

An Introduction to Effective Data Management Planning

Antonín Zita <zita@ics.muni.cz> & EOOSC CZ Secretariat
Masaryk University | CERIT-SC

Institute of Physics of the Czech Academy of Sciences
28 November 2025

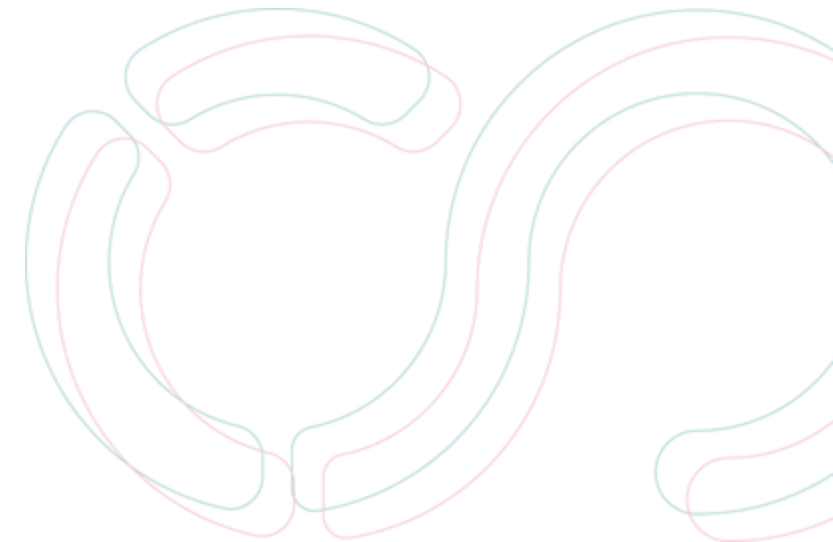


Spolufinancováno
Evropskou unií



Outline

1. The Road to FAIR Data
2. FAIR Data: A Revision
3. Data Management Plan (DMP) and Data Management Planning (DMPlanning)
4. Introduction to Data Stewardship Wizard (DSW)
5. Practical Part: DSW – First Steps
6. Practical Part: DSW – Advanced Functions
7. Future Development of DSW
8. Support for Data Management



The Road to FAIR Data

Chapter

Supporting the replication of your research

2023, Research Ethics in Behavior Analysis

Heather J. Forbes, ... Jeneé Vickers Johnson

The replication crisis

The **replication crisis** is a term used to refer to widespread concern about the accumulation of erroneous studies in scientific literature. Psychology researchers arguably have played a central role in the replication crisis (Wiggins & Chrisopherson, 2019), but other fields including medicine (Gehr et al., 2006; Ioannidis, 2005), economics (Duvendack et al., 2017), political science (Gerber et al., 2001), and education (Makel & Plucker, 2014) also appear to be affected. Although much

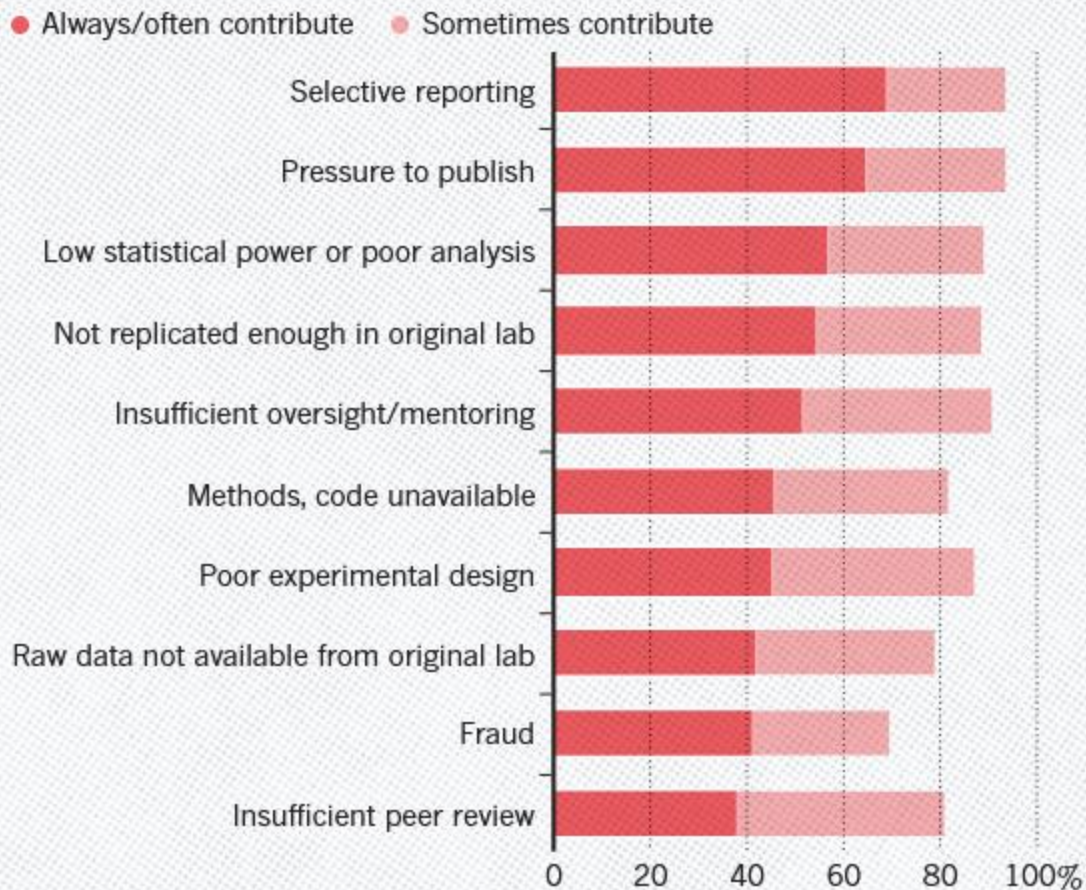


News Feature | Published: 25 May 2016

1,500 scientists lift the lid on reproducibility

WHAT FACTORS CONTRIBUTE TO IRREPRODUCIBLE RESEARCH?

Many top-rated factors relate to intense competition and time pressure.



IS THERE A REPRODUCIBILITY CRISIS?

A *Nature* survey lifts the lid on
how researchers view the 'crisis'
rocking science and what they
think will help.

