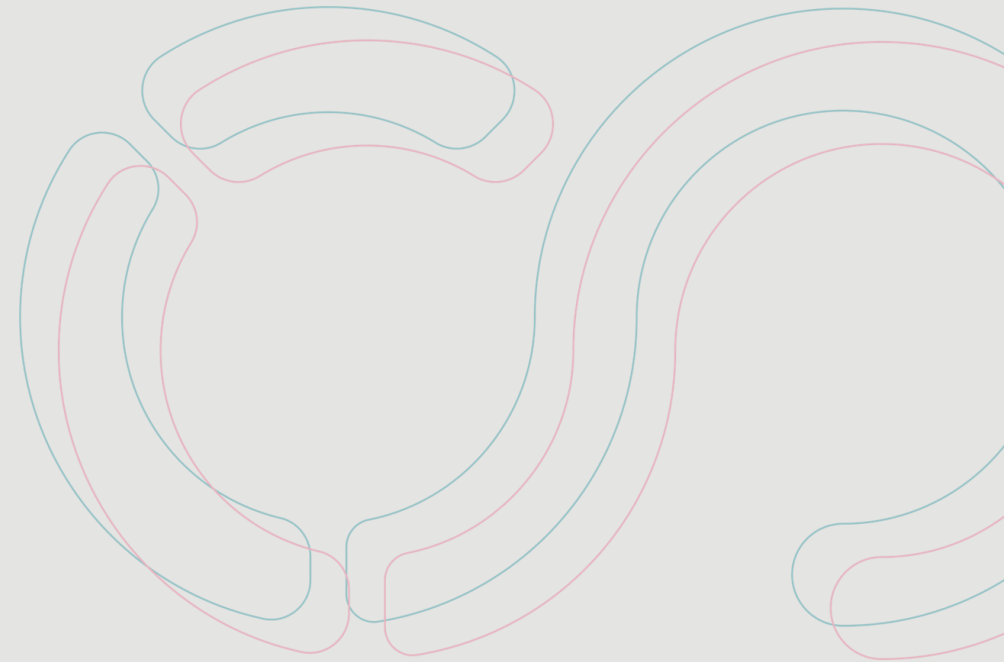
- **EAN13:** <https://www.gs1.org/standards/barcodes/ean-upc>
  - Example: [5901234123457](https://www.gs1.org/standards/barcodes/ean-upc/5901234123457)



- Inhabitants of the Czech Republic

- **Birth Number** (~Social Security Number): <https://www.zakonyprolidi.cz/cs/2004-302/>
  - Example: [736028/5163](https://www.zakonyprolidi.cz/cs/2004-302/736028/5163)

# Repositories for Research Data



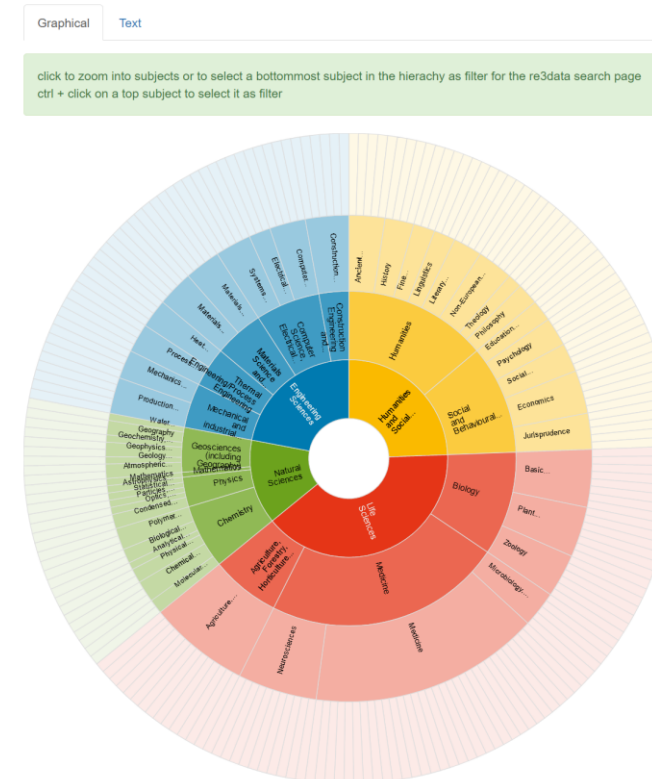
# Types of Repositories

- Topical, national, institutional, catch-the-all, ...
- A common procedure for working with research data is to divide it by disciplines, not by country or institution.
  - We do not consider national/institutional/... specifics, but research topic specifics.
- If we are dealing with a project/institutional/national standard, we must be compatible with international industry practices.
  - Industry standards are addressed by, for example, EOSC Task Forces:  
<https://www.eosc.eu/task-force-faq>

# How to Find/Select a Data Repository

- Repository **directories**:
  - Open Access Repositories: [OpenDOAR](#)
  - Data repositories: [re3data.org](#)
- OpenAIRE: [How to find a trustworthy repository for your data](#)
  - Trusted certified repositories are preferred.
  - [CoreTrustSeal](#) (list of certified repositories).
  - [Nestor Seal](#) (verification according to DIN 31644).
  - [ISO 16363](#).
  - For example, the well-known [Zenodo](#) has no certification...
- The most used **general repositories** include
  - [Zenodo](#),
  - [Figshare](#), or
  - [Dryad](#).

## Browse by subject

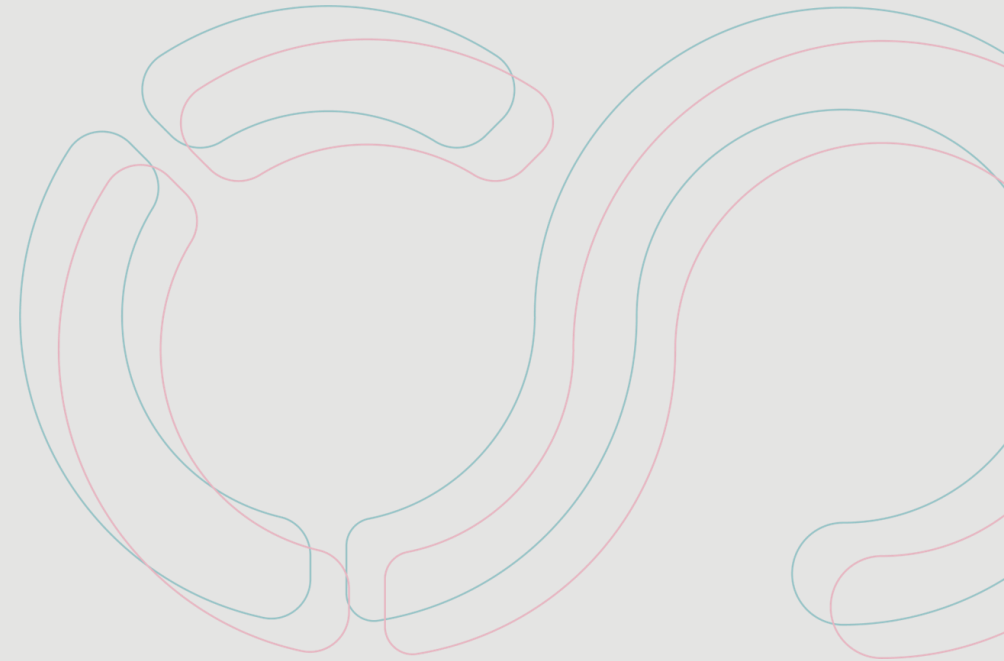


Source: <https://www.re3data.org/browse/by-subject/>

# What to Look Out for when Choosing a Repository

- Will the repository assign a **persistent identifier** (e.g. DOI) to your data?  
Persistent identifiers make your data easier to find and cite.
- Is the repository **trustworthy/certified**?  
With certified repositories, you can be more sure that your data is well taken care of.
- Does the selected repository provide **open access to the stored data**?  
If you want to share your data openly, then this is the key information.
- Will the repository **license** your data? Does it state clear conditions under which the data stored in the repository can be used?  
It's important that users of your data know how they can handle it.
- Will the repository provide a **landing page** for your data?  
Metadata will help others find the data, find out what the data is, and also how to cite it.
- Does the repository allow **versioning**?  
If you update your dataset, you can upload the updated version as a new version to the original dataset. The new dataset is assigned its own persistent identifier and users can easily find out what the latest version is or which version was used in the original study.

# European Open Science Cloud (EOSC)

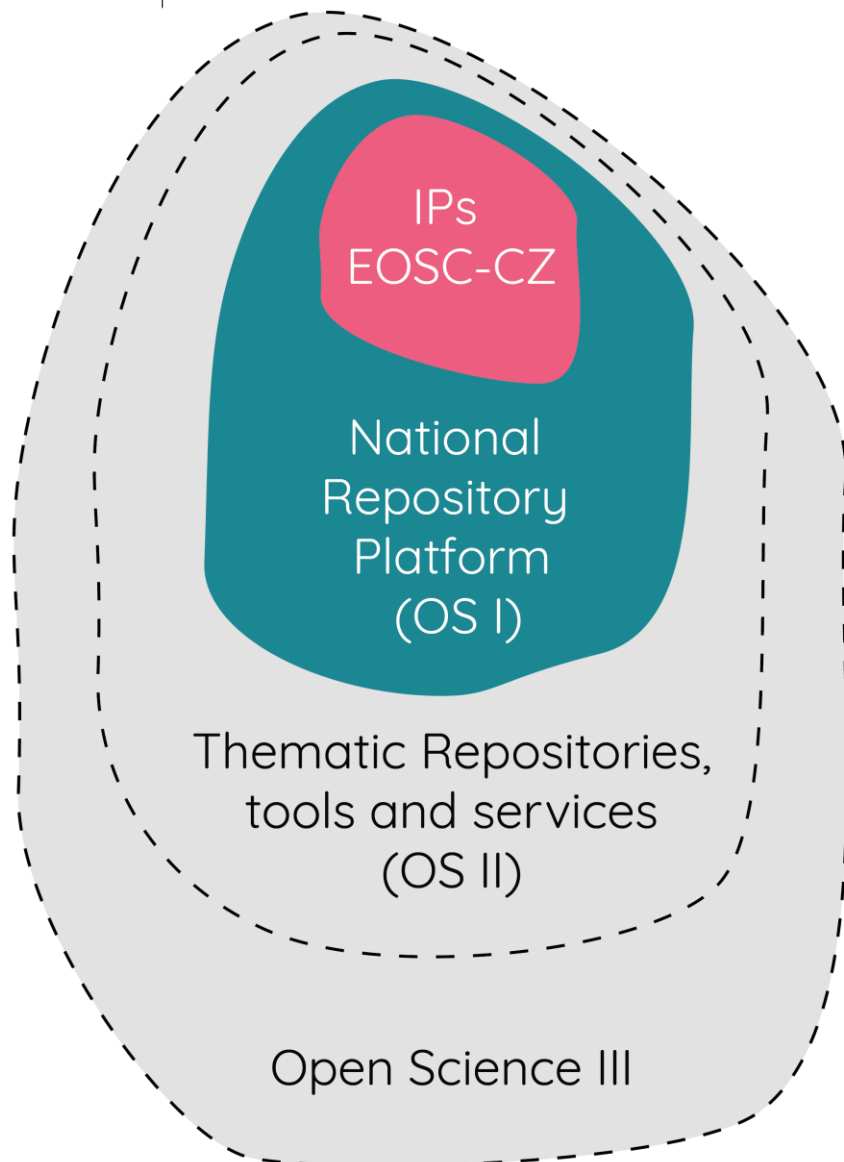


# European Open Science Cloud (EOSC)

- Initiative to **support data management** – the entire data life cycle
  - Since 2016, mainly at the EU level, the search for a suitable setting through a series of projects
  - Currently the concept of EOSC Federation, de facto a large meta-infrastructure
- The key concept is data management
  - Not just any data, but **FAIR data** – i.e. well managed, described, reliably stored and reusable
  - FAIR data is not necessarily Open data – support for data access management is included
- Ecosystem of data and related services
  - Built on existing foundations, i.e. mainly large research infrastructures and e-infrastructures