BY MONYA BAKER

52%
Yes, a significant
crisis

38%
Yes, a slight
crisis

7%
Don't know

3%
No, there is no crisis

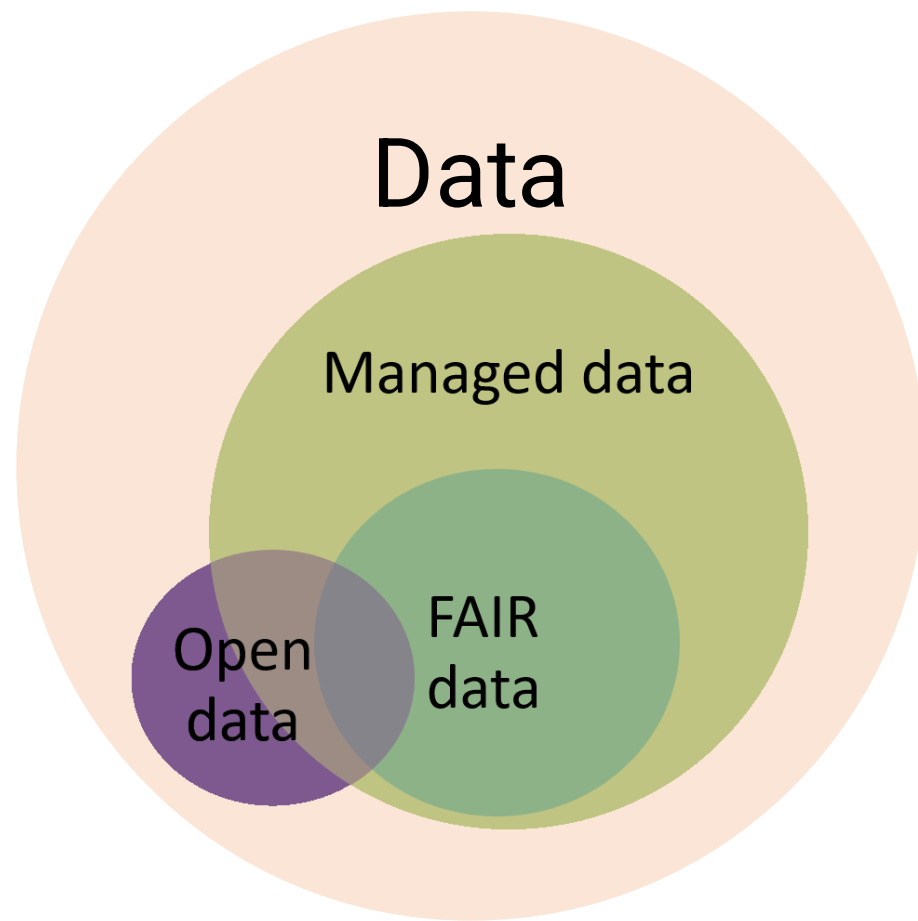
1,576
RESEARCHERS SURVEYED

The Road to FAIR Data II

“The ‘replication crisis’ has introduced a number of considerable challenges, including compromising the public’s trust in science¹⁵ and undermining the role of science and scientists as reliable sources to inform evidence-based policy and practice¹⁶. **At the same time, the crisis has provided a unique opportunity for scientific development and reform.** In this narrative review, we focus on the latter, exploring the replication crisis through the lens of a credibility revolution¹⁷...”

The replication crisis has led to positive structural, procedural, and community changes.
Korbacher et al. *Communications Psychology*, 1 (2023): 2.

FAIR Data vs Open Data



- FAIR principles

- Findable
- Accessible
- Interoperable
- Reusable

“As open as possible, as closed as necessary.”

Source: FAIR Handbook for the Data Steward Community in the Czech Republic (ENG), <https://doi.org/10.71495/hxfc-6f57>

FAIR Principles I

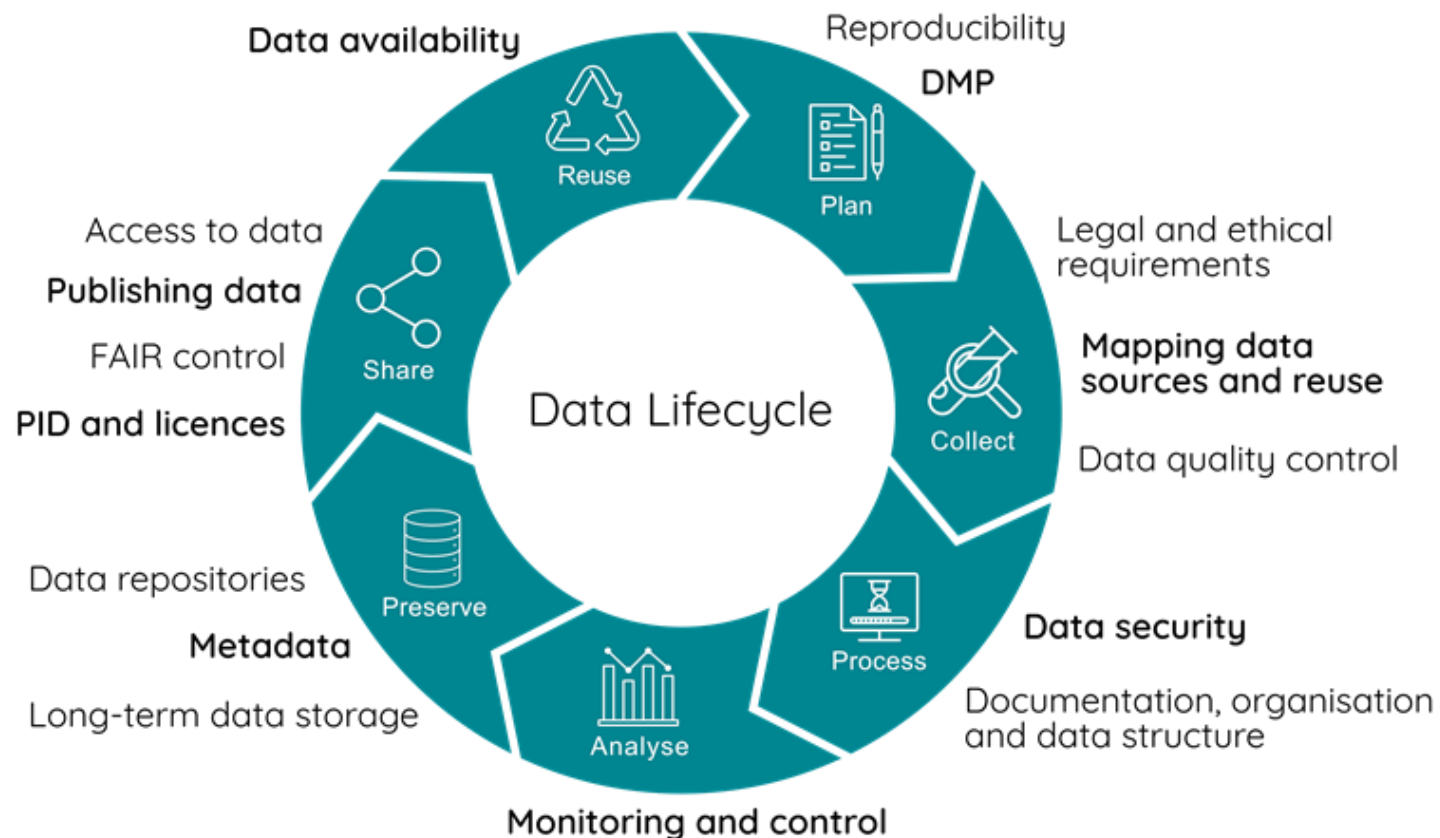
- **FAIR** is an abbreviation of the following:
- **Findable**
 - For **people and machines**.
 - We need **(machine-readable) metadata** that is available to search engine/users.
 - We need **unique persistent identifiers** (PIDs).
- **Accessible**
 - Technical solution for **access to data and metadata** – 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**
 - **the goal** 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.