# EOSC CZ

- Supported through OP JAK (MEYS)
- **Series of projects:**
  - EOSC-CZ, CARDS, NRP, partly Research Environments, OS II (submitted), OS III (planned)
- Related to Data Management Plans (DMPs)
- Supports objectives:
  - Create an environment for working with **FAIR research data**
  - Primarily **Preserve, Share, Reuse**
  - **Plan and collect** is a matter of research
  - **Process and Analyse** are related to (e-)infrastructures

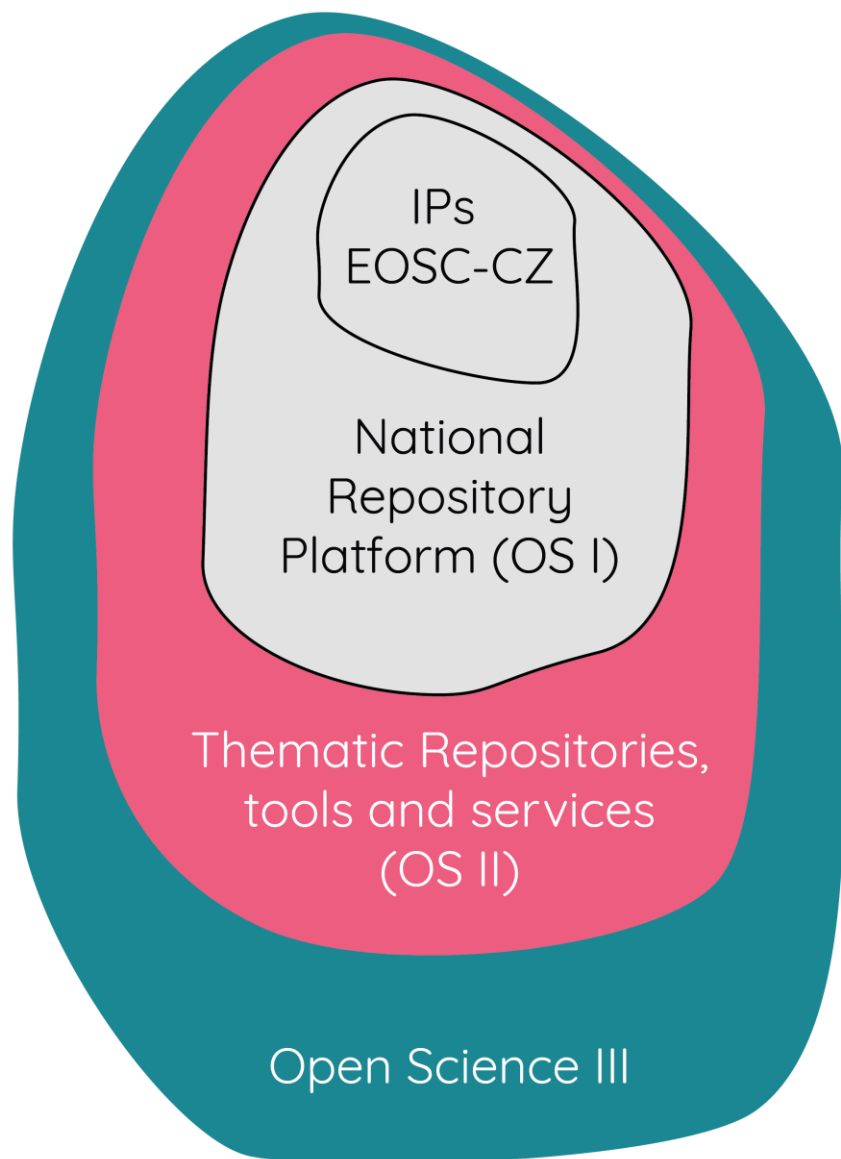


## IPs EOSC-CZ – fundamentals for EOSC implementation in CZ

- Organizational (**Secretariat**) – <https://www.eosc.cz/en/about-eosc-cz/secretariat-eosc-cz>
- Technical (**National Metadata Directory**) – <https://nma.eosc.cz/>
- Educational (**Training Centre**) – <https://www.eosc.cz/en/training>

## National repository platform – „technical core”

- **Repository systems** (DSpace, CESNET Invenio, ASEP-ARL)
- **Pilot repositories**
- **Core services** (PIDs, DSW, licenses, ...)
- **Compliance** and UX (Cybersecurity, ServiceDesk, ...)
- **Training** – technical aspects



## OS II – “domain-specific outputs”

### Based on the inputs from 8 thematic working groups

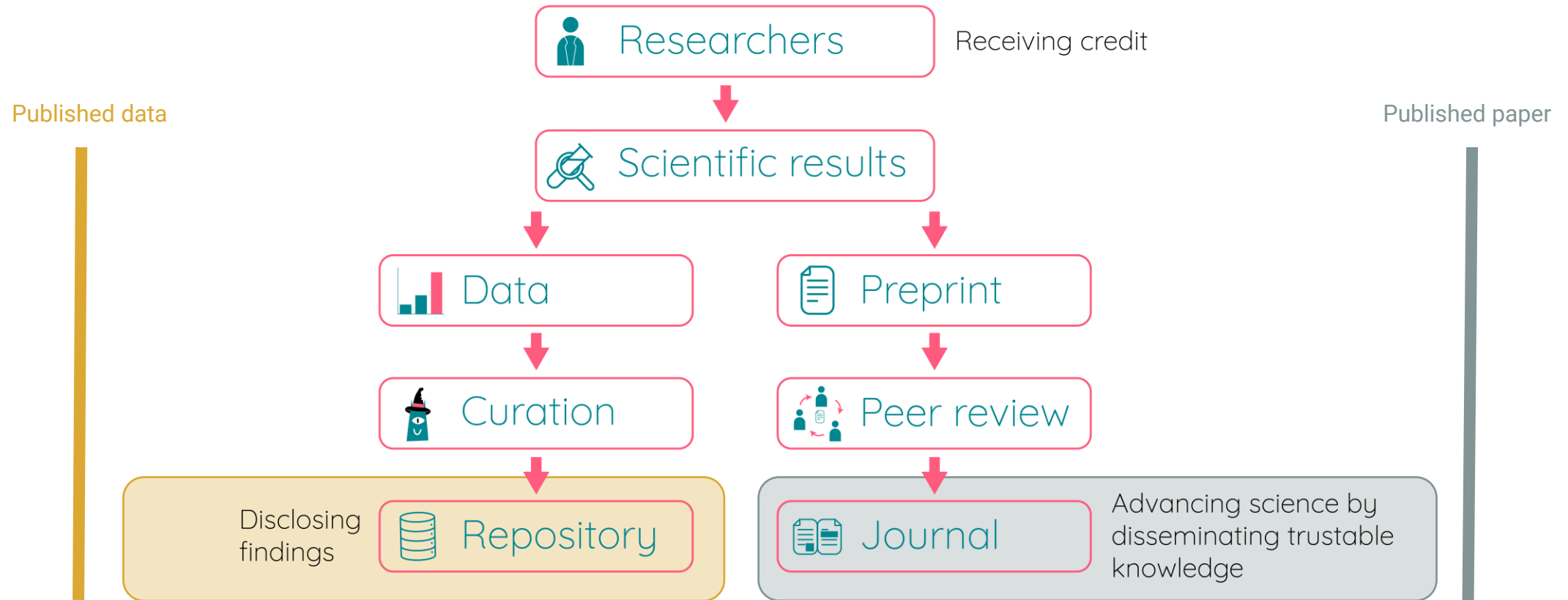
- Domain-specific and interdisciplinary activities
- Development of thematic repositories and tools for FAIR data management
- Bio/Health/Food, Motech, AI & ML, Social Sciences, Physics, Humanities & Arts, Environmental Sciences, Sensitive Data
- <https://www.eosc.cz/en/working-groups>

## OS III – “development of knowledge and skills”

### Development of human resources and competencies in research data management

- Involvement of institutions where the topic of data management is in early stages
- Communication and educational activities
- Establishing institutional support for FAIR data management
- Funded through mini-projects

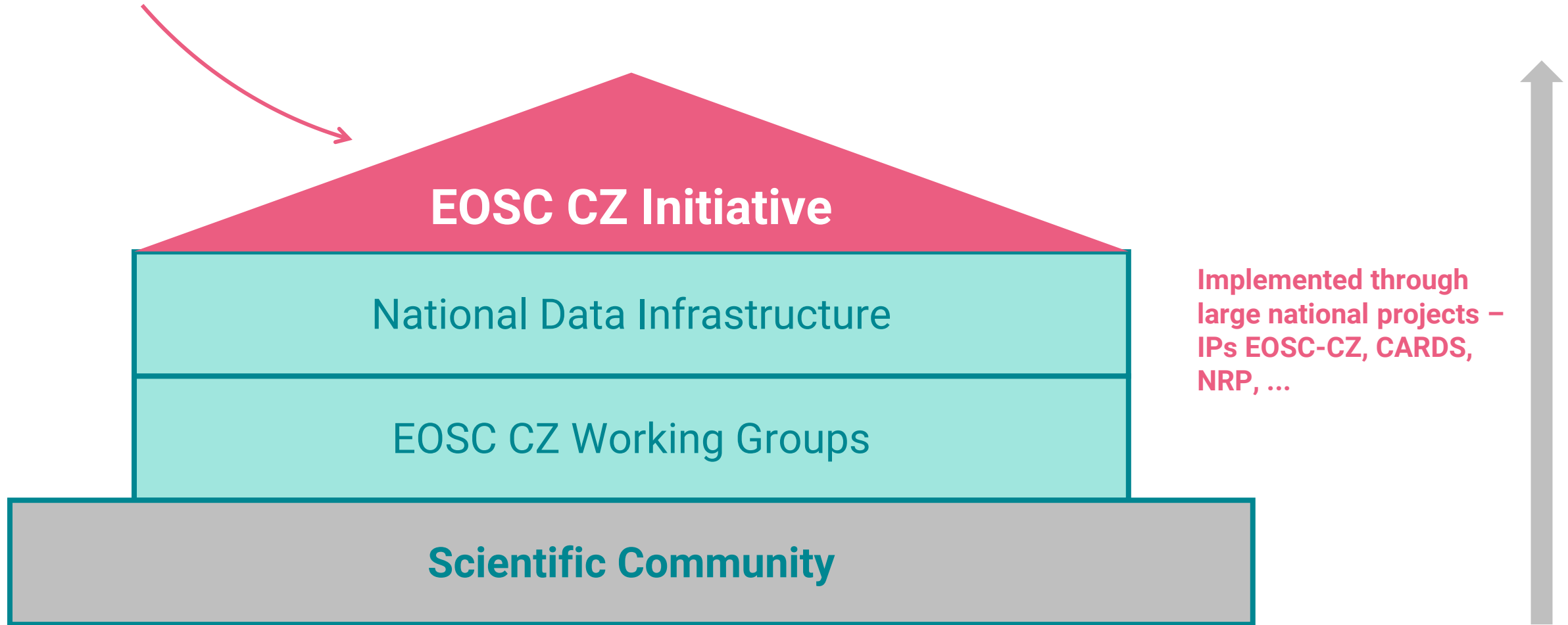
## Peer reviewed research



This **ALSO** is a figurative checkmark in your career

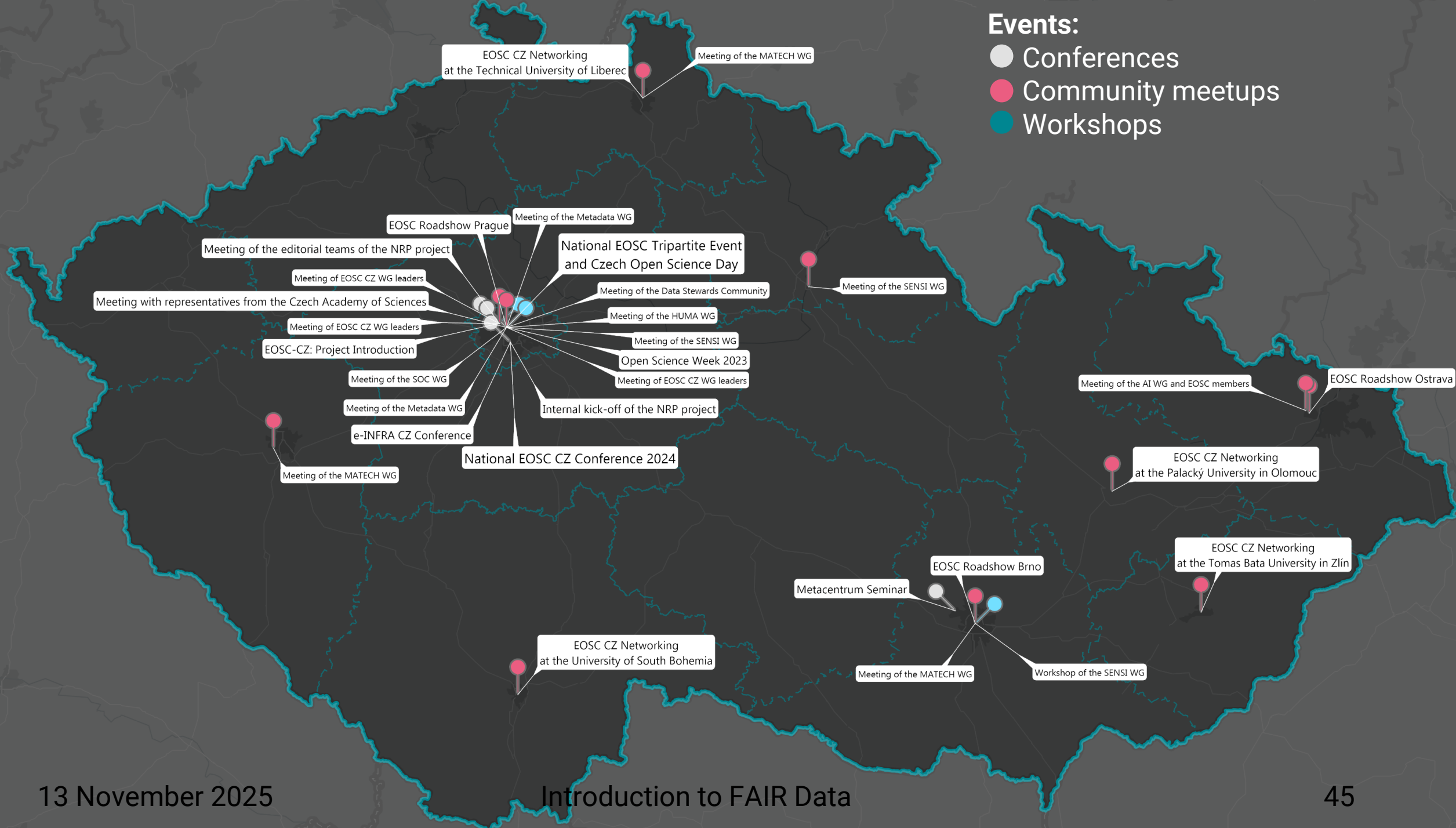
This is a figurative checkmark in your career

# FAIR Support in Czechia

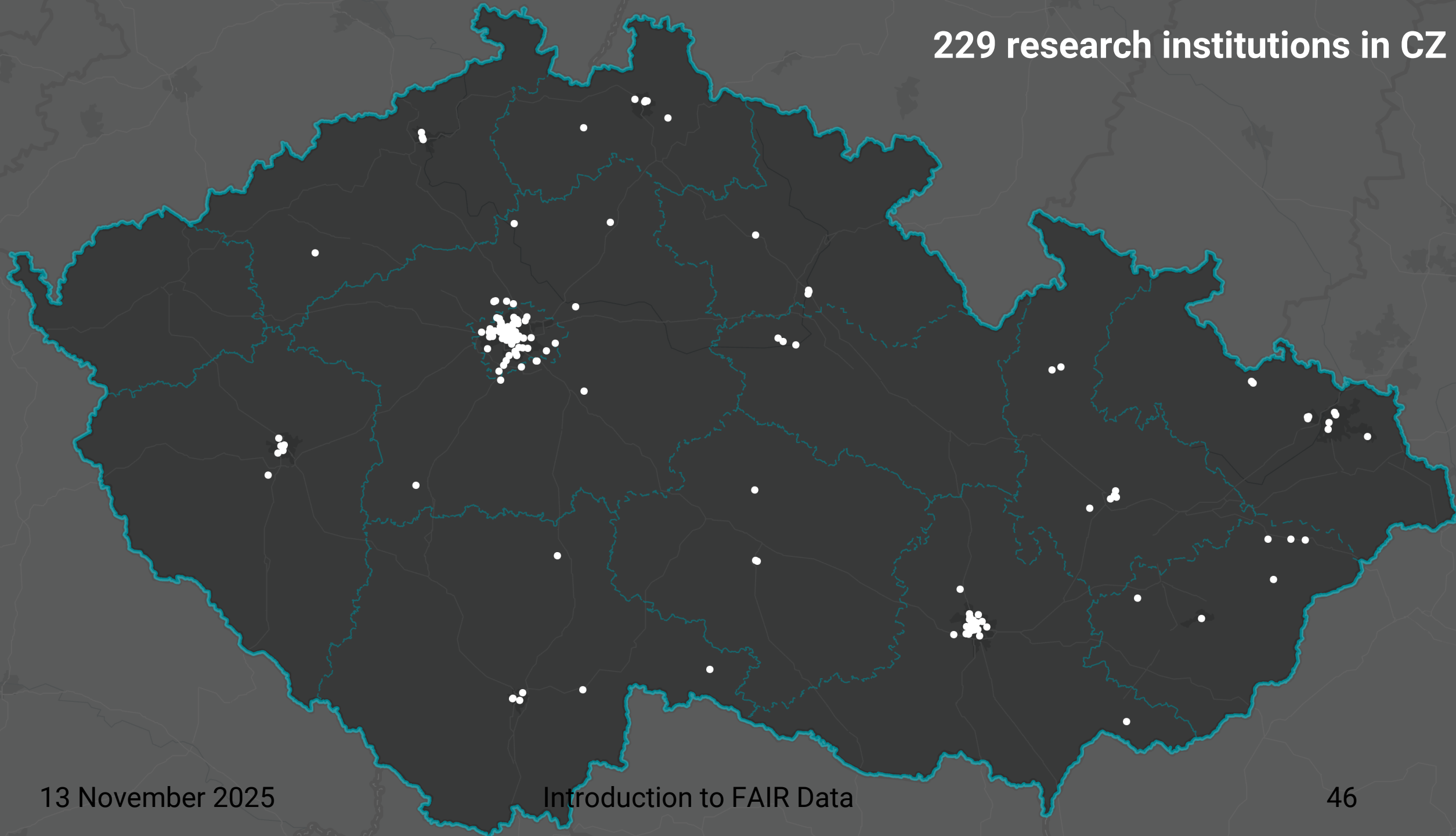


## Events:

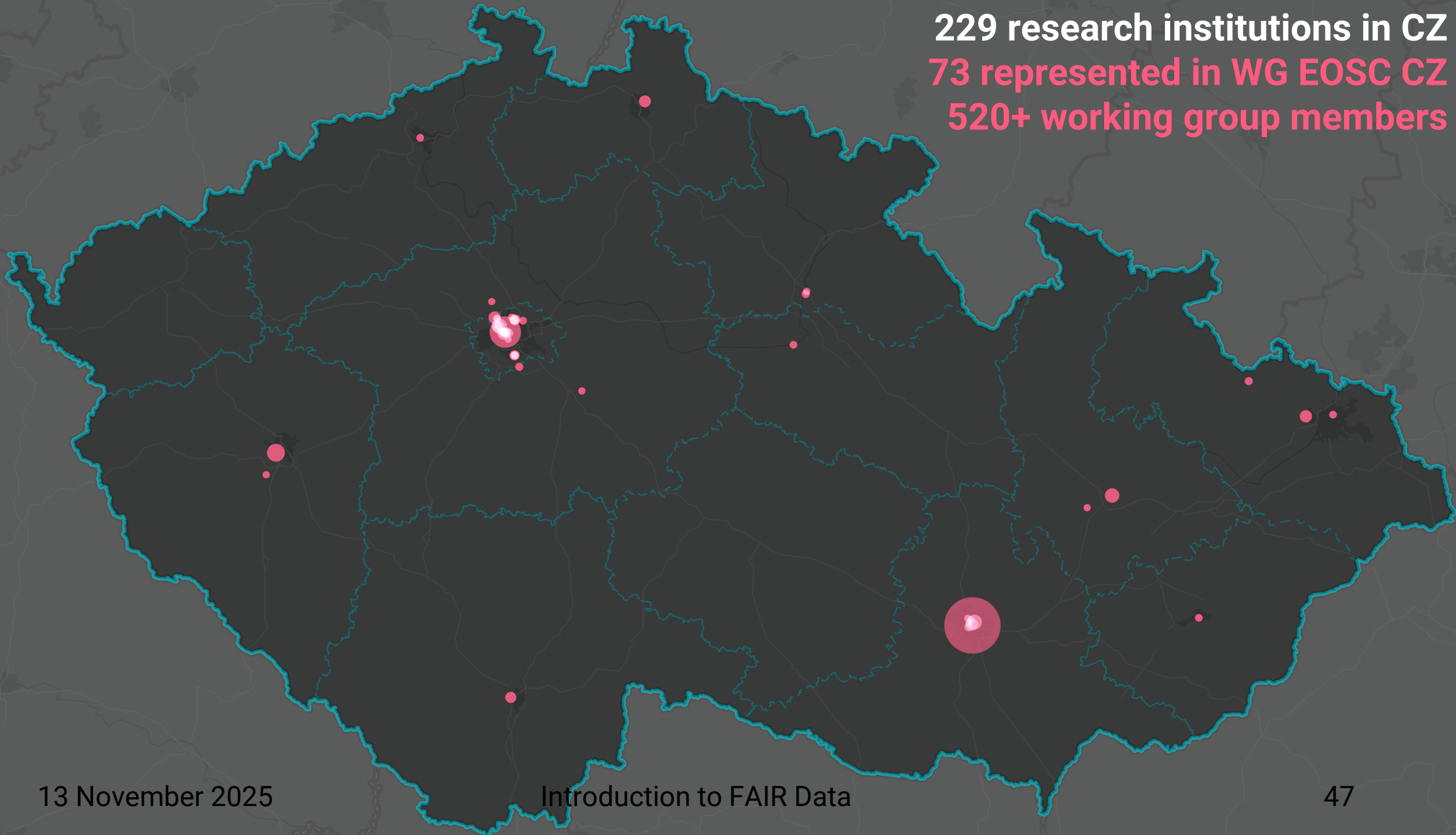
- Conferences
- Community meetups
- Workshops