FAIR Principles II

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

FAIR principle	Question	FAIR enabling resource types	Your answers
F1	What globally unique, persistent, resolvable identifier service do you use for metadata records?	Identifier service	e.g. PURL, DOI
F1	What globally unique, persistent, resolvable identifier service do you use for datasets?	Identifier service	
F2	Which metadata schemas do you use for findability?	Metadata schema	
F3	What is the schema that links the persistent identifiers of your data to the metadata description?	Metadata-Data linking schema	
F4	In which registry are your metadata records indexed?	Registry	
F4	In which registry are your datasets indexed?	Registry	
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 service do you use for metadata records?	Authentication & authorisation service	
A1.2	Which authentication & authorisation service do you use for datasets?	Authentication & authorisation service	
A2	Which metadata preservation policy do you use?	Metadata preservation policy	
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 vocabulary	
I2	Which structured vocabularies do you use to encode your datasets?	Structured vocabulary	
I3	Which models, schema(s) do you use for your metadata records?	Semantic model	
I3	Which models, schema(s) do you use for your datasets?	Semantic model	
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	
R1.3	Who is the community, and what are their domain-relevant community standards?	FAIR Implementation Profile	

Data Lifecycle



Source: Data Stewards, EOSC CZ <https://www.eosc.cz/en/communities/data-stewards>

- Managing data a complex task
- Good practices vital – ideally from the very beginning!

FAIR Data

OTEVŘENÁ DATA (OPEN DATA)

Výzkumná data jsou nezbytná k možnosti ověření a dalšího využití výsledků výzkumu. Aby GA ČR dostála svým zákonným povinnostem v oblasti otevřených dat, zahrnula správu výzkumných dat do politiky podpory grantových projektů. Cíle této politiky:

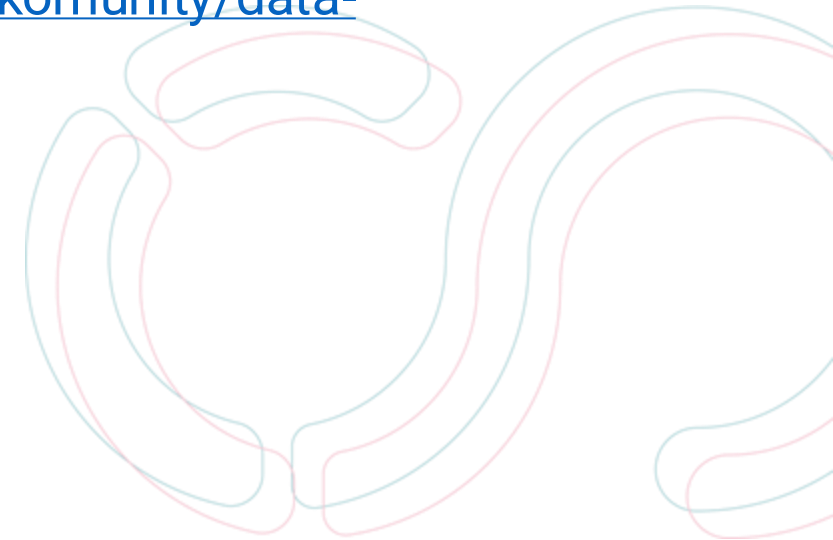
- Required
 - MEYS, GA CR, ...
- Open Data vs FAIR data
 - Somewhat incorrect definition
 - <https://gacr.cz/ga-cr-otevrena-veda/>
- všechny vědecké publikace vzniklé v rámci projektů financovaných GA ČR mají dohledatelná otevřená data (se zákonnými výjimkami jako ochrana duševního vlastnictví, obchodního tajemství, bezpečnosti státu nebo oprávněných obchodních zájmů příjemce nebo třetí osoby)
- řešitelé projektů GA ČR předkládají plán správy dat (Data Management Plan – DMP) nejpozději k datu doručení první dílčí zprávy projektu a pravidelně tento plán aktualizují
- otevřená data naplňují principy FAIR, tj. jsou dohledatelná, přístupná, interoperabilní a opětovně využitelná (Findable, Accessible, Interoperable, and Reusable)
- náklady spojené s otevřenými daty jsou způsobilými náklady projektu, GA ČR doporučuje zřídit v rámci projektu pozici odborníka na správu dat (data stewarda)

zákonem č. 130/2002 Sb. a s ohledem na výzkumná data podporují zásady FAIR.

Otevřená věda v ČR. MŠMT.

Data Management Plans

- Currently required:
 - By funders (GA CR, TA CR, AS CR, ...)
 - By law (130/2002 Coll., amended on 2022)
 - [DMP Requirements of Major Funders](https://www.eosc.cz/komunity/data-stewardi/data-stewardi-v-cr) (at <https://www.eosc.cz/komunity/data-stewardi/data-stewardi-v-cr>)
- Generally submitted as:
 - Initial DMP
 - Updates during the project (interim reports)
 - Final DMP



DMPlanning vs DMP

1. Data Management Plan

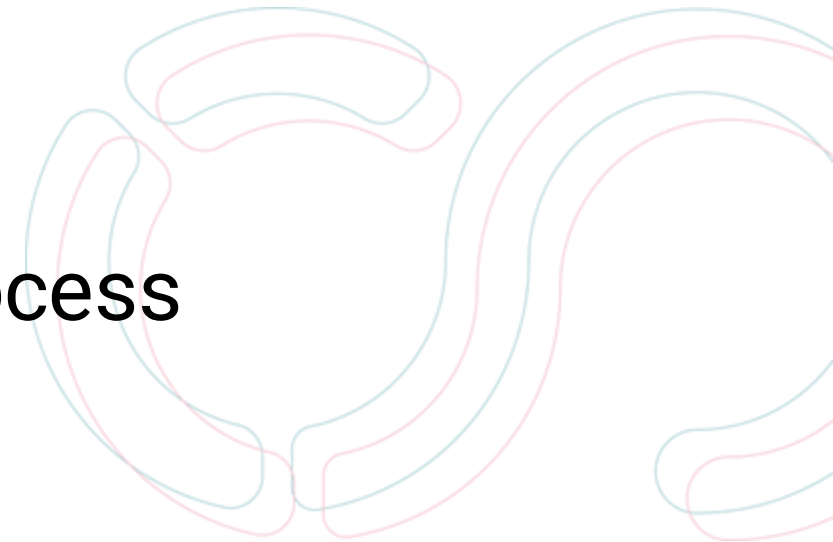
vs.

2. Data Management Planning

Dokument

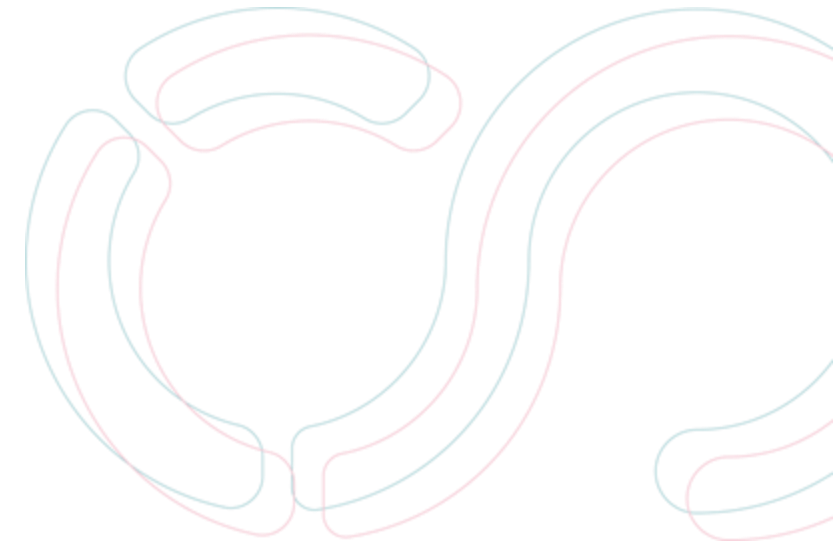
vs.

Process



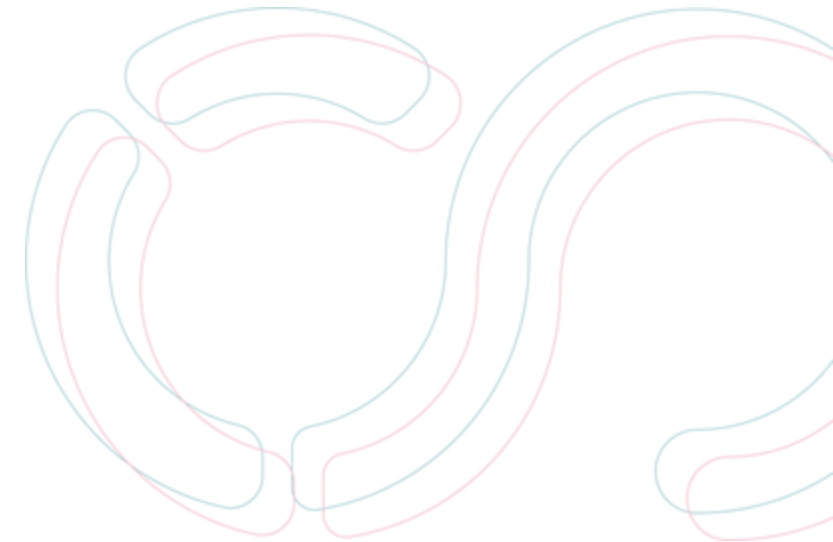
DMP Myth No. 1

- Information in the DMP cannot be changed
- DMP is a **living document**



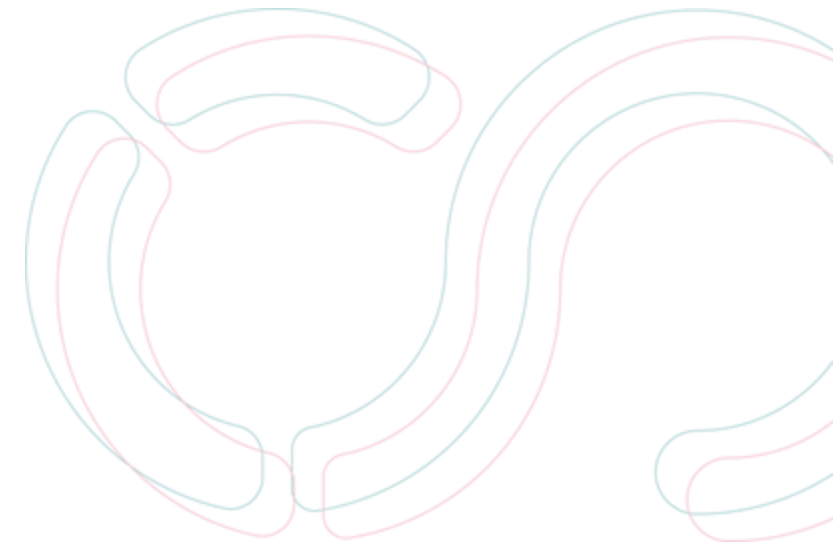
DMP Myth No. 2

- DMP should contain only the “correct” solutions
- DMP should contain the **actual solutions**



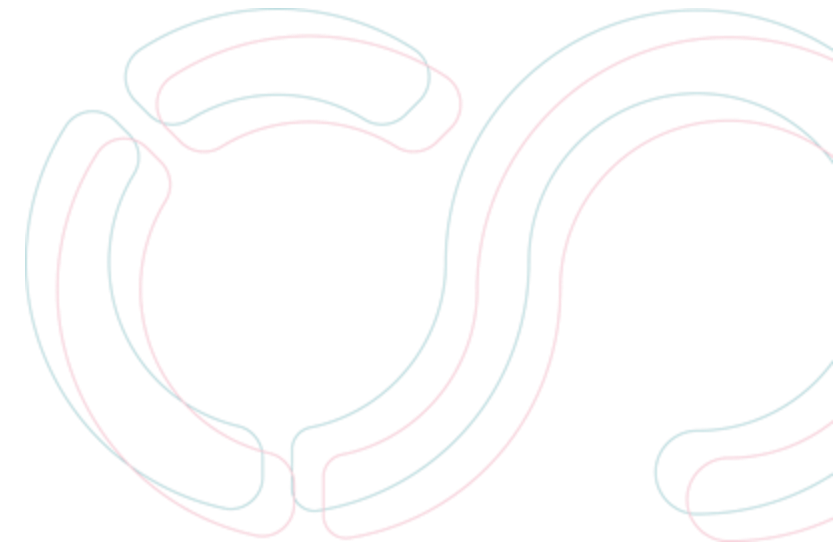
DMP Myth No. 3

- **FAIR data** must always be **open data**
- Option of **embargo/not publishing data**



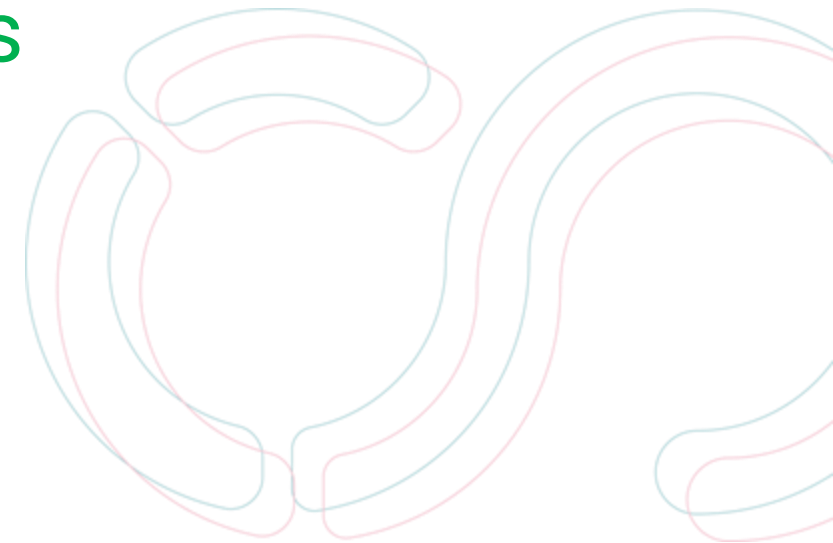
DMP Myth No. 4

- Completing the DMP in MS Word is easier
- Open-ended answers can **lead to issues**



DMP Myth No. 5

- DMP is useful only for the funding agency
- **DMP(lanning)** is useful for the researchers



DMP Put Simply

The following must be absolutely clear from the DMP:

Yes, we know how to handle our data.