229 research institutions in CZ



**229 research institutions in CZ**  
**73 represented in WG EOSC CZ**  
**520+ working group members**



# EOSC CZ Working Groups



National Data Infrastructure  
(NDI) Architecture



Metadata



Core Services



Education and Human  
Resources



Bio/Health/Food



Materials Sciences and  
Engineering



Data Management for Artificial  
Intelligence and Machine  
Learning



Social Sciences



Physical Sciences



Humanities and the Arts



Environmental Sciences



Sensitive Data

## Open Platform for EOSC CZ Implementation

- Main EOSC CZ Building Blocks  
(scientists for scientists)
- 4 cross-sectional and 8 field-specific groups
- Open to new members at any time
- Operating throughout the duration of the initiative

# EOSC CZ Working Groups

## How do WGs work?

- Regular meetings
- Notes and recordings
- Open membership on a voluntary basis

## Why get involved?

- Expertise and awareness
- Building connections
- Opportunity to make an impact

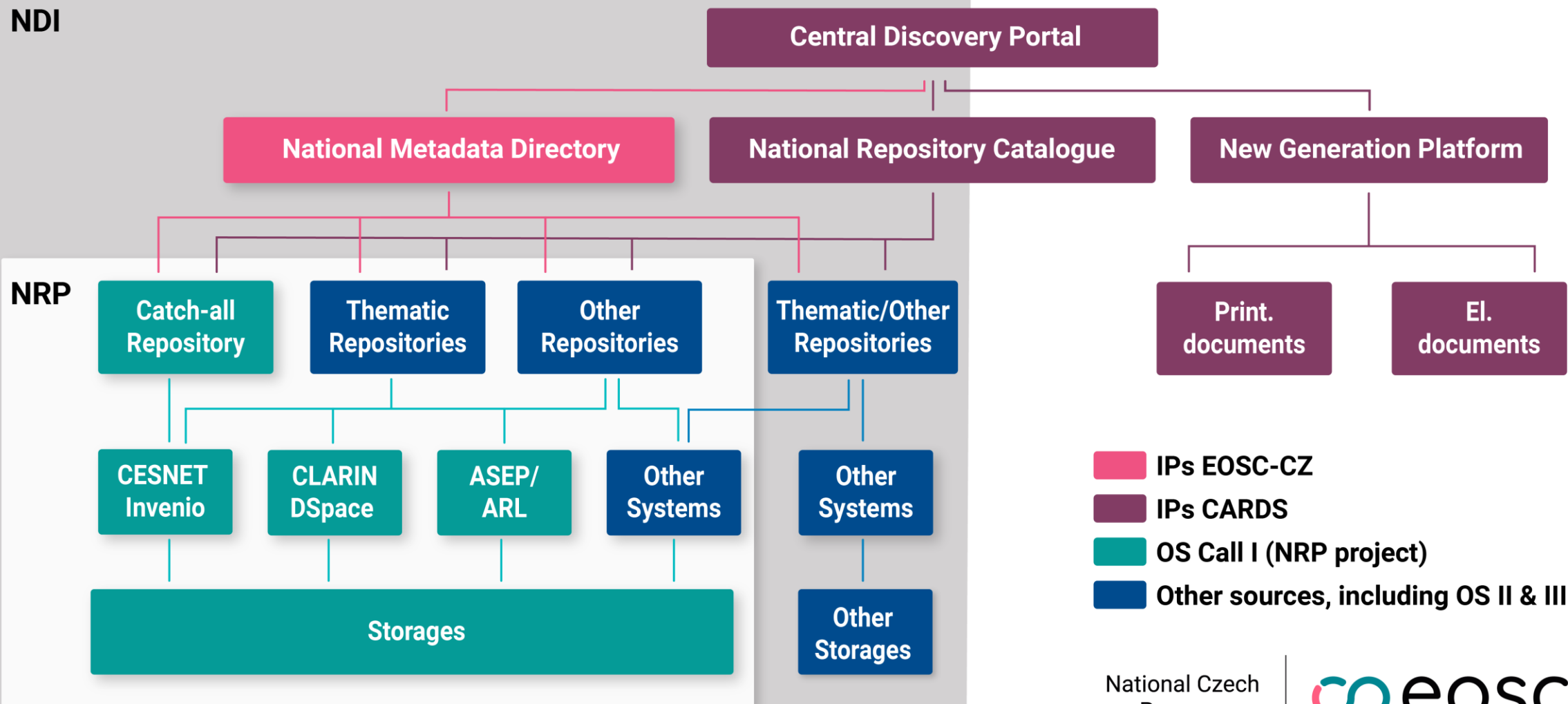
**Are you interested in becoming a member?**



<https://www.eosc.cz/en/working-groups>

# National Data Infrastructure (NDI)

NDI



IPs EOSC-CZ

IPs CARDS

OS Call I (NRP project)

Other sources, including OS II & III

# NDI Outputs: Storage Capacities

## Repositories

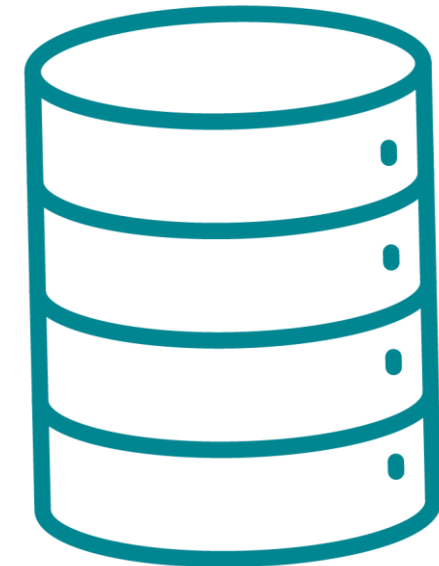
- [Catch-all repository](#) (2025)
- **Thematic (domain-specific) repositories**  
4 pilots: Molecular Biophysics Database\*, National Repository for Biodiversity Data, Repository for Biological Imaging Data, ArchaeoVault (end of 2025)
  - Others from 2025/2026
- [National Metadata Directory](#)

## Repository systems

- CESNET Invenio, CLARIN-DSpace, ASEP-ARL

## Hardware

- Physical, distributed storage infrastructure
- Total of 50+ PB of user data storage capacity




\* in production mode already




# AI-ready (Valuable) Data


[← BACK TO COLLECTION](#)

## Polycaprolactone nanofibers for construction of the alveolar-capillary interface model: Detailed data

**License:**  


**Attachments:**










-  Dataset Cell co-culture scaffolds production.zip
-  Dataset Cell culture and co-culture analysis.zip
-  Dataset Nanofibers production and characterization.zip

**Object identifier:**  
 [10.48700/datst.wmbbb-xhc25](https://doi.org/10.48700/datst.wmbbb-xhc25)

**Record status:**  
Published

**In community:**  
[General community](#)

**Subtitle:** [English](#) Nanofibers production and characterization [English](#) Cell co-culture scaffolds production [English](#) Cell culture and co-culture analysis

**Creators:**  
[Capandova, Michaela](#)  | [Sedlakova, Veronika](#)  | [Vorac, Zbynek](#)  | [Kotasova, Hana](#)  | [Antol, Matej](#)  | [Moran, Lukas](#)  | [Tomáš Bárta](#)  | [Dasa Bohaciakova](#)  | [Ales Hampel](#) 

**Date available:** 2024-11-04

**Dataset creation date:** 2024/2024

**Data collection date:** 2014/2024

**Language:** English

**Publisher:** [Masaryk University](#)

**Keywords:** [en nanofibers](#) [en electrospinning](#) [en polycaprolactone](#) [en tissue engineering](#) [en scaffold](#) [en alveolar-capillary interface](#)

**Subject categories:** [Engineering and technology](#) || [Nano-technology](#) || [Medical and health sciences](#) || [Medical biotechnology](#) || [Nano-materials \(production and properties\)](#) || [Technologies involving the manipulation of cells, tissues, organs or the whole organism \(assisted reproduction\)](#) || [Biomaterials \(as related to medical implants, devices, sensors\)](#)

**Abstract:** [English](#)

This data collection contains the datasets showing the preparation and characterization of polycaprolactone nanofibers for the proof-of-concept construction of the alveolar-capillary interface. We include parameters of nanofibers manufacturing as well as their characterization. We prepared nanofibers from polycaprolactone, polylactic acid and polyamide. We used polycaprolactone nanofibers to model the alveolar-capillary interface of human lung: We electrospun the nanofibers onto supporting mesh and incorporated the whole structure into 3D-printed insert to create the nanofibrous cell co-culture scaffold. For reproducing the 3D-printing of 24-well plate co-culture insert, see also the GitHub repository <https://github.com/Grindyd/Nanofiber-holder-insert/>. We seeded the scaffold with capillary endothelial cells (HUVEC) and alveolar epithelial cells (ELEP) to mimic the alveolar-capillary interface. For reproducing our protocol for differentiation of ELEP (Expandable lung epithelium) from hESCs (Human embryonic stem cells) see our protocol in the publication, DOI: 10.1007/s13770-022-00458-0. Importantly, we include detailed data from cell culture and co-culture experiments leading to construction of the in vitro alveolar-capillary interface proof-of-concept model. Some conclusions based on these data have been summarized in this publication: <https://doi.org/10.1002/jbm.a.37824>.

**Methods:** [English](#)

Nanofibers production and characterization: We produced nanofibers by electrospinning method, using Nanospider technology. The nanofibers were electrospun onto supporting polyamide mesh. The characterization of nanofibrous structures provided in this dataset is based on advanced microscopic techniques (SEM). Cell co-culture scaffolds production: The nanofibrous structures electrospun onto supporting polyamide mesh were mounted into 3D-printed polyamide insert and used for cell culture and co-culture. We provide detailed description of this arrangement and the methodology used to get it. Cell culture and co-culture analysis: The analysis of cell culture and co-culture provided in this dataset is based on standard microscopic techniques (brightfield microscopy), advanced microscopic techniques (SEM), biochemical methods (MTT and CV assay).