Yes, we will create data in these formats and quantities, and yes, we want our data to be FAIR because we want to preserve it or increase its value. Yes, we will use secondary data. Yes, we will use repositories. Yes, we will publish our data under a certain license. ...

Examples of DMPlanning

- 1) Researchers had gaps in the process of pseudonymization of their sensitive data
-> **confidentiality agreements necessary**
- 2) Researchers unknowingly chose a server without backups for their archiving
-> **found different archival solution-**
- 3) Researchers did not know how to handle sharing of data with potential commercial value
-> **“due to the sensitive nature of the outputs, their publication is not planned. The data will serve as a basis for obtaining a utility model”.**
- 4) Researchers were worried about storage costs and persistent identifiers
-> **solved by choosing a repository**

DMP – Horizon Europe

DATA MANAGEMENT PLAN

Date:	[dd/mm/yyyy]
Version:	[DMP version]

1. Data Summary

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

What types and formats of data will the project generate or re-use?

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

What is the expected size of the data that you intend to generate or re-use?

What is the origin/provenance of the data, either generated or re-used?

To whom might your data be useful ('data utility'), outside your project?

2. FAIR data

2.1. Making data findable, including provisions for metadata

Will data be identified by a persistent identifier?

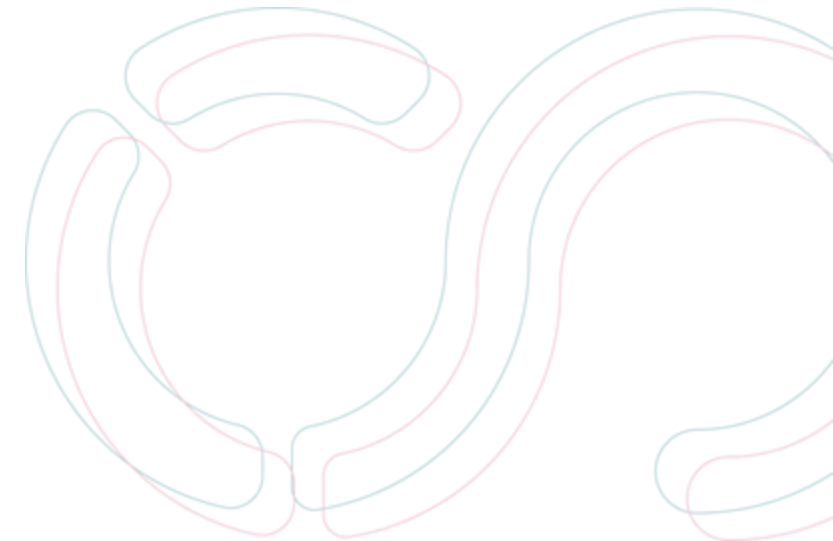
Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

Will metadata be offered in such a way that it can be harvested and indexed?

Data Stewardship Wizard – General Information

- Questionnaire open source tool
- Recommended in [Horizon Europe Programme Guide](#)
- Tracks FAIR metrics
- Filters questions by using tags
- External validation (email, ORCID, DOI, ...)
- Templates
- Creating the DMP document



Gaining Access to DSW (I)

ELIXIR DSW

- DSW Cloud within the ELIXIR infrastructure (dsw.elixir-europe.org)

Institutional Instance

- DSW is an open-source tool with publicly available documentation
- Institutions can have their own instances in which they maintain the current **Knowledge Model** and can make further modifications

FAIR Wizard

- Commercial service for institutions outside of ELIXIR or that do not wish to maintain their own instance

Gaining Access to DSW (II)

DSW NRP (dmp.eosc.cz)

- Part of the **National Repository Platform (NRP)**
- KM and other content developed specifically for the needs of the NRP
- KM different from the “regular” KM DSW, but still meeting the needs of the given calls (HE, MEYS, GA CR, ...)
- Still in pilot mode, but already accessible to **everyone*** (***within the e-INFRA infrastructure**)

Data Management Planning within National Repository Platform for Research Data



Data Stewardship Wizard (DSW) is an open-source tool for effective data management planning. Within the NRP, researchers can use a localized instance of DSW for their projects.

Currently, a **pilot operation is underway**.

[Start planning](#)



Pilotní provoz služby

Tato služba pro plánování správy dat je pilotní součástí Národní repozitářové platformy pro výzkumná data. Obsah i funkce průběžně vylepšujeme v rámci projektu a budeme rádi za vaši zpětnou vazbu.

Pilot Service Notice

This data management planning service is a pilot within the National Repository Platform for research data in the Czech Republic. We are continuously improving its content and features over the course of the project and welcome your feedback.



Co-funded by
the European Union



Provoz této služby je podporován projektem Národní repozitářová platforma pro výzkumná data (grant číslo CZ.02.01.01/00/23_014/0008787).

Log In

[Forgot your password?](#)

Log In

Or connect with

 e-INFRA CZ

Pro přihlášení se do této služby je nutné se nejprve v rámci e-INFRA CZ AAI registrovat [ZDE](#).

To log in into this service it is first required to register within e-INFRA CZ AAI organization [HERE](#).

Welcome, Antonín!

As a researcher, you create and collaborate on data management plans.



You have no projects yet, start by creating some.



Create Project

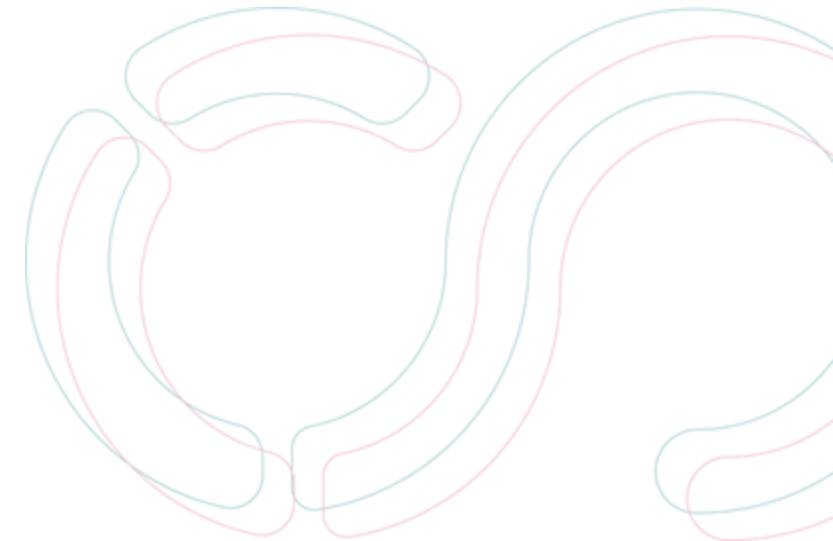
Project is a workspace where you create your DMP. It is based on a knowledge model, which contains knowledge about what should be asked and how based on the research field or organization's needs. You can use document templates to transform the answers into a document. This document can be anything, from PDF to machine-actionable JSON.

You can create a new project from a project template that data stewards prepare for you to have an easier start or from scratch where you set up everything yourself.