Your (author)

Citable (DOI) + Findable

Accessible + Interoperable

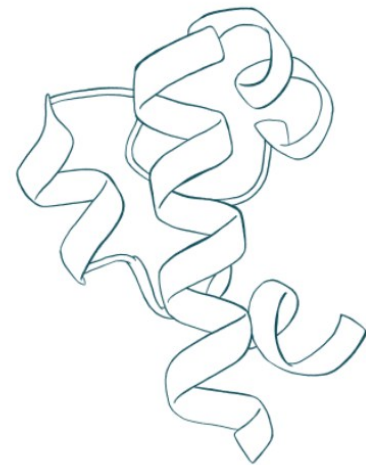
Reusable (licence)


Machine actionable (metadata)


= AI-ready record (dataset)

**VALUABLE SCIENTIFIC RESULT**



# Molecular Biophysics Database




 Documentation Tutorials [Deposit](#) [sign in with ID](#)



## Molecular Biophysics Database

 [All records](#) 

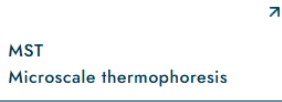
Example: Lysozyme, NaCl, K\_D, homo sapiens, xyhds-adf3t (record id) [Advanced search](#)




**Molecular Biophysics Database** collects raw data produced in experiments with biomolecular samples, biological material or other material, using molecular biophysics methods, such as Microscale Thermophoresis (MST), Biolayer interferometry (BLI), Surface Plasmon Resonance (SPR) and others.

**MBDB** development is supported by the project **MOSBRI** - Molecular Scale Biophysics Research Infrastructure of the European Commission, no. 101004806


### Currently supported techniques



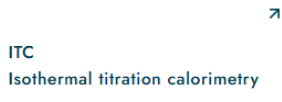
**MST**  
Microscale thermophoresis



**BLI**  
Bio-layer interferometry



**SPR**  
Surface plasmon resonance



**ITC**  
Isothermal titration calorimetry



<https://mbdb-data.org/>

# National Metadata Directory


[← BACK TO SEARCH RESULTS](#)

Search datasets..

Search

Mar 3, 2025

## Data for Strain-Induced Decoupling Drives Gold-Assisted Exfoliation of Large-Area Monolayer 2D Crystals

Ziewer, Jakob  ; Ghosh, Abyay  ; Hanušová, Michaela  ; Pirker, Luka  ; Frank, Otakar  ; Velický, Matěj  ; Grüning, Marta  ; Huang, Fumin Ústav fyzikální chemie J. Heyrovského AV ČR  (Publisher)Dataset 

Fyzika kondenzovaných látek (zahrnuje fyziku pevných látek, supravodivost)

Condensed matter physics (including formerly solid state physics, supercond.)

Nanomateriály (výroba a vlastnosti)

Nano-materials (production and properties)

2D materials

decoupling

gold-assisted exfoliation

MoS2

Raman spectroscopy

strain

### Descriptions

angličtina

This is a dataset supporting a published article "Strain-Induced Decoupling Drives Gold-Assisted Exfoliation of Large-Area Monolayer 2D Crystals". Gold-assisted exfoliation (GAE) is a groundbreaking mechanical exfoliation technique that produces centimeter-scale single-crystal monolayers of 2D materials. Such large, high-quality films offer unparalleled advantages over the micron-sized flakes typically produced by conventional exfoliation techniques, significantly accelerating the research and technological advancements in the field of 2D materials. Despite its wide applications, the fundamental mechanism of GAE remains poorly understood. In this study, using MoS2 on Au as a model system, ultra-low frequency Raman spectroscopy is employed to elucidate how the interlayer interactions within MoS2 crystals are impacted by the gold substrate. The results reveal that the coupling at the first MoS2-MoS2 interface between the adhered layer on the gold substrate and the adjacent layer is substantially weakened, with the binding force being reduced to nearly zero. This renders the first interface the weakest point in the system, thereby the crystal preferentially cleaves at this junction, generating large-area monolayers with sizes comparable to the parent crystal. Biaxial strain in the adhered layer, induced by the gold substrate, is identified as the driving factor for the decoupling effect. The strain-induced decoupling effect is established as the primary mechanism of GAE, which can also play a significant role in general mechanical exfoliations.



Open access

License:


CS

Creative Commons Uvedte původ 4.0

Mezinárodní licence

[More info](#)

Dataset identifiers:

DOI [10.48700/datst.hcmd1-jp979](https://doi.org/10.48700/datst.hcmd1-jp979) Original sources 


Publisher:

Ústav fyzikální chemie J. Heyrovského AV ČR [Link to original record](#)

Published in:



EOSC CZ Data Repo

Export Native JSON 

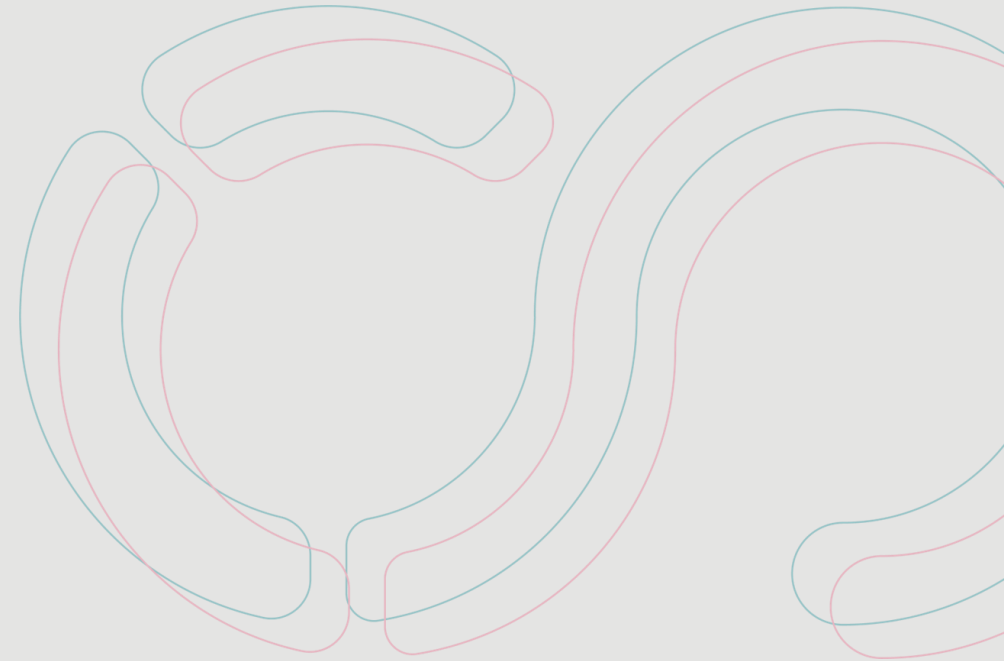
Export

<https://nma.eosc.cz/>



# National Repository Platform (NRP) Project

Focusing



# NRP for End-Users

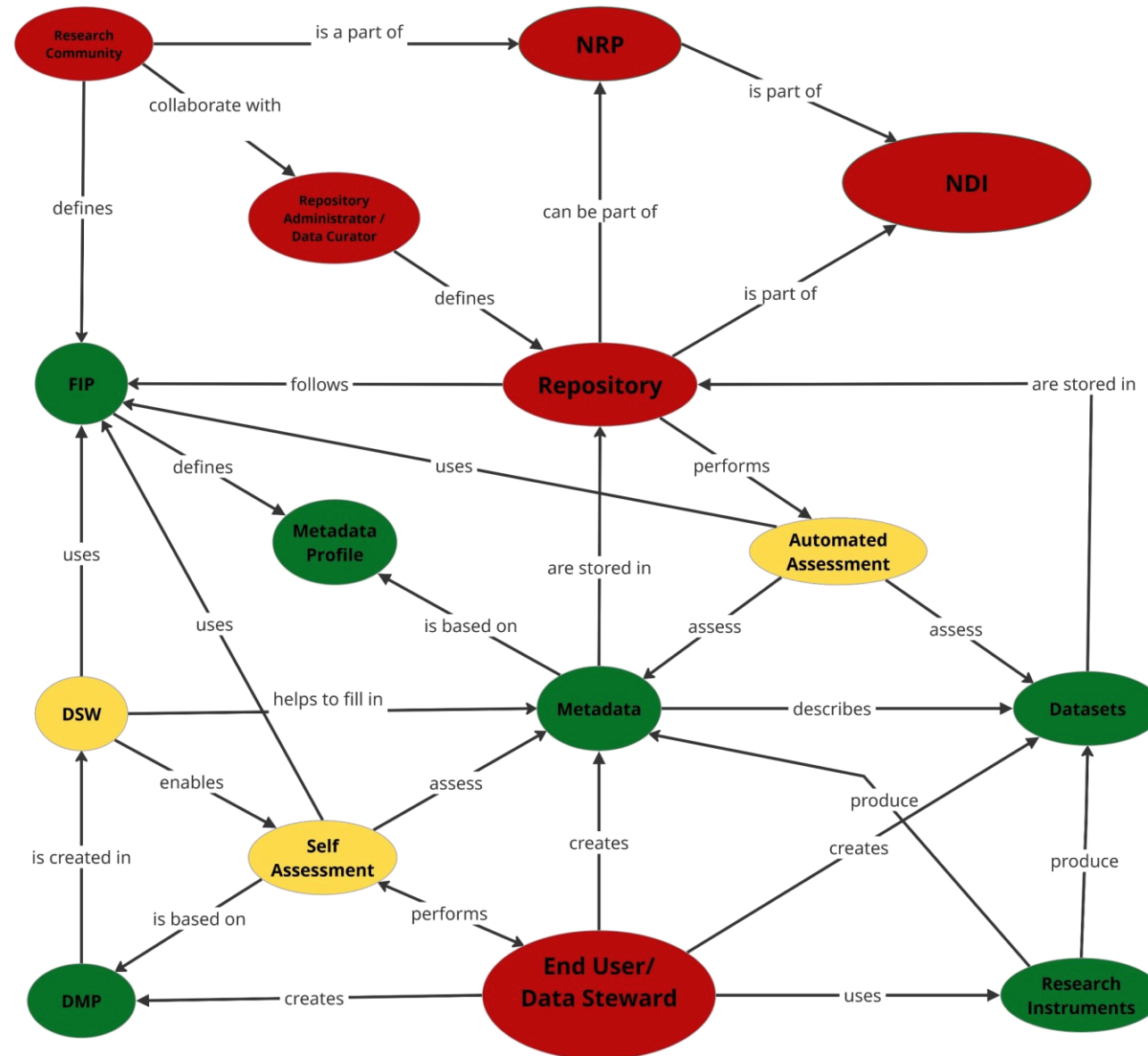
- The main objective of NRP:
  - provide various research data-related services for NRP users:
    - academic institutions and
    - individual scientists and researchers.
- What NRP will provide
  - to enhance their abilities to fulfil FAIR principles and
  - data management requirements
  - through the life cycle of research data in projects.

# NDI Outputs: Tools and Services

- Support for data management planning (DMP).
- Metadata profile management.
- Support for license handling.
- Support for working with persistent identifiers.
- Support for FAIRification of research data.
- Automation of data collection.
- Electronic laboratory notebooks.
- Overall cybersecurity and system compliance.



# Schema of FAIR Implementation in NDI/NRP







# Data Access Control

- FAIR Data, “As open as possible, as closed as necessary”.
  - We need precise access control to data, metadata, services.
- Federative system – use your institutional identity, your well-known authentication web page.
  - Effective collaboration across institutions, individuals, ...

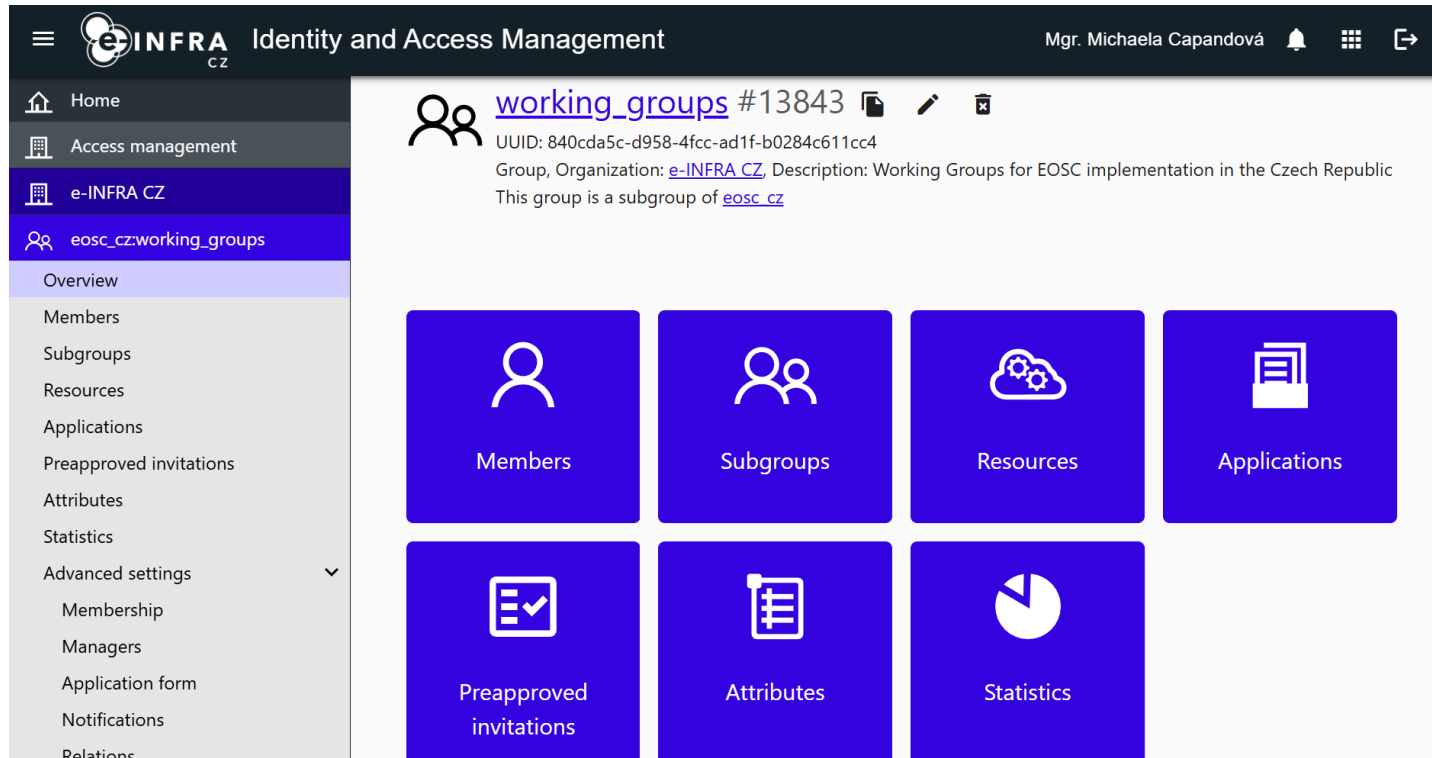
Log in with

---

Masaryk Memorial Cancer Institute (MMCI)	
Masaryk Public Library	
Masaryk University	
Municipal library T.G. Masaryk Sumperk	

# Authentication and Authorization Infrastructure

Enabling **users** from different institutions to **easily access data and services**.



- Access and **identity** management
- Group and **role** management
- **Permission** assignment



<https://perun-aai.org/>

# SensitiveCloud

Secure environment for storing, sharing  
and processing sensitive data.

- Primarily designed for work with **your own sensitive data**.
- Gradually extended to support **controlled data sharing**.
  - Main technical component for handling sensitive data within NDI.
- Includes **storage**, **computing** resources and support for ready-to-use web applications.

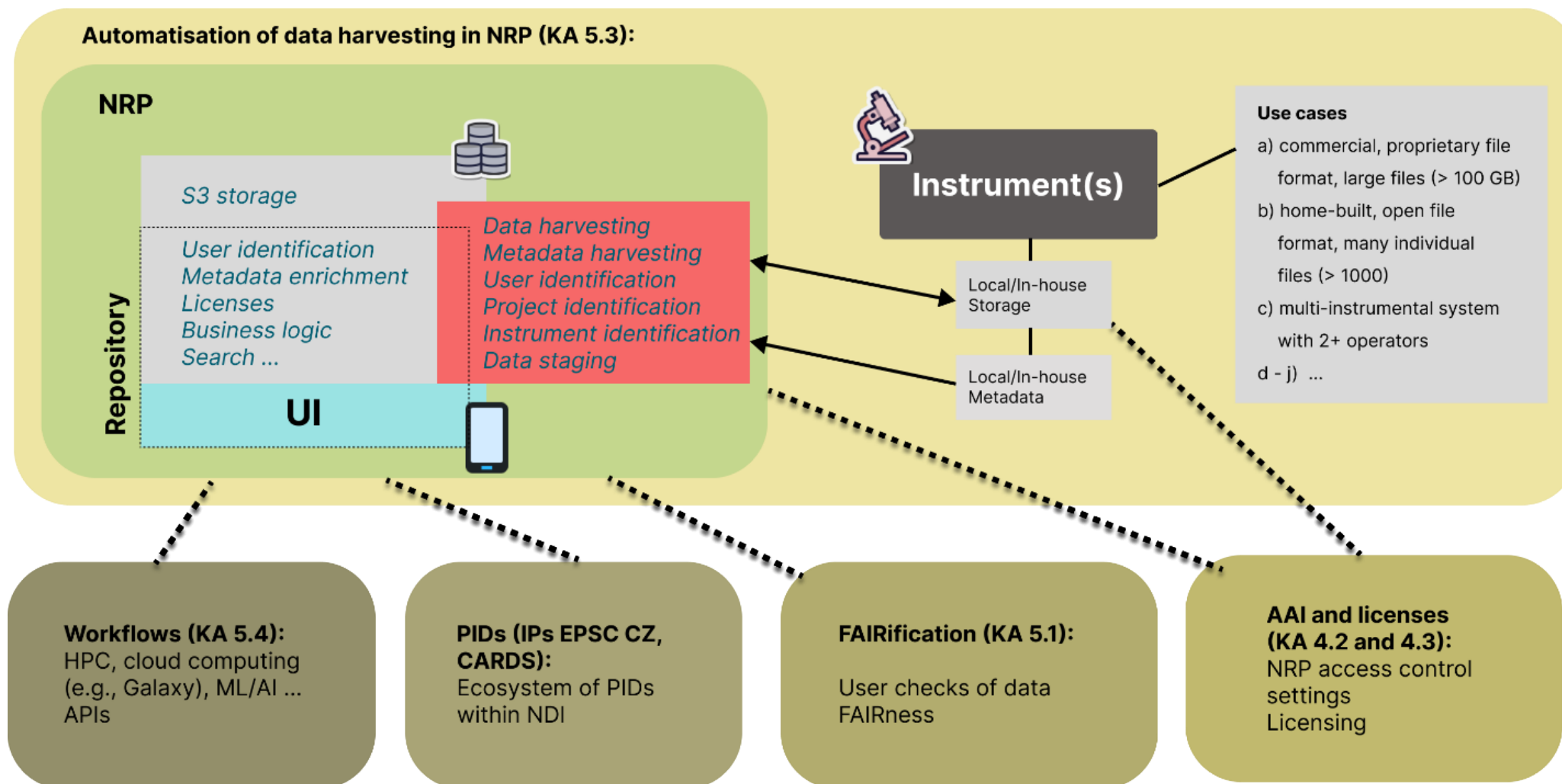


[SensitiveCloud](#)



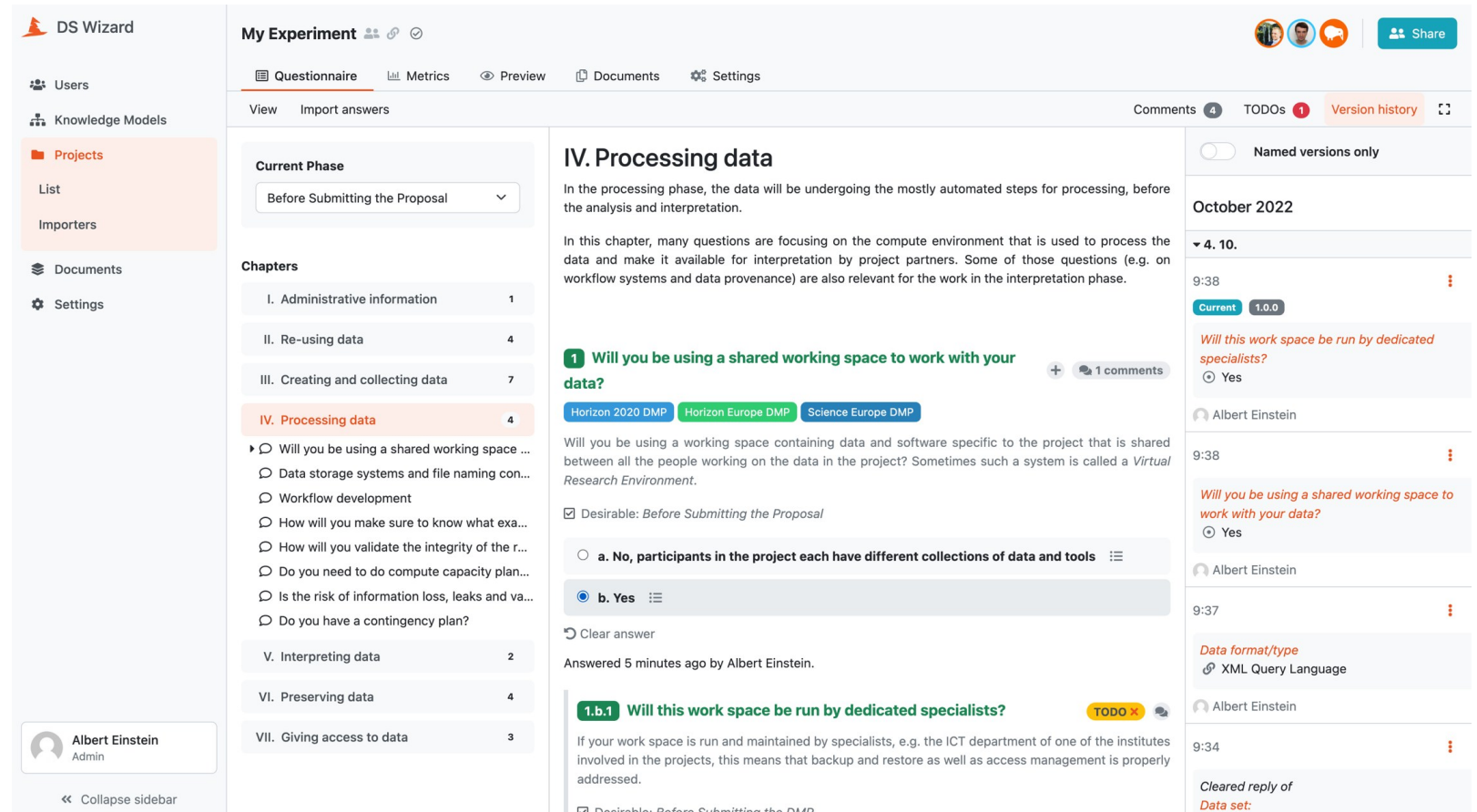
# eosc Automation of Data and Metadata Collection

**FAIR:** Accessible,  
Interoperable,  
Reusable



# Support for Data Management Planning

- We expect integration of tools like Data Stewardship Wizard (DSW) directly to the platform.
  - <https://dmp.eosc.cz/>
- Integration allows effective re-use of available (meta)data.

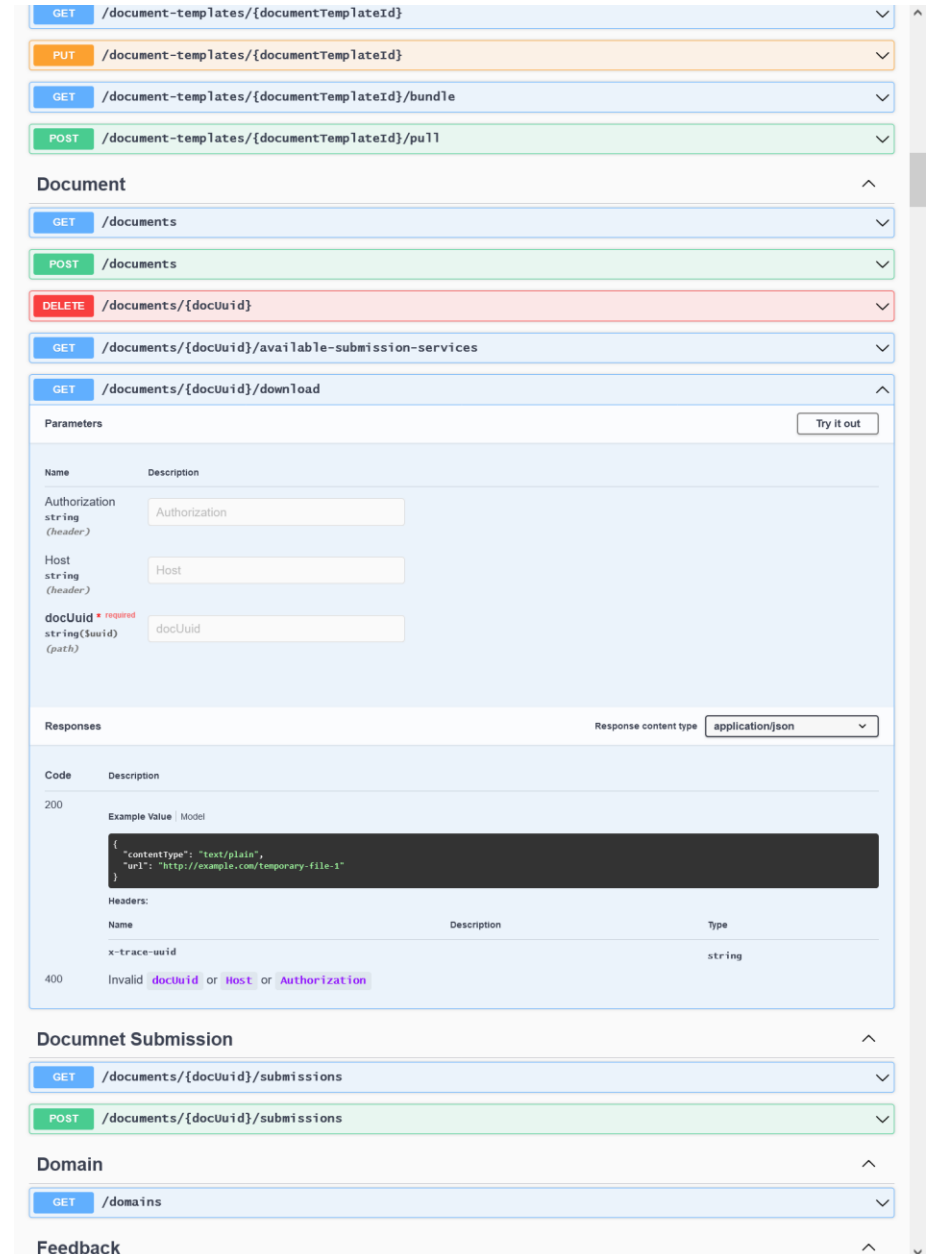


The screenshot displays the DS Wizard interface for 'My Experiment'. The sidebar on the left contains navigation links: Users, Knowledge Models, Projects (highlighted), Documents, and Settings. The main content area is titled 'My Experiment' and includes tabs for Questionnaire, Metrics, Preview, Documents, and Settings. The 'Current Phase' is set to 'Before Submitting the Proposal'. A list of chapters is shown, with 'IV. Processing data' selected. This chapter contains a question: '1 Will you be using a shared working space to work with your data?'. The question has two options: 'a. No, participants in the project each have different collections of data and tools' and 'b. Yes'. Option 'b' is selected. The right sidebar shows a version history for October 2022, with a current version 1.0.0. The interface also includes a 'Share' button and a 'Version history' link.

# Machine-Actionable

FAIR: Findable,  
Accessible,  
Interoperable

- Automate as much as possible.
  - API and machine-readability for data and metadata.
- As much as possible read from metadata of the datasets, from infrastructure configuration, ...



The screenshot shows an API documentation interface with a sidebar on the left containing a list of endpoints categorized by resource: document-templates, Document, Documnet Submission, Domain, and Feedback. The main panel displays the details for the endpoint `GET /documents/{docUuid}/download`. It includes a 'Parameters' section with input fields for 'Authorization' (string, header), 'Host' (string, header), and 'docUuid' (string, path, marked as required). Below this is a 'Responses' section showing a 200 status code with an example JSON response and a table of headers, including 'x-trace-uuid' (string). A 400 status code response is also listed as 'Invalid docUuid or Host or Authorization'. The 'Response content type' is set to 'application/json'.

GET /document-templates/{documentTemplateId}

PUT /document-templates/{documentTemplateId}

GET /document-templates/{documentTemplateId}/bundle

POST /document-templates/{documentTemplateId}/pull

Document

GET /documents

POST /documents

DELETE /documents/{docUuid}

GET /documents/{docUuid}/available-submission-services

GET /documents/{docUuid}/download

Parameters

Try it out

Name	Description
Authorization string (header)	Authorization
Host string (header)	Host
docUuid * required string(docUuid) (path)	docUuid

Responses

Response content type: application/json

Code	Description						
200	<p>Example Value   Model</p> <pre>{  "contentType": "text/plain",  "url": "http://example.com/temporary-file-1"}</pre> <p>Headers:</p> <table><thead><tr><th>Name</th><th>Description</th><th>Type</th></tr></thead><tbody><tr><td>x-trace-uuid</td><td></td><td>string</td></tr></tbody></table>	Name	Description	Type	x-trace-uuid		string
Name	Description	Type					
x-trace-uuid		string					
400	Invalid docUuid or Host or Authorization						

Documnet Submission

GET /documents/{docUuid}/submissions

POST /documents/{docUuid}/submissions

Domain

GET /domains

Feedback

# Persistent Identifiers

 **identifikatory.cz**

Persistent Identifiers

Persistent Identifiers ▾ Services ▾ About us News



Home / Persistent Identifiers

## Persistent Identifiers

Learn more about each persistent identifier (PID). Persistent identifiers are tools that are used to uniquely identify people, organisations, and other objects (e.g., books, articles, datasets) in a scholarly communication system.

**ORCID iD for  
researchers**

**DOI for objects**

**ISBN for books**

**ISSN for periodicals**

**ISMN for notated  
music**

**ROR for  
organizations**

**IGSN for samples**

**Other PIDs**

**NTK**

50°6'14.083"N, 14°23'26.265"E  
Národní technická knihovna  
National Library of Technology



Persistent Identifiers

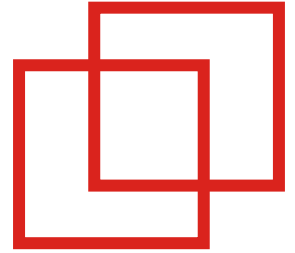
ORCID iD for researchers  
DOI for objects

Services

National ORCID Centre  
FAQs – ORCID

About us

News



<https://identifikatory.cz/en/>

# Support Work with Licenses

- Templates of **deposition licenses** governing the rules for upload of datasets to repositories.
- **License chooser** for users' datasets on upload.
- Framework for dataset's **license-based access control** to dataset with limited access.
- **Integration and machine actionability**.
  - Connection to data management, access control, ...

### Choose a License

Answer the questions or use the search to find the license you want



[Start again](#) [←](#) [→](#)

Is your data within the scope of copyright and related rights?

[Yes](#) [No](#)



#### Public Domain Dedication (CC Zero)

CC Zero enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.

[Publicly Available](#)   [OPEN DATA](#)

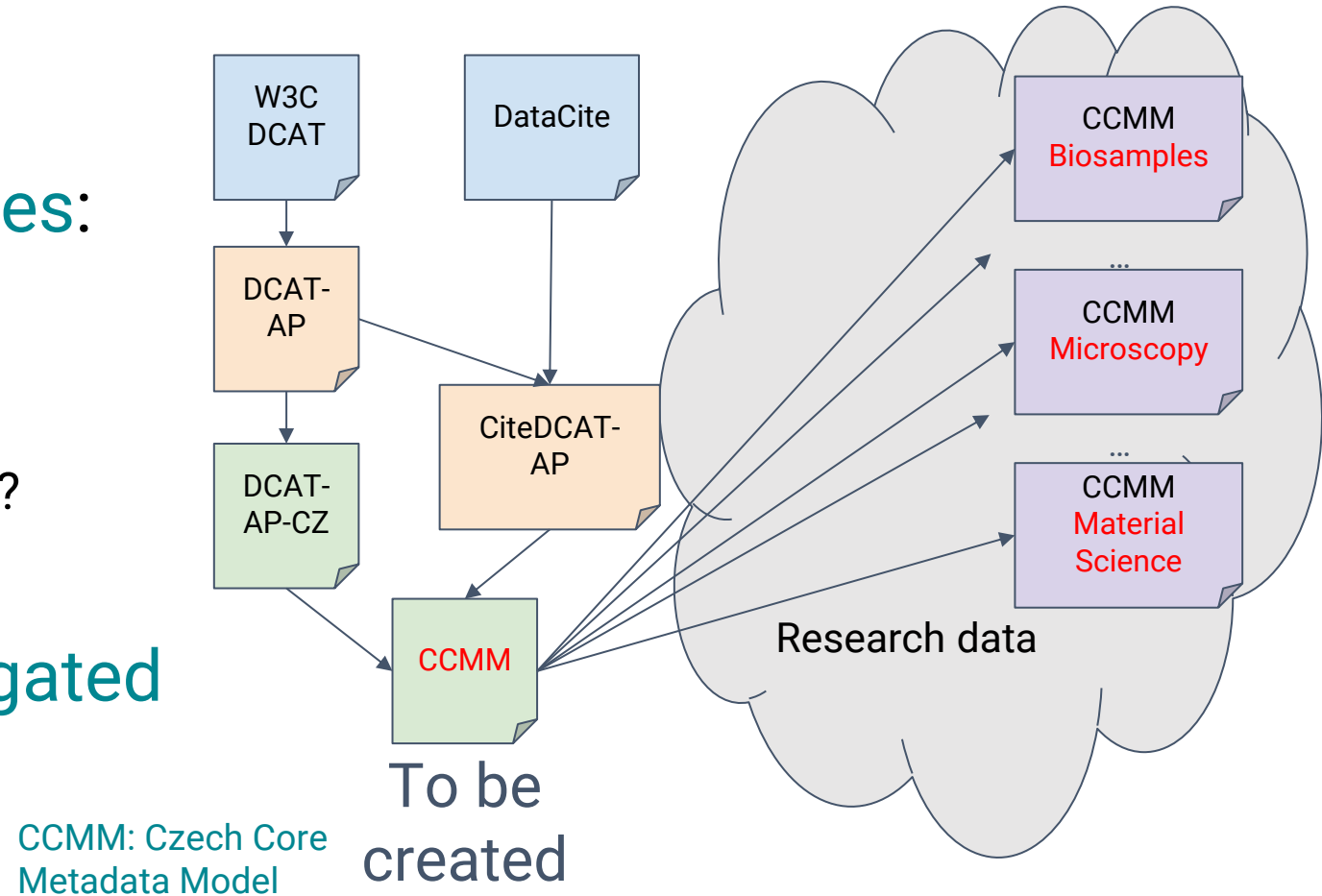
#### Creative Commons Attribution (CC-BY)

This is the standard creative commons license that gives others maximum freedom to do what they want with your work.

[Publicly Available](#)   [OPEN DATA](#)

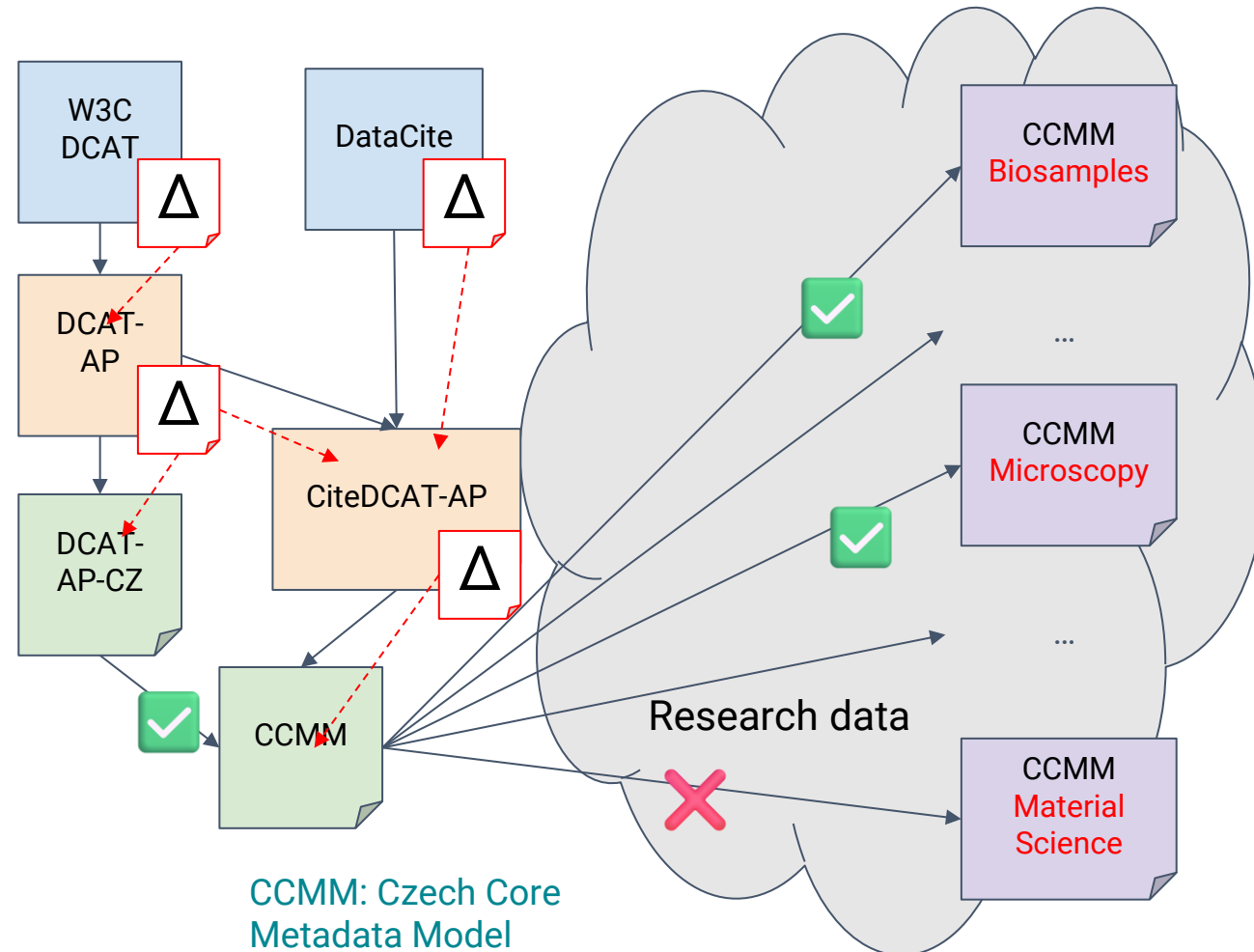
# Managing Metadata Profiles in NRP

- <https://dataspecer.com/>
- Research data metadata profiles:
  - What happens, when
    - DCAT v2 → DCAT v3?
    - DCAT-AP v2.1.1 → DCAT-AP 3.0.1?
    - DataCite 4.4 → DataCite 4.5?
- We want changes to be propagated automatically.

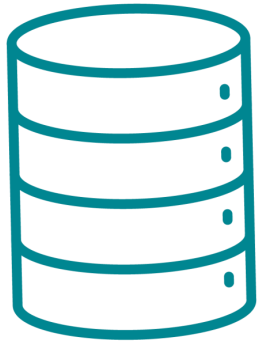


# Managing Metadata Profiles in NRP

- <https://dataspecer.com/>
- But also
  - profile compliance validation (✓ ✗),
  - description of changes in specifications (Δ),
  - change propagation mechanism (----->),
  - implementation in tools.



# NDI Outputs: Summary



**Storage capacities**



**Tools and services**



**Computing capacities**

# Useful Links and Contacts

## [EOSC CZ Website](#)



## [EOSC CZ Newsletter](#)



- Any questions?
  - [info@eosc.cz](mailto:info@eosc.cz)
- Ideas for a lecture or a training?
  - [events@eosc.cz](mailto:events@eosc.cz)
- Get in touch with our PR
  - [pr@eosc.cz](mailto:pr@eosc.cz)



@EOSC Czech Republic

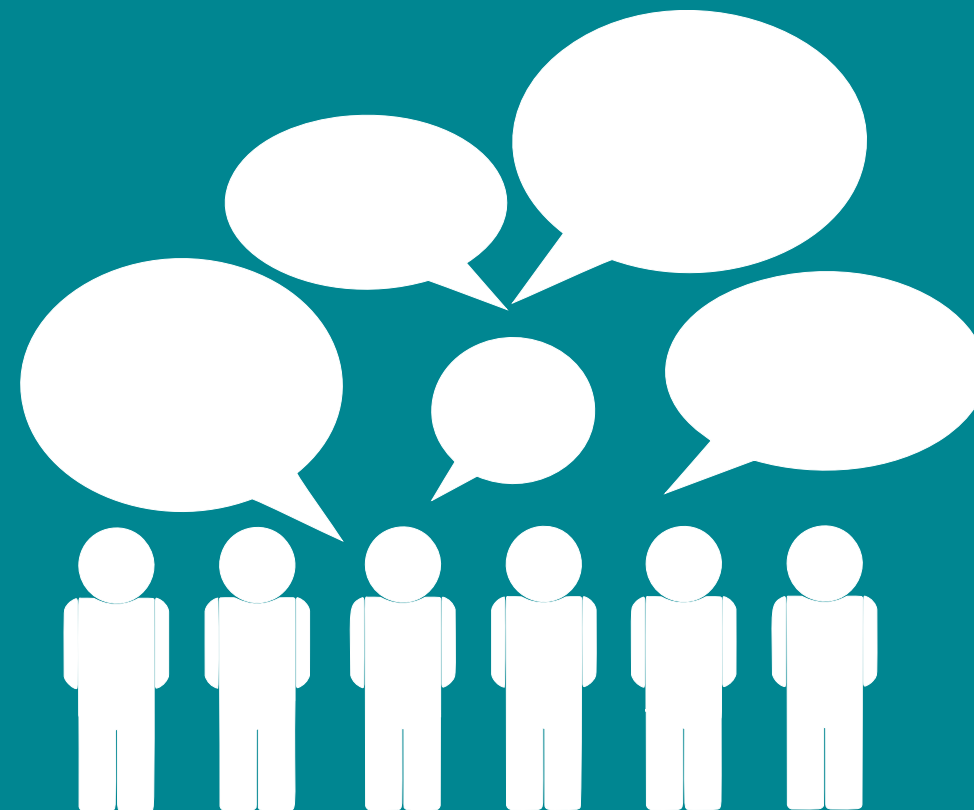


@eosccz.bsky.social

# Thank you for your attention

E: [ruzicka@ics.muni.cz](mailto:ruzicka@ics.muni.cz)

E: [info@eosc.cz](mailto:info@eosc.cz) W: [www.eosc.cz/en](http://www.eosc.cz/en)



Source: Communicate\_communication\_conference\_2028004 by OpenClipart-Vectors from Pixabay



Spolufinancováno  
Evropskou unií



MINISTERSTVO ŠKOLSTVÍ,  
MLÁDEŽE A TĚLOVÝCHOVY

MUNI  
ICS

cesnet  
....

VŠB TECHNICKÁ  
UNIVERZITA  
OSTRAVA

IT4INNOVATIONS  
NÁRODNÍ SUPERPOČÍTAČOVÉ  
CENTRUM