Create

Data Stewardship Wizard – Other Functions

- Project Ownership
 - Management multiple users
 - Sharing, commenting, ...
- Project Phases
 - Required and optional items
 - Visual feedback for questionnaire tracking
- Document Export Different
 - File types and templates
 - Migration between new KM versions
- Versioning



 DS Wizard

Dashboard

Projects

Storage Costs Evaluator

DSW Guide

Antonin Zita
Researcher

⏪ Collapse sidebar

Test 5 

Questionnaire

Metrics

Preview

Documents

Settings

View

Import replies

TODOs

Comments

Version history





Current phase

Before Submitting the DMP

Chapters

I. Administrative information3

II. Re-using data✓

III. Creating and collecting data5

IV. Processing data1

V. Interpreting data✓

VI. Preserving data6

IV. Processing data

In the processing phase, the data will be undergoing the mostly automated steps for processing, before the analysis and interpretation.

In this chapter, many questions are focusing on the compute environment that is used to process the data and make it available for interpretation by project partners. Some of those questions (e.g. on workflow systems and data provenance) are also relevant for the work in the interpretation phase.

✓ IV.1 Will you be using a shared working space to work with your data?

Horizon Europe DMP

Will you be using a working space containing data and software specific to the project that is shared between all the people working on the data in the project? Sometimes such a system is called a *Virtual Research Environment*.

☒ Desirable: Before Submitting the Proposal

☒ a. No, participants in the project each have different collections of data and tools

☐ b. Yes

 Clear answer

Answered 2 minutes ago by Antonin Zita.

✓ IV.2 How will you validate the integrity of the results?

Horizon Europe DMP

☒ Desirable: Before Submitting the DMP

☒ a. Explore

 Clear answer

Answered 2 minutes ago by Antonin Zita.

IV.3 Will you run a subset of your jobs several times across the different compute infrastructures you are using?

28 November 2025

29

User Interface – Basics

- DMP phases
- DMP chapters
- Question types:
 - Mandatory questions
 - Optional questions
- Answer types:
 - Multiple choice
 - Open text
 - Text completion

Current phase

Before Submitting the DMP

 VI.1.a.6

 VI.1.a.6

 VI.1.a.5

Chapters

I. Administrative information



- ▶  Contributors
- ▶  Research Project(s)
 -  To execute the DMP, is additional specialist e...
 -  Do you require hardware or software in addi...
- ▶  Describe national / funder / sectorial / depar...

II. Re-using data



III. Creating and collecting data



IV. Processing data



V. Interpreting data

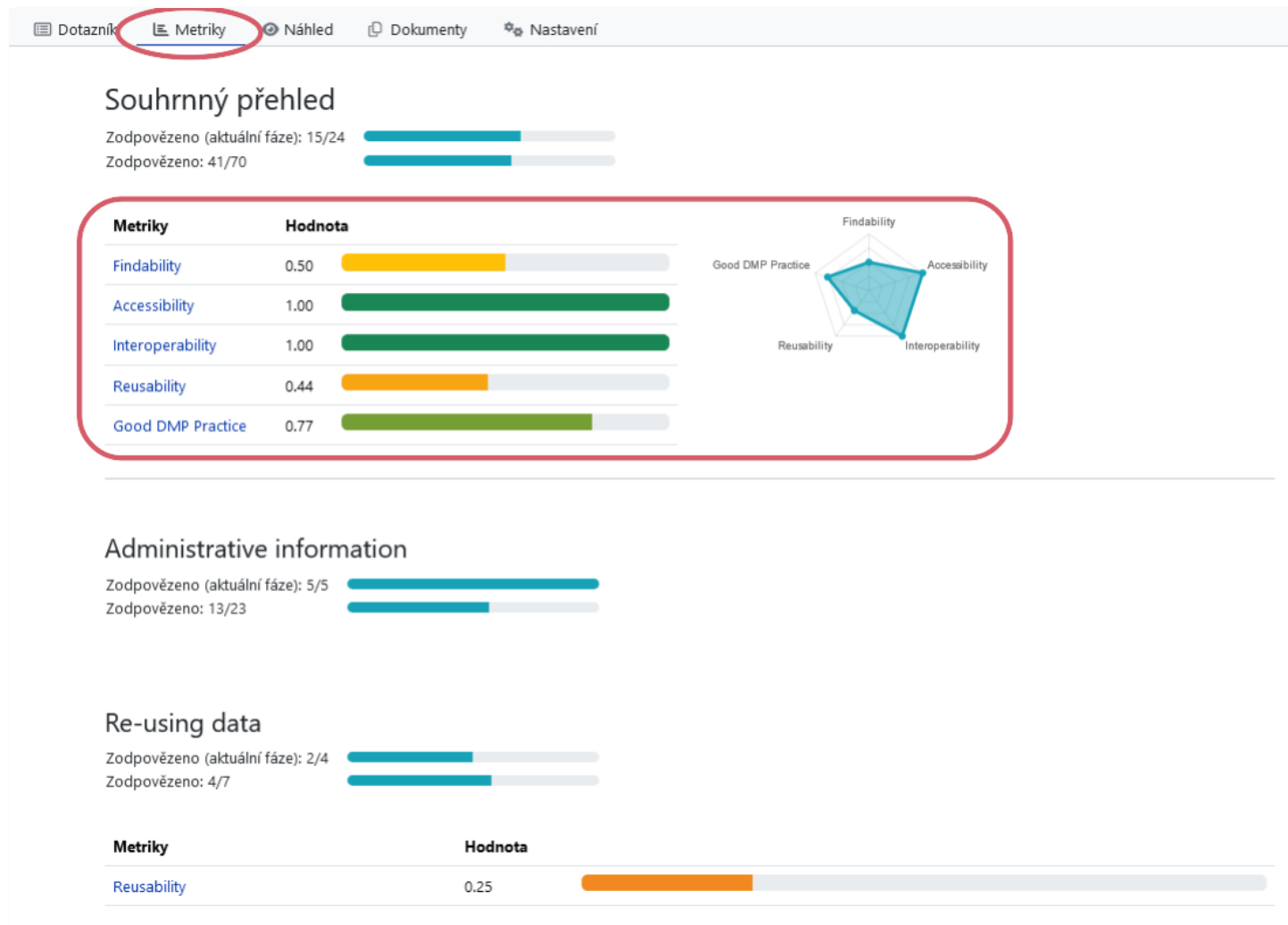


VI. Preserving data

2

VII. Giving access to data





Did you consider the possible impact to the organization if information is lost?

Good DMP Practice

e fázi: *Before Submitting the Proposal*

Good DMP Practice

the effect is small Good DMP Practice

the risk is acceptably low Good DMP Practice

we will need to work on this. Good DMP Practice

Task No. 1

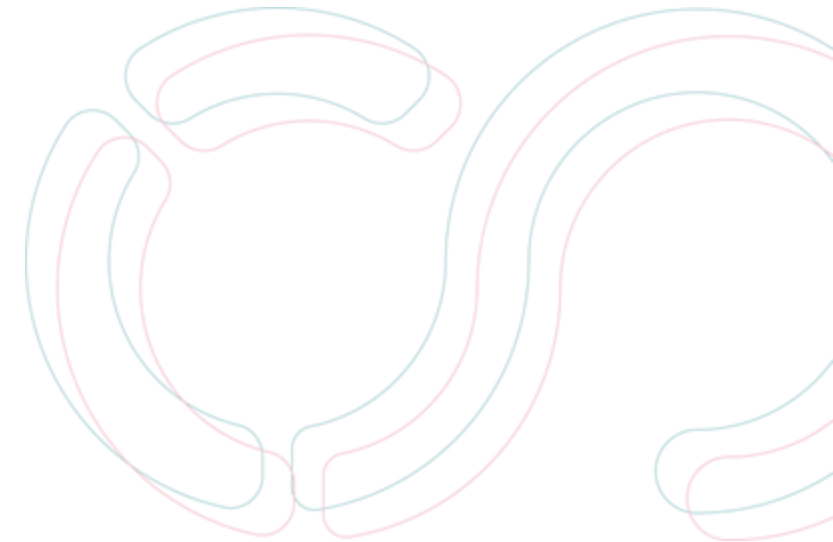
- Enter the following information into **Chapter 2 Re-using data**:
 - Freely available dataset from **Clouds and the Earth's Radiant Energy System** (<https://ceres.larc.nasa.gov/>),
 - National Aeronautics and Space Administration
 - The dataset is in a format allowing **immediate use**, is **versioned**, you will be using **version 4.2.1** of the dataset **“CERES Energy Balanced and Filled (EBAF) TOA Monthly means data in netCDF”**
 - If new version of the dataset is available, you will continue your work with it
 - Bonus round:
 - Who is the owner of the dataset?
 - What is the license of the dataset?
 - What is the format of the data? And the amount?

Task No. 2

- Complete the information in **Chapter 10 – Preserving data** according to the following:
 - You have only **one data preservation strategy** across all datasets
 - Data will be stored **on tape** on a **remote storage site** with **physical access protection** and a **clearly defined group** of people who have access
 - The data **do not need to be encrypted** and will be stored for **10 years**.
 - Cold storage will be maintained for **15 years** and will be paid for by one of the **project partners**, who is also **responsible for all cold-storage policies**.
 - The budget **allows for the extension** of cold-storage **if the data is being used**.
 - The budget **allows for updating storage media** if necessary but **will not change data formats**.

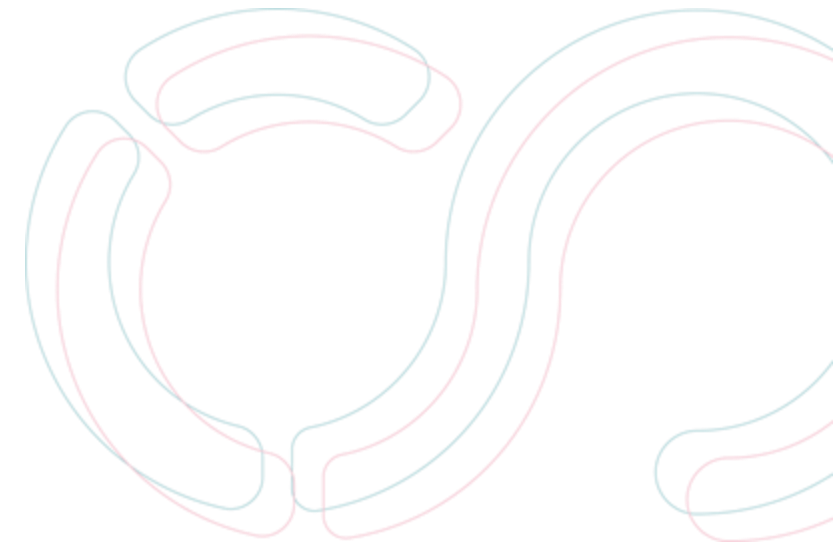
Completed DMPs: An Example

- [DMP Example: Alliance for Life Sciences: From Strategies to Actions in Central and Eastern Europe](#)
- [DMP Example: Ashy layers: Monjukli Depe](#)



Future Development of DSW NRP

- Modification of KM NRP
 - Better work with AAI
 - Translations of KM and templates
- Integration with NRP services and repositories
- Search available
- Cross-referencing questions



Future Development of DSW NRP

Questionnaire

Metrics

Preview

Documents

Settings

View

Import replies

Actions

TODOs 1

Comments

Version history

Search

Fullscreen

Current phase

Before Submitting the DMP

Chapters

I. Administrative information6

II. Re-using data1

Contributors

Research Project(s)

To execute the DMP, is additional spe...

Do you require hardware or software i...

Describe national / funder / sectorial / ...

I. Administrative information

✓ I.1 Contributors

Horizon 2020 DMPHorizon Europe DMPScience Europe DMPmaDMP

Each person contributing to creating or executing the data management plan should be added as a contributor. A project probably should have a Contact Person, and a Data Curator.

☒ Desirable: Before Submitting the DMP

✓ Albert Einstein

+ Add

re-use

Results for: re-use

Question

...rom elsewhere, and facilitate re-use of your own data later, it is...

Chapter

...ctice to check all options to re-use existing available data, eith... ..eful) make them available for re-use by others. And the circle islete if such data is actually re-used.

Chapter

...ion needed by people who will re-use your data, and with the acces...

Future Development of DSW NRP

✓ I.1 What datasets are you collecting?

⇅ Expand all ⇅ Collapse all

✓ **Butterfly specimen dataset**



✓ **Flowers dataset**



+ Add

Future Development of DSW NRP

Reference

 c9baed0a

 Move

 Delete

Reference Type

Cross Reference

Question

What datasets are you collecting?

[Go to related question](#)

Description

Consider all datasets you are collecting

Future Development of DSW NRP

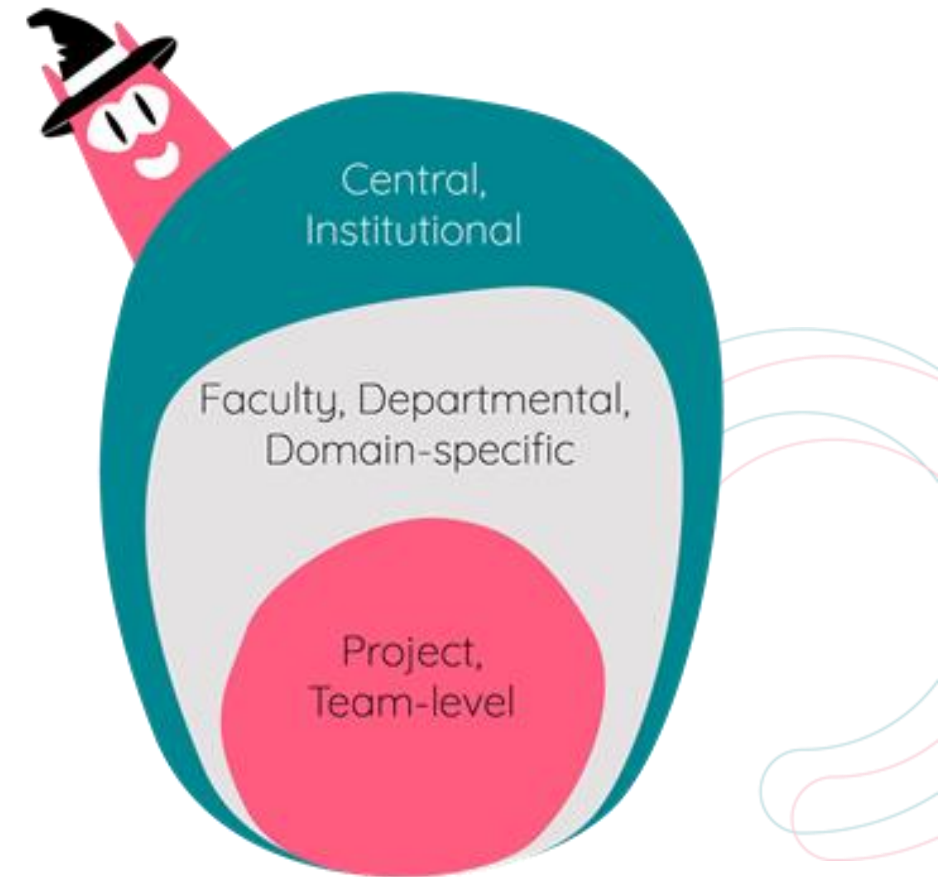
II.1 How big storage do you need (in GBs)?

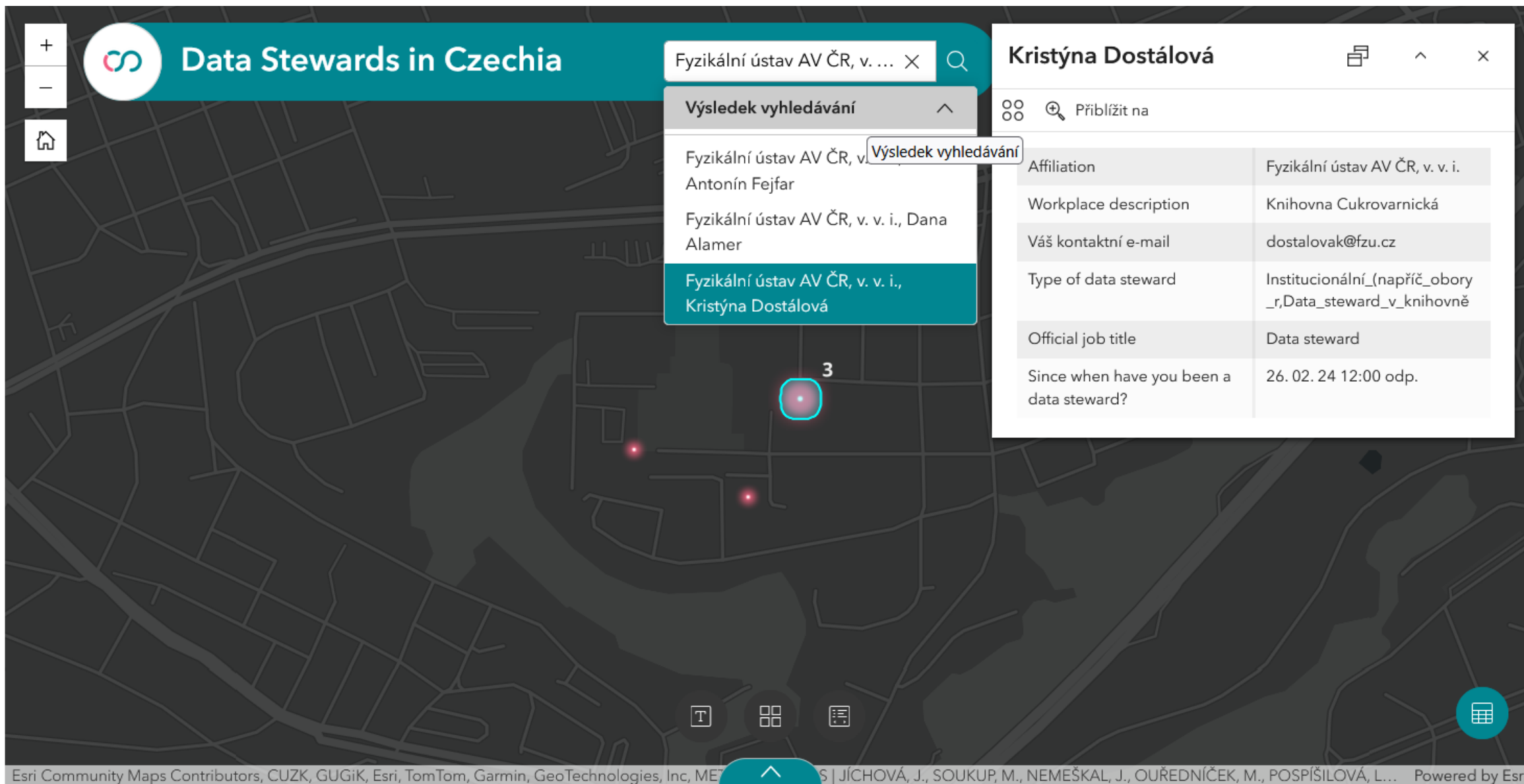
💬 Related questions: What datasets are you collecting? (Consider all datasets you are collecting)

New Research Role – Data Stewards

- Research data management specialist
- Bridge between researchers, institutions, and technology
- Ensures data quality, compliance with legislation and policies, increases research visibility, reduces risk of data loss, ...
- <https://www.eosc.cz/en/communities/data-stewards>

Types of Data Stewards





Data Stewards in Czechia

Fyzikální ústav AV ČR, v. ...

Výsledek vyhledávání

- Fyzikální ústav AV ČR, v. i., Antonín Fejfar
- Fyzikální ústav AV ČR, v. v. i., Dana Alamer
- Fyzikální ústav AV ČR, v. v. i., Kristýna Dostálová**

Kristýna Dostálová

Affiliation	Fyzikální ústav AV ČR, v. v. i.
Workplace description	Knihovna Cukrovarnická
Váš kontaktní e-mail	dostalovak@fzu.cz
Type of data steward	Institucionální_(napříč_obory_r,Data_steward_v_knihovně
Official job title	Data steward
Since when have you been a data steward?	26. 02. 24 12:00 odp.

Esri Community Maps Contributors, CUZK, GUGiK, Esri, TomTom, Garmin, GeoTechnologies, Inc, ME S | JÍCHOVÁ, J., SOUKUP, M., NEMEŠKAL, J., OUŘEDNÍČEK, M., POSPÍŠILOVÁ, L... Powered by Esri

Co od Vás očekáváme

- VŠ vzdělání (praxe se správou dat výhodou)
- Analytické myšlení
- Zájem o problematiku Open Science
- Znalost konceptu Open Science a požadavků na archivaci dat v souladu s principy FAIR výhodou
- Dobrá znalost angličtiny (min. úroveň B2)
- Výborné komunikační, organizační a prezentační dovednosti
- Pečlivost a spolehlivost
- Proaktivita a týmovost

Useful Links and Contacts

[EOSC CZ Website](#)



[EOSC CZ Newsletter](#)



- Any questions?
 - info@eossc.cz
- Ideas for a lecture or a training?
 - events@eossc.cz
- Get in touch with our PR
 - pr@eossc.cz



@EOSC Czech Republic



@eosccz.bsky.social

Thank you for your attention

E: zita@ics.muni.cz

E: info@eosc.cz

W: www.eosc.cz/en



Spolufinancováno
Evropskou unií



MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY