

Prepare your teams for effective data management (plans)

version 2

Georgia Koutentaki, MEng.

Data Steward for STEM fields, National Repository Platform, FIT ČVUT

Data Steward, Faculty/Institution, VŠCHT

This presentation was made with the kind help of A. Soldánová,
Data Steward for Natural Sciences, Library of the CAS, CZ



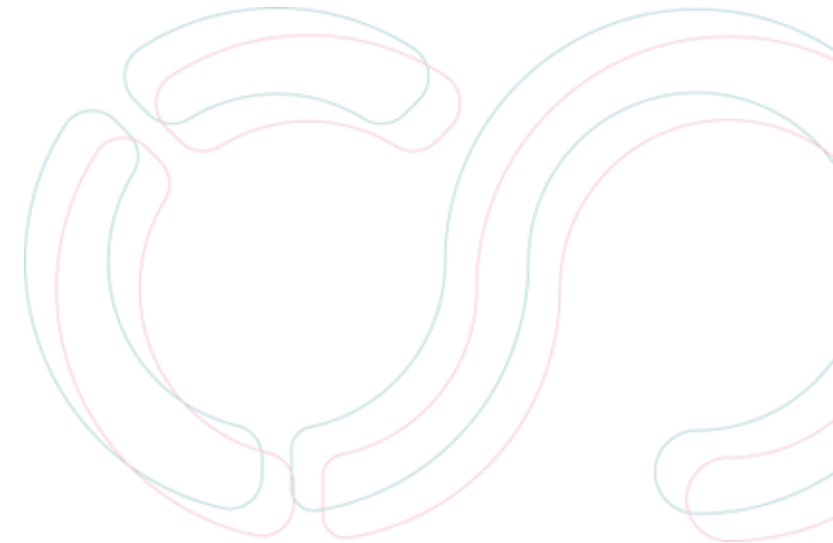
Co-funded by
the European Union



License: Except where otherwise noted, the text of this presentation is licensed under [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/). This license does not apply to third-party logos, branding, or system icons, which remain the property of their respective copyright holders.

Program

First part	9:00-9:50
Break	9:50-10:00
Second part	10:00-10:50
Break	10:50-11:00
Q & A	11:00-11:30



Introduction & basics of data management



Open science & data management vocabulary

- **OS:** open science
- **DS:** data steward
- **DMP:** data management plan
- **Metadata:** “data about the data”, aka documentation for the data
- **FAIR:** Findable, Accessible, Interoperable, Reusable
- **Metadata standard:** standardized structure for metadata, with specific list of requirements
- **CC or Creative Commons:** open licenses used for data and publications. CC is the abbreviation. *There are several CC licenses and several versions of them.*
- **PID:** persistent identifier is a long-lasting, unique digital reference to a person, place, or thing.

What do we mean when we say...

Data

- Numbers, dates, results, names, locations, images... received by machines and literature
- The (data) points of a graph – x, y, z, t... combos
- Personal data (name, date of birth, phone no., email, address etc.)
- Computational data (settings, inputs, outputs)
- and more 😊

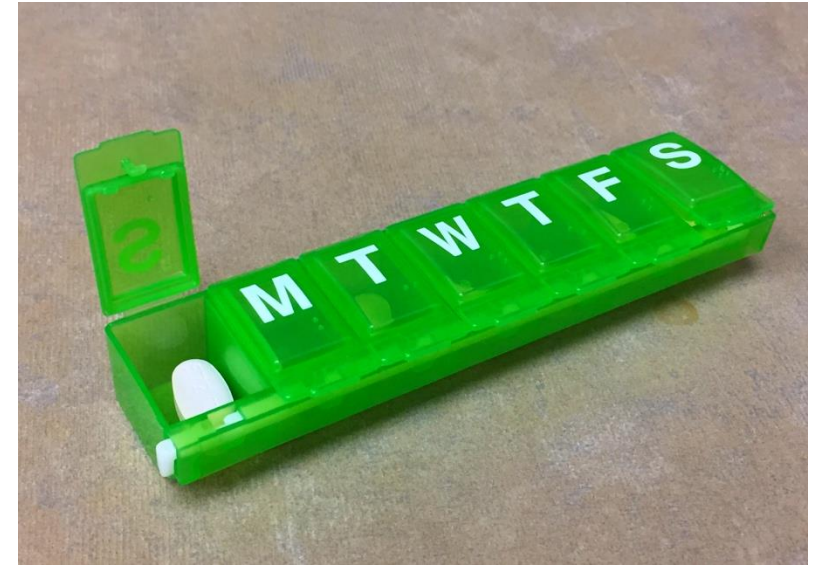
Metadata

- How/where did you get those data?
- Did you use specific software and instruments?
- Who collected the data and who analyzed them?
- Where are those data?
- Are they in a format that I can open in 10 years?
- Are the data available for reuse? Where?
- How can I open these data if I do not have your software?
- Can I share or adapt these data?

Data management

**Data management is an everyday task,
we just don't realize that we are doing it.**

It is hidden under *labeling* laboratory samples and *dating* them, adding our *name* on our sample boxes, *storing* results of a measurement in a school computer, *making a folder* for pictures from our family in our flash drive.



"Pill Box (32636946385).jpg" by NIAID, licensed under [CC BY 2.0](https://creativecommons.org/licenses/by/2.0/). Source: [Wikimedia Commons](https://commons.wikimedia.org/wiki/File:Pill_Box_(32636946385).jpg)

Data management – what do we mean in research?

Data management involves several steps:

- Finding out **what data** are you using or creating
- Creating a plan that shows **how** you would expect the data to be managed:
naming, dates, which data are you keeping, which data are you deleting
- **Deciding** which data will be **published** and how
- Calculating possible **costs** from operating and analyzing data, to making them available and open...

Data management plans (DMPs)

- Not just a funder requirement
- DMPs are and should be a **reflective, living document** for you and your team
- They should include **all management processes** related to data
- They can save you time and resources

How can DMPs be useful?

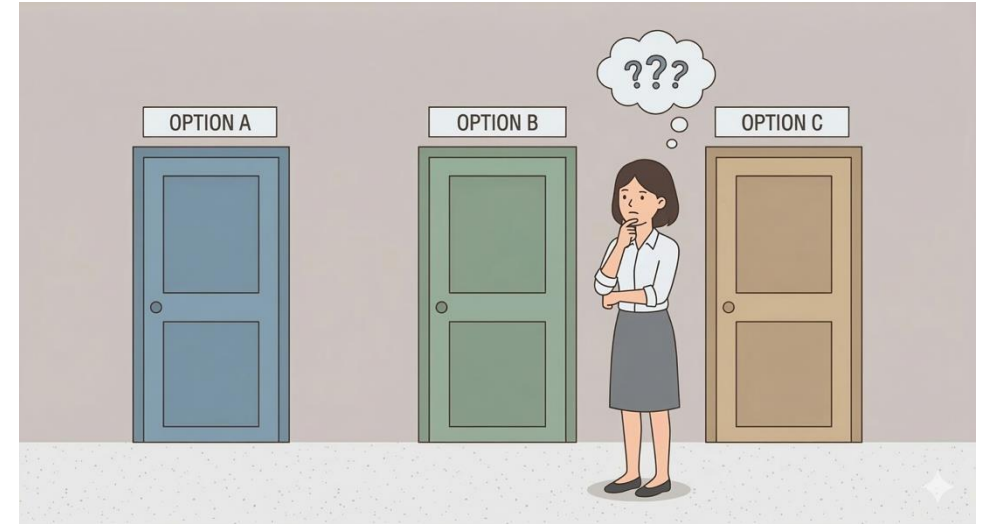


Scenario: *A funded project is running for 5 years. You start the project with 7 master and 3 bachelor students, 5 PhD students and 1 post-doc. You are the PI and the responsible person for reporting the outputs of the grant.*

On the 3rd year, the post-doc finds another university to do research in, and the 3 bachelor students are getting their diploma. Two of the PhD students are leaving too.

What are your next steps?

You look to hire new students, and especially a post-doc to help with the workload, and possibly PhD students that will be also supervising researchers in earlier stages.



- *How do you explain your project to them without losing information?*
- *How do you explain the data they find from previous work that they need to understand and work on?*
- *How will you make sure that GDL_Jan1 is the last sample of a student and not GDL1_Jan?*

Saving time and resources through DMPs

An internal DMP can save you as the team PI when it comes to **introducing new people in the group**

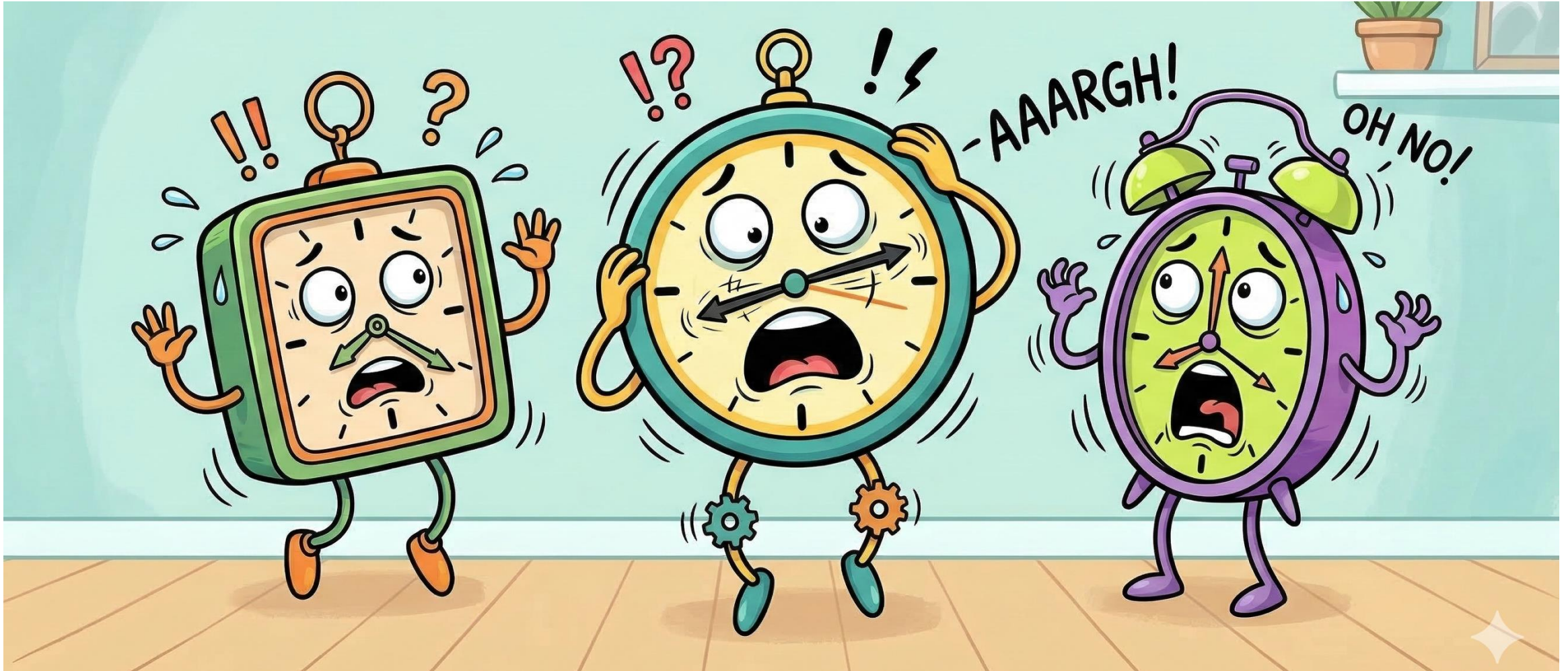
- Giving all the important information through DMP
- No need to have the whole team discuss with the researchers the moment you hire them
- All the researchers have now the same information and work in an organized and understandable way
- Any time you do not know what a folder is, you can go to the DMP and find out, even when the students and researchers are no longer in the team

Saving time and resources through DMPs

An internal DMP can save you as the team PI **when it comes to reporting**

- You have access to information related to all the students simply by looking the DMP
- Updates are essential – **A DMP must be updated often** internally, so everyone can keep up with the changes
- As the projects evolves, so does the DMP
- When funder reporting comes, you only need to review!

How to not freak out?



When to start preparing?



- When conceptualizing
- During the writing of proposals
- If not during the proposal writing, then **immediately** after acceptance of a grant

Not last minute

What to consider?



Get to know your OS
requirements



Get to know your team



Get to know your data

What to not skip in a team meeting



Inform them of the **OS requirements** for the project they work on



Give them **information** on possible **help** from the institute
(aka. Data Steward, open science office, project center)

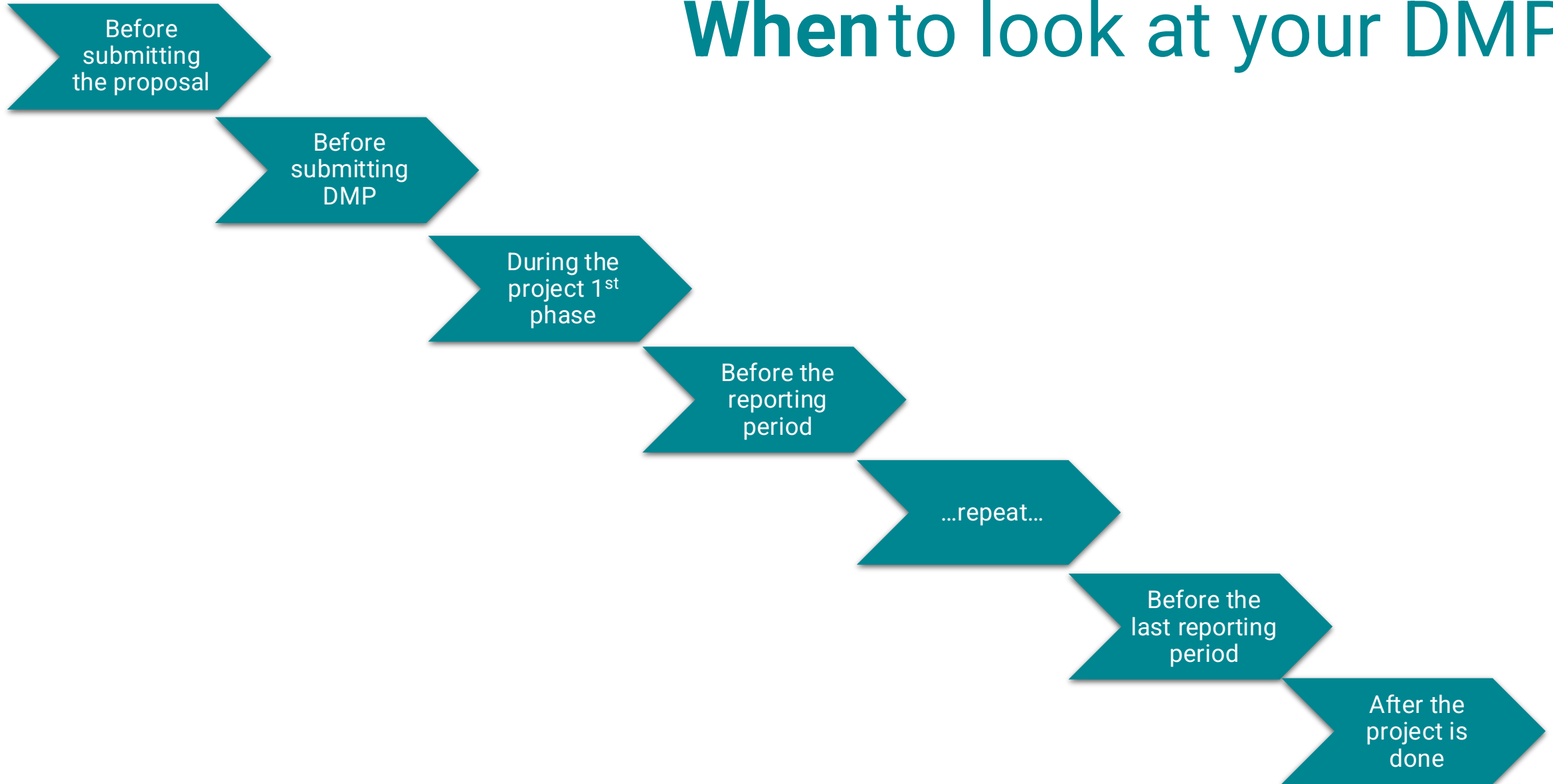


Give them the **DMP** with the information that you wish their data to have
(metadata, names, formats, publishing possibilities)

DMP timeline



When to look at your DMP ?



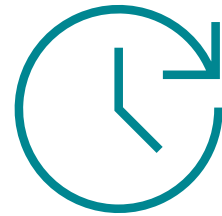
DMP after the project – for the manager



A detailed scheme of all data available from the project



Where are the data,
what are the data



An easy way to create new plans for future projects

DMP after the project – for the team



Easy access to
what was done



A guide for
publishing & thesis
organization



Correctly citing
people's work,
no "guessing"
who did what



Easy repeatability,
help with
continuation of works

Commonly used tools for DMP

Data Stewardship Wizard

Question and answer, collaborative environment, prefilled answers, personalization

(open source, hosted locally or hosted by an infrastructure)

FAIR Wizard

Extension on top of DSW, build in mind of RDM needs for organizations.

(software as a service in cloud environment)

ARGOS

Choose from a catalogue of funders, grants, and possible templates and answer questions for your DMP

(OpenAIRE service)

DMPOnline

Simple UI, reply to the main points that need to be in the DMP based on funders' requirements

(by the Digital Curation Centre (DCC))

Data Stewardship Wizard NRP - through e-INFRA CZ

(dmp.eosc.cz)

- New Knowledge Model (KM) and other functions developed specifically for the needs of the NRP & the Czech RDM infrastructure
- The new KM fulfills the needs of common funder requirements (HE, MŠMT, GACR, ...)
- Still in pilot mode, but available to all*

Data Management Planning within National Data Infrastructure (NDI)



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

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

[Start Planning](#)



Services and Support

Explore the services and support options tailored for different user groups and discover the possibilities available to you.



Data Management Planning

Efficient data management begins with a quality Data Management Plan (DMP). We will assist you in its creation and provide best practices for managing research data. With our guide, you will gain an overview of key principles and recommendations.

[Start Planning](#)

(pilot operation)



Documentation

The user guide is a comprehensive manual that covers all functionalities and options of the Data Stewardship Wizard. It includes detailed instructions, tips for effective use, and recommended practices. Whether you are a beginner or an advanced user, the documentation will help you.

[Read Guide](#)



FAIR Implementation Profiles

FAIR Implementation Profiles (FIP) help communities define how to apply the FAIR principles (Findable, Accessible, Interoperable, Reusable). We are currently preparing support for their creation within the NDI. Stay tuned for updates and get involved in building better research data management.

[In Preparation](#)



Training

Do you want to try the Data Stewardship Wizard in practice? We are preparing the option to create a training instance of DSW, where you can safely test data management planning. Soon you will have the opportunity to access a test environment for teaching and training.

[In Preparation](#)

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).

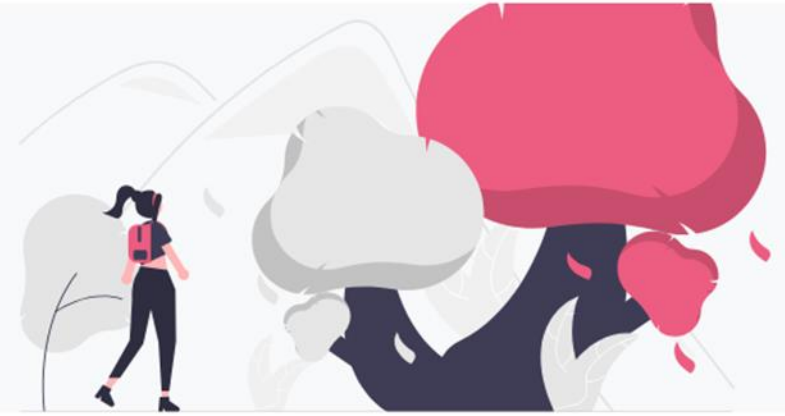
Log In

[Forgot your password?](#)

Or connect with

Welcome, [redacted] !

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

Collaborate and communicate with co-workers

The screenshot displays the DMP.eosc.cz web application interface. The main content area is titled "I. Research Projects" and contains a "Share Project" dialog box. The dialog box has a "Users" section with a search input and a list of users. Two users are listed: Georgia Koutentaki and Jana Martínková. The role dropdown for Jana Martínková is open, showing options: Owner, Viewer (highlighted), Commenter, Editor, and Owner. The dialog also includes checkboxes for "Visible by all other logged-in users" and "Public link", a "Copy link" button, and a "Done" button. In the top right corner of the application, a "Share" button is circled in red. The left sidebar shows a navigation menu with items like "Dashboard", "Projects", "eosc.cz", and "DSW Guide". The top navigation bar includes "Questionnaire", "Metrics", "Preview", "Documents", and "Settings".

Nested questions

IV.1.a.1 Label of this data collection and processing

Horizon Europe DMP

Please provide a label of this data collection and processing approach. This label is used solely to distinguish this entry within the questionnaire.

Desirable: Before Submitting the DMP

label 1

Clear answer

Answered less than a minute ago by Georgia Koutentaki.

IV.1.a.2 Is this data collection and processing approach consistent across all datasets you've produced?

Horizon Europe DMP

Desirable: Before Finishing the Project

a. No, different datasets were collected and processed using different approach

b. Yes, the same data collection and processing approach applies to all datasets

Clear answer

Answered less than a minute ago by Georgia Koutentaki.

To accurately describe the **data collection and processing approach for all datasets** across the project, please create **only one entry** to describe the approach.

IV.1.a.3 Start date of the data collection period

IV.1.a.1 Label of this data collection and processing

Horizon Europe DMP

Please provide a label of this data collection and processing approach. This label is used solely to distinguish this entry within the questionnaire.

Desirable: Before Submitting the DMP

label 1

Clear answer

Answered less than a minute ago by Georgia Koutentaki.

IV.1.a.2 Is this data collection and processing approach consistent across all datasets you've produced?

Horizon Europe DMP

Desirable: Before Finishing the Project

a. No, different datasets were collected and processed using different approach

b. Yes, the same data collection and processing approach applies to all datasets

Clear answer

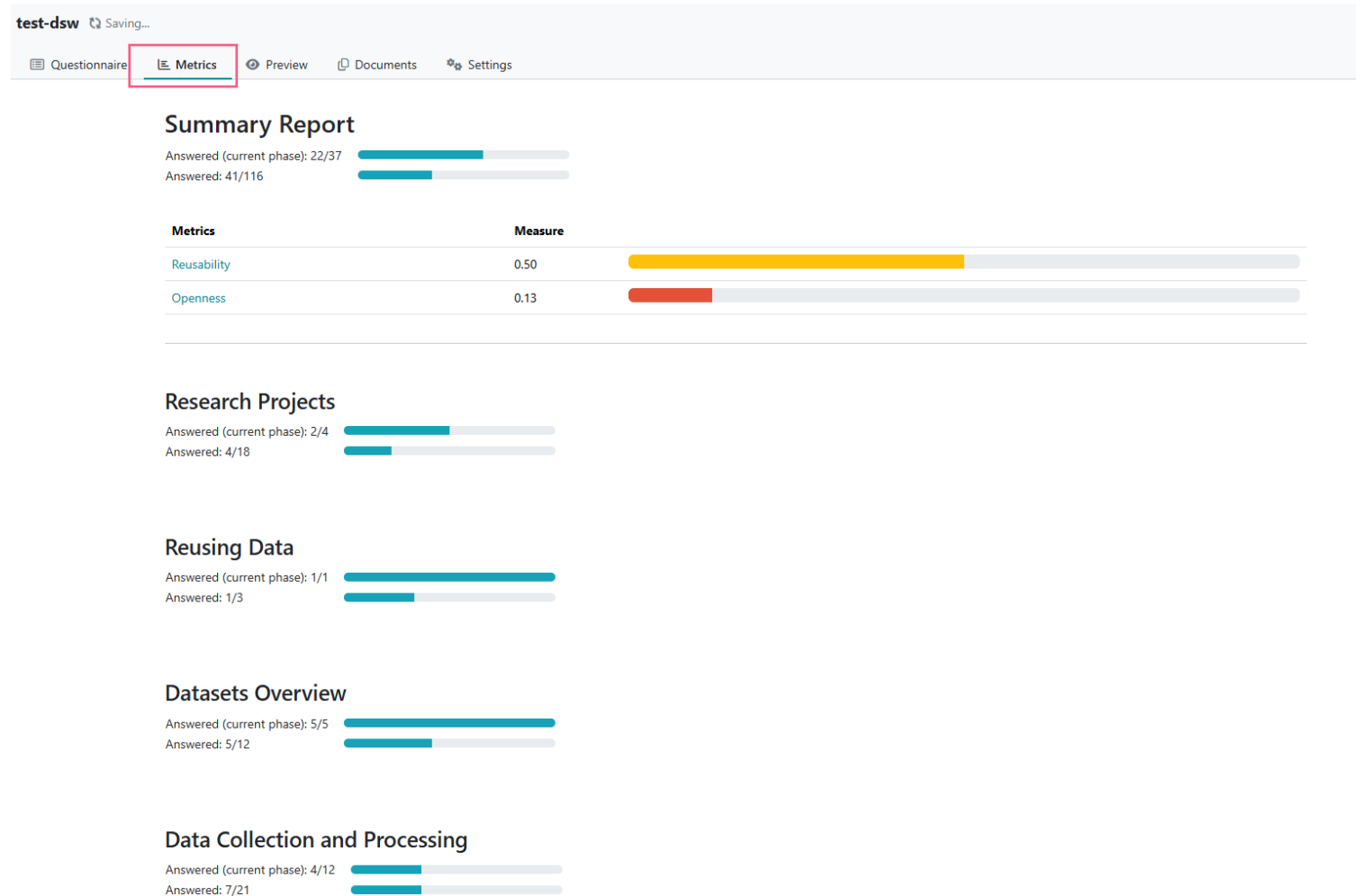
Answered less than a minute ago by Georgia Koutentaki.

To accurately describe the **data collection and processing approach for each dataset** (or group of datasets that share the same approach), please create a **separate entry for each distinct data collection and processing approach**.

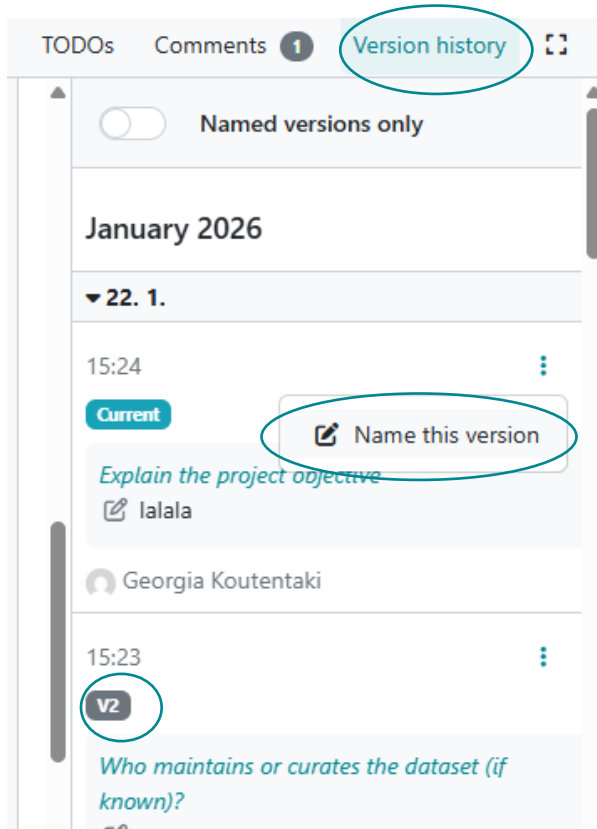
Collapse

IV.1.a.2.a.1 Please specify which datasets follow this process described here in this entry

Get metrics of FAIRness and more




Versioning DMPs



In DSW

History of changes



Version	Publication date	Changes
V2	22 Jan 2026	Addition of existing datasets info
V1_3	22 Jan 2026	Added metadata description
V1_2	22 Jan 2026	Addition of data formats
V1_1	22 Jan 2026	Initial version DMP

In the document

Funding Open science



Common requirements from funders



Open Access publications
under an open license



Submission of DMP
(usually 3x in the project)



Submission of datasets in a
trusted repository



Ethics mentioned in DMP

Common requirements from journals

- Data related to publication in a repository
- Data of publication in the journal's repository
- Pre-print in a repository related to the journal
- Data Availability Statement
- CRediT (Contributor Roles Taxonomy)

Common mistakes & questions



Common mistakes & questions

- Data formats for long term archiving
- Licensing of publications
- Persistent identifiers for data and publications
- Creating a record in a repository – publication and data together?
- Data in repository – all data for a paper together? Data vs. software repository.

File formats – Desirable file formats

Common requirements	Desirable? (examples)			
	NO/ Maybe		YES!	
1. complete and open documentation	Excel 2003 or older (*.xls)	Excel 2007 or newer (*.xlsx)	Comma Separated Values (*.csv)	Delimited Text (*.txt)
2. platform-independence				
3. non-proprietary	Executable files (.EXE)		Uncompiled computer program source code (*.c, *.java)	
4. no embedded files, programs or scripts				
5. no full or partial encryption	PDF (*.pdf) (encrypted)		PDF/A-1	
6. no password protection				

File formats for digital content: Probability for full long-term preservation

Content type	High	Medium	Low
Text	<ul style="list-style-type: none"> Plain text (encoding: USASCII, UTF-8, UTF-16 with BOM) XML (includes XSD/XSL/XHTML, etc.; with included or accessible schema) PDF/A-1 (ISO 19005-1) (*.pdf) 	<ul style="list-style-type: none"> Cascading Style Sheets (*.css) DTD (*.dtd) Plain text (ISO 8859-1 encoding) PDF (*.pdf) (embedded fonts) Rich Text Format 1.x (*.rtf) HTML (include a DOCTYPE declaration) SGML (*.sgml) Open Office (*.sxw/*.odt) OOXML (ISO/IEC DIS 29500) (*.docx) Microsoft Word 2007 or newer (*.docx) 	<ul style="list-style-type: none"> PDF (*.pdf) (encrypted) Microsoft Word 2003 or older (*.doc) WordPerfect (*.wpd) DVI (*.dvi) All other text formats not listed
Raster image	<ul style="list-style-type: none"> TIFF (uncompressed) JPEG2000 (lossless) (*.jp2) PNG (*.png) 	<ul style="list-style-type: none"> BMP (*.bmp) JPEG/JFIF (*.jpg) JPEG2000 (lossy) (*.jp2) TIFF (compressed) GIF (*.gif) Digital Negative DNG (*.dng) 	<ul style="list-style-type: none"> MrSID (*.sid) TIFF (in Planar format) FlashPix (*.fpx) PhotoShop (*.psd) RAW JPEG 2000 Part 2 (*.jpf, *.jpx) All other raster image formats not listed
Vector graphics	<ul style="list-style-type: none"> SVG (no Java script binding) (*.svg) 	<ul style="list-style-type: none"> Computer Graphic Metafile (CGM, WebCGM) (*.cgm) 	<ul style="list-style-type: none"> Encapsulated Postscript (EPS) Macromedia Flash (*.swf) All other vector image formats not listed
Audio	<ul style="list-style-type: none"> AIFF (96kHz 16bit PCM) (*.aif, *.aiff) WAV (96kHz 24bit PCM) (*.wav) 	<ul style="list-style-type: none"> SUN Audio (uncompressed) (*.au) Standard MIDI (*.mid, *.midi) Ogg Vorbis (*.ogg) Free Lossless Audio Codec (*.flac) Advance Audio Coding (*.mp4, *.m4a, *.aac) MP3 (MPEG-1/2, Layer 3) (*.mp3) 	<ul style="list-style-type: none"> AIFC (compressed) (*.aifc) NeXT SND (*.snd) RealNetworks 'Real Audio' (*.ra, *.rm, *.ram) Windows Media Audio (*.wma) Protected AAC (*.m4p) WAV (compressed) (*.wav) All other audio formats not listed
Video	<ul style="list-style-type: none"> Motion JPEG 2000 (ISO/IEC 15444-4)??*.mj2) AVI (uncompressed/native, motion JPEG) (*.avi) QuickTime Movie (uncompressed/native, motion JPEG) (*.mov) 	<ul style="list-style-type: none"> Ogg Theora (*.ogg) MPEG-1, MPEG-2 (*.mpg, *.mpeg, wrapped in AVI, MOV) MPEG-4 (H.263, H.264) (*.mp4, wrapped in AVI, MOV) 	<ul style="list-style-type: none"> AVI (others) (*.avi) QuickTime Movie (others) (*.mov) RealNetworks 'Real Video' (*.rv) Windows Media Video (*.wmv) All other video formats not listed

Data formats for long term archiving – sources

ETH-Bibliothek, Fachstelle Forschungsdatenmanagement und Datenerhalt (FDD). Confluence of ETH Zurich, [File formats for archiving](#)

Library of Congress, [Recommended Formats Statement 2025-2026](#)

MIT Libraries. [File formats for long-term access](#)

Cornell University Research Data Management Service Group. [File formats for preservation](#)



**BREAK
TIME**

Common questions – data

- How to upload a large data set? – Curate, Minimize, Repository
- Who owns my data? – Contracts
- Does my university accept my deposit of data online? Data steward
- Does my industry collaborator need to know the deposit of data online? YES
- I do not know if I work with sensitive data – Data steward & Team
- What is metadata??? – Data steward

Let's look at some data

Disclaimer: This is a **modified excerpt** of the original dataset. It contains a **selection of original data points** alongside **fictionalized entries** created for demonstration purposes.. For the authentic dataset, please refer to the original resource: **Data source:** Doyle, L. (2021). Fluorescent polarization data for Figure 7 [Data set]. Harvard Dataverse. <https://doi.org/10.7910/DVN/ROSFLC>



Example dataset: “Fluorescent Polarization Data for Figure 7: Sheet1”

Working Conc. (uM)	Final Conc. (uM)		SH2 (+His)	24x_25/28_sub8_3	24x25/28_3_2SH 2	Tir10			
23.000	20.700	mg/mL	5.2	18.6	17.4	Stock (mM)	5.7		
11.500	10.350	Stock (uM)	381.1	112.0	69.0	1:1000 dilution(uM)	5.7		
5.750	5.175					Final conc (nM)	50		
2.875	2.588	To get [A2]	57	194	79				
1.438	1.294	Buffer	888	751	866				
0.719	0.647		945	945	945				
0.359	0.323								
0.180	0.162					To get 0.5 uM	43	86	129
0.090	0.081					Buffer	447	894	1341
0.045	0.040						490	980	1470
0.022	0.020		Dil. prot. (uL)	0.5 uM Tir10		# FP runs	1	2	3
0.011	0.010	3.5 rxn	315	35	350				
0.006	0.005								
0.003	0.003								
		Mw	Stock (uM)	Work conc. (uM)	Final conc. (uM)				
		SH2	13645	381.1	23.000	20.700			
		24x_25/28_sub8_3	166113	112.0	23.000	20.700			
		24x25/28_3_2SH 2	252240	69.0	5.750	5.175			

Data Version II: One workbook SEMI-Vertical orientation

Choosing one workbook with rows to indicate separate parts of analysis and data

Organized data – sheet 1

row number

1	Dissolution measurements					
2		Working Conc. from absorbance at 200 nm	Final= Working – Initial Conc. (uM)			
3	time (min)	Working Conc. (uM)	Final Conc. (uM)			
4	0	23.000	20.700			
5	5	11.500	10.350			
6	10	5.750	5.175			
7	15	2.875	2.588			
8	20	1.438	1.294			
9	25	0.719	0.647			
10	30	0.359	0.323			
11	35	0.180	0.162			
12	40	0.090	0.081			
13	45	0.045	0.040			
14	50	0.022	0.020			
15	55	0.011	0.010			
16	60	0.006	0.005			
17	65	0.003	0.003			
18						
19						
20						
21	Calculating concentration of samples					
22	Sample name [A2]	mg/mL	Mw g/mol	Stock (uM)	Work conc. (uM)	Final conc. (uM)
23	SH2	5.2	13645	381.1	23.000	20.700
24	24x_25/28_sub8_3	18.6	166113	112.0	23.000	20.700
25	24x25/28_3_2SH2	17.4	252240	69.0	5.750	5.175
				:		
				(next slide for next rows)		

Organized data – sheet 1

row number



26		:			
27					
28					
29	Calculating volumes for samples				
30	To get [A2] mL	Buffer mL	Total vol. mL		
31	57	888	945		
32	194	751	945		
33	79	866	945		
34					
35					
36					
37	Calculating concentration of samples				
38	Sample	Stock Conc. (mM)	Stock Conc. after 1:1000 dilution (uM)	Final Conc. (nM)	
39	Tir10	5.7	5.7	50	
40					
41					
42					
43	Calculating volumes for samples				
44	To get 0.5 uM Tir10	Buffer mL	Total vol. mL		
45	43	447	490		
46	86	894	980		
47	129	1341	1470		

Common questions – Metadata



What is it?



How to create metadata for my project?

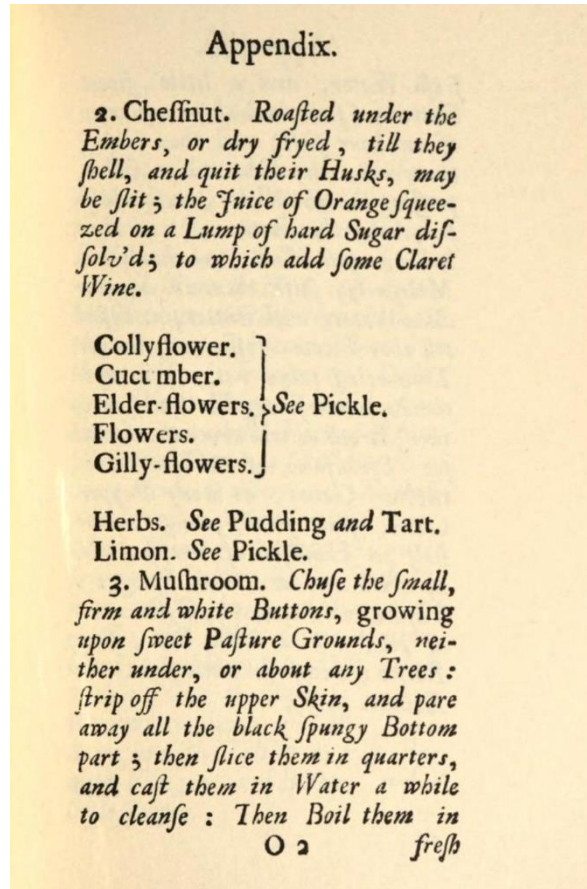


What do I need to include?



How does it look like?

Metadata



Examples of metadata

1. Name of authors /collectors/ rightsholders
2. Contact person
3. Abstract
4. Explanation of names of the files
5. Methodology
6. Terms of use (License)
7. ...

Example Data & Metadata

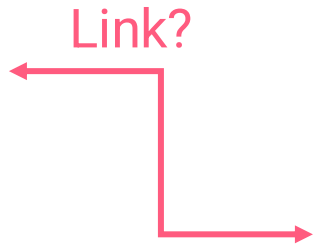
Data

Working Conc. (uM)	Final Conc. (uM)		SH2 (+His)	24x_25/28_sub8_3	24x25/28_3_2SH2	
23.000	20.700	mg/mL	5.2	18.6	17.4	
11.500	10.350	Stock (uM)	381.1	112.0	69.0	
5.750	5.175					
2.875	2.588	To get [A ₂]	57	194	79	
1.438	1.294	Buffer	888	751	866	
0.719	0.647		945	945	945	
0.359	0.323					
0.180	0.162					
<ul style="list-style-type: none"> SH2 (medicine): Src homology 2 domains - protein structure SH2 (chemistry): H₂S hydrogen sulfide – colorless gas 						
0.006	0.005				350	
0.003	0.003					
		Mw	Stock (uM)	Work conc. (uM)	Final conc. (uM)	
		SH2	13645	381.1	23.000	20.700
		24x_25/28_sub8_3	166113	112.0	23.000	20.700
		24x25/28_3_2SH2	252240	69.0	5.750	5.175

Naming of sheets: e.g. 20190510_24x2528-2SH2

Example Data & Metadata

Data



Metadata

Fluorescent Polarization Data for Figure 7 (doi:10.7910/DVN/ROSFLC)
<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ROSFLC>

View:	Part 1: Document Description Part 2: Study Description Part 3: Data Files Description Part 4: Variable Description Entire Codebook
Document Description Citation	
Title:	Fluorescent Polarization Data for Figure 7
Identification Number:	doi:10.7910/DVN/ROSFLC
Distributor:	Harvard Dataverse
Date of Distribution:	7/9/2021
Version:	1
Bibliographic Citation:	Doyle, Lindsey, 2021, "Fluorescent Polarization Data for Figure 7", https://doi.org/10.7910/DVN/ROSFLC , Harvard Dataverse, V1, UNF:6:PhSdiPa5rzTex2tZaPP6FQ== [fileUNF]
Study Description Citation	
Title:	Fluorescent Polarization Data for Figure 7
Identification Number:	doi:10.7910/DVN/ROSFLC
Authoring Entity:	Doyle, Lindsey (Fred Hutchinson Cancer Research Center)
Distributor:	Harvard Dataverse
Access Authority:	Doyle, Lindsey
Depositor:	Doyle, Lindsey
Date of Deposit:	7/9/2021
Holdings Information:	https://doi.org/10.7910/DVN/ROSFLC
Study Scope	
Keywords:	Medicine, Health and Life Sciences Raw Fluorescent Polarization Data for Figure 7
Abstract:	
Methodology and Processing	
Sources Statement	
Data Access	
Notes:	http://creativecommons.org/publicdomain/zero/1.0/ >CC0 1.0
Other Study Description Materials	
File Description--f4884994	
File: Figure7_FluorPolarize.tab	
	Number of cases: 35
	No. of variables per record: 13
	Type of File: text/tab-separated-values
Notes:	UNF:6:PhSdiPa5rzTex2tZaPP6FQ==

Variable Description	WorkingConc.(uM) - WorkingConc.(uM) FinalConc.(uM) - FinalConc.(uM) C - C D - D SH2(+His) - SH2(+His) 24x25/28sub83 - 24x25/28sub83 24x25/2832SH2 - 24x25/2832SH2 H - H I - I Tir10 - Tir10 K - K L - L M - M
List of Variables:	
Variables	
WorkingConc.(uM)	Summary Statistics: StDev 6.529276806788478; Max. 23.0; Valid 14.0; Min. 0.0028076171875; Mean 3.2855137416294635
f4884994 Location:	Variable Format: numeric
FinalConc.(uM)	Notes: UNF:6:DWVGzsTb6rAQnQhDCP PWig==
f4884994 Location:	Summary Statistics: Valid 14.0; Min. 0.00252685546875; Max. 20.7; StDev 5.876349126109631; Mean 2.9569623674665175
f4884994 Location:	Variable Format: character

Example Data & Metadata

Data

Link?

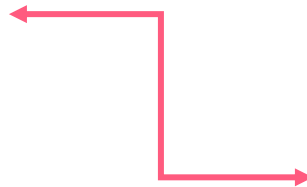


Figure 7 of what?

Where is this Figure 7 shown?

Is this connected to an article?

Fluorescent Polarization Data for Figure 7 (doi:10.7910/DVN/ROSFLC) https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ROSFLC	
View:	Part 1: Document Description Part 2: Study Description Part 3: Data Files Description Part 4: Variable Description Entire Codebook
Document Description Citation	
Title:	Fluorescent Polarization Data for Figure 7
Identification Number:	doi:10.7910/DVN/ROSFLC
Distributor:	Harvard Dataverse
Date of Distribution:	7/9/2021
Version:	1
Bibliographic Citation:	Doyle, Lindsey, 2021, "Fluorescent Polarization Data for Figure 7", https://doi.org/10.7910/DVN/ROSFLC , Harvard Dataverse, V1, UNF:6:PhSdiPa5rzTex2tZaPP6FQ== [fileUNF]
Study Description Citation	
Title:	Fluorescent Polarization Data for Figure 7
Identification Number:	doi:10.7910/DVN/ROSFLC
Authoring Entity:	Doyle, Lindsey (Fred Hutchinson Cancer Research Center)
Distributor:	Harvard Dataverse
Access Authority:	Doyle, Lindsey
Depositor:	Doyle, Lindsey
Date of Deposit:	7/9/2021
Holdings Information:	https://doi.org/10.7910/DVN/ROSFLC
Study Scope	
Keywords:	Medicine, Health and Life Sciences Raw Fluorescent Polarization Data for Figure 7
Abstract:	
Methodology and Processing	
Sources Statement	
Data Access	CC0 1.0
Notes:	
Other Study Description Materials	
File Description--f4884994	
File: Figure7_FluorPolarize.tab	Number of cases: 35
	No. of variables per record: 13
	Type of File: text/tab-separated-values
Notes:	UNF:6:PhSdiPa5rzTex2tZaPP6FQ==

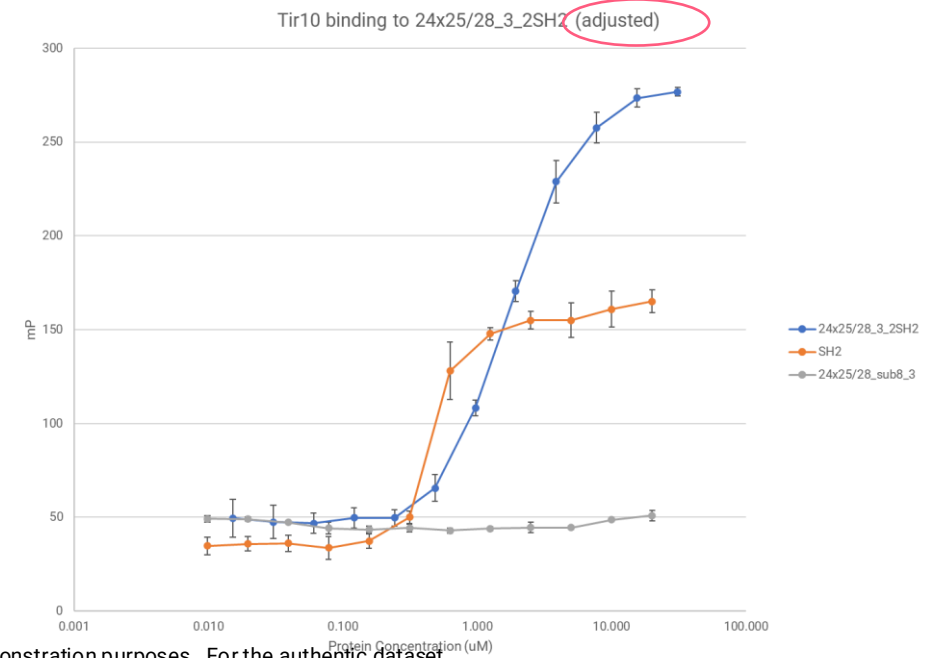
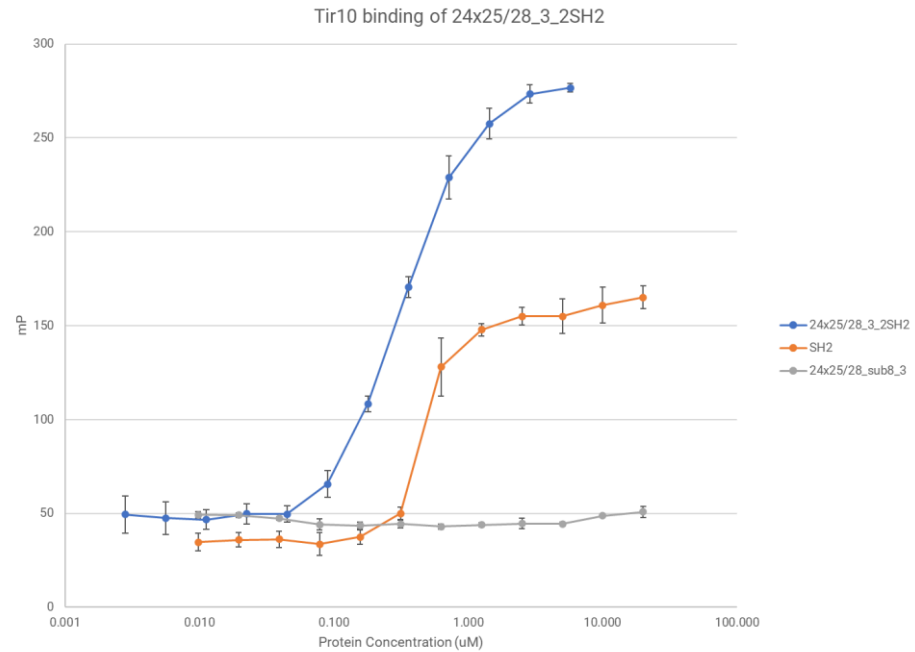
Variable Description	WorkingConc.(uM) - WorkingConc.(uM) FinalConc.(uM) - FinalConc.(uM) C - C D - D SH2(+His) - SH2(+His) 24x25/28sub83 - 24x25/28sub83 24x25/2832SH2 - 24x25/2832SH2 H - H I - I Tir10 - Tir10 K - K L - L M - M
List of Variables:	
Variables	
WorkingConc.(uM)	Summary Statistics: StDev 6.529276806788478; Max. 23.0; Valid 14.0; Min. 0.0028076171875; Mean 3.2855137416294635
f4884994 Location:	Variable Format: numeric
FinalConc.(uM)	Notes: UNF:6:DWVGzsTb6rAQnQhDCP PWig==
f4884994 Location:	Summary Statistics: Valid 14.0; Min. 0.00252685546875; Max. 20.7; StDev 5.876349126109631; Mean 2.9569623674665175
f4884994 Location:	Variable Format: character

Example Data & Metadata

Data

Which one is supposed to be Figure 7?

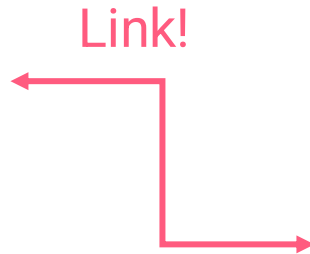
What is the second figure adjusted to?



Example Data & Metadata

Data

Working Conc. (uM)
23.000
11.500
5.750
2.875
1.438
0.719
0.359
0.180
0.090
0.045
0.022
0.011
0.006
0.003



Fluorescent Polarization Data for Figure 7 (doi:10.7910/DVN/ROSFLC) https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/ROSFLC	
View:	Part 1: Document Description Part 2: Study Description Part 3: Data Files Description Part 4: Variable Description Entire Codebook
Document Description Citation	
Title:	Fluorescent Polarization Data for Figure 7
Identification Number:	doi:10.7910/DVN/ROSFLC
Distributor:	Harvard Dataverse
Date of Distribution:	7/9/2021
Version:	1
Bibliographic Citation:	Doyle, Lindsey, 2021, "Fluorescent Polarization Data for Figure 7", https://doi.org/10.7910/DVN/ROSFLC , Harvard Dataverse, V1, UNF:6:PhSdiPa5rzTex2tZaPP6FQ== [fileUNF]
Study Description Citation	
Title:	Fluorescent Polarization Data for Figure 7
Identification Number:	doi:10.7910/DVN/ROSFLC
Authoring Entity:	Doyle, Lindsey (Fred Hutchinson Cancer Research Center)
Distributor:	Harvard Dataverse
Access Authority:	Doyle, Lindsey
Depositor:	Doyle, Lindsey
Date of Deposit:	7/9/2021
Holdings Information:	https://doi.org/10.7910/DVN/ROSFLC
Study Scope	
Keywords:	Medicine, Health and Life Sciences
Abstract:	Raw Fluorescent Polarization Data for Figure 7
Methodology and Processing	
Sources Statement	
Data Access	
Notes:	CC0 1.0
Other Study Description Materials	
File Description--f4884994	
File: Figure7_FluorPolarize.tab	
Notes:	Number of cases: 35 No. of variables per record: 13 Type of File: text/tab-separated-values UNF:6:PhSdiPa5rzTex2tZaPP6FQ==

Variable Description	
	WorkingConc.(uM) - WorkingConc.(uM) FinalConc.(uM) - FinalConc.(uM) C - C D - D SH2(+His) - SH2(+His) 24x25/28sub83 - 24x25/28sub83 24x25/2832SH2 - 24x25/2832SH2 H - H I - I Tir10 - Tir10 K - K L - L M - M
List of Variables:	
	WorkingConc.(uM)
Variables	
	WorkingConc.(uM)
	Summary Statistics: StDev 6.529276806788478; Max. 23.0; Valid 14.0; Min. 0.0028076171875; Mean 3.2855137416294635
f4884994 Location:	Variable Format: numeric
	Notes: UNF:6:DWVGzsTb6rAQnQhDCP PWig==

Example Data & Metadata

Data

Metadata

What was missing?

- Description of what is this Figure 7
- PID of other connected works (article, preprint, internal research)
- Explanation of the work done in the worksheet
- Differentiation of the two charts in the data worksheet
- Documentation of methodology
- What are the different columns in the different sheets of the file

Readme: How would it look like?

Authors: [names +ORCID]

Contributors: people & affiliations with ORCID preferably

Contact person: name & email

License: The data are provided as public domain.

Data file: Figure7_FluorPolarize_version1.xlsx

Readme: How would it look like?

Authors: [names +ORCID]

Contributors: people & affiliations with ORCID preferably

Contact person: name & email

License: The data are provided as public domain.

Data file: [Figure7_FluorPolarize_version1.xlsx](#)

[Figures are part of the publication \[doi\]. Figure7 title: Fluorescent Polarization](#)

Readme: How would it look like?

The data structure in the data file “Figure7_FluorPolarize_version1.xlsx” is based on row separation for different sets of data. **All values are in SI system.**

Rows discussing the same experiment or substance are placed one next to another, while for rows containing data from separate experiments and or substances separation is done **with 3 blank rows between them.**

Each row set includes several columns with titles. Each main experiment also includes a unique title in column A above the rows including the data, without any other columns filled.

For example, cell A1 includes the first main experiment called “Dissolution measurements”. Row 2 includes notes on columns, and rows 3 to 17 include column titles and data. The next experiment called “Calculating concentration of samples” follows up after 3 blank rows, at row 21.

All columns have a title explaining the values of the below rows.

Readme: How would it look like?

Abbreviation

[A2]
SH2

Explanation:

Sample group A trail 2
Src Homology 2 domains

Naming convention of samples:

ax_b/c_yi_j
b
c
subi
j

“ax”: x=hours, where a=1,2,3....
Test number
Total number of tests
sub= subject, i=number of subject
Repetition number where j=1,2 or 3.

EXAMPLE:

24x_25/28_sub8_3

24hours_test 25 of 28_subject8_repetition3

Data Re-use



Reused data – Common Issues

“License??
I don’t see it,
so I’ll just use it”

BIG no-no

All data
for one
publication
together

Not ideal for
versioning,
metadata,
or licensing

Data & publication
together

Not clear
which PID is
for what

Data & software
together

Not all licenses
fit everything

Reused data – solutions of issues

Prefer **open data with license** than those without one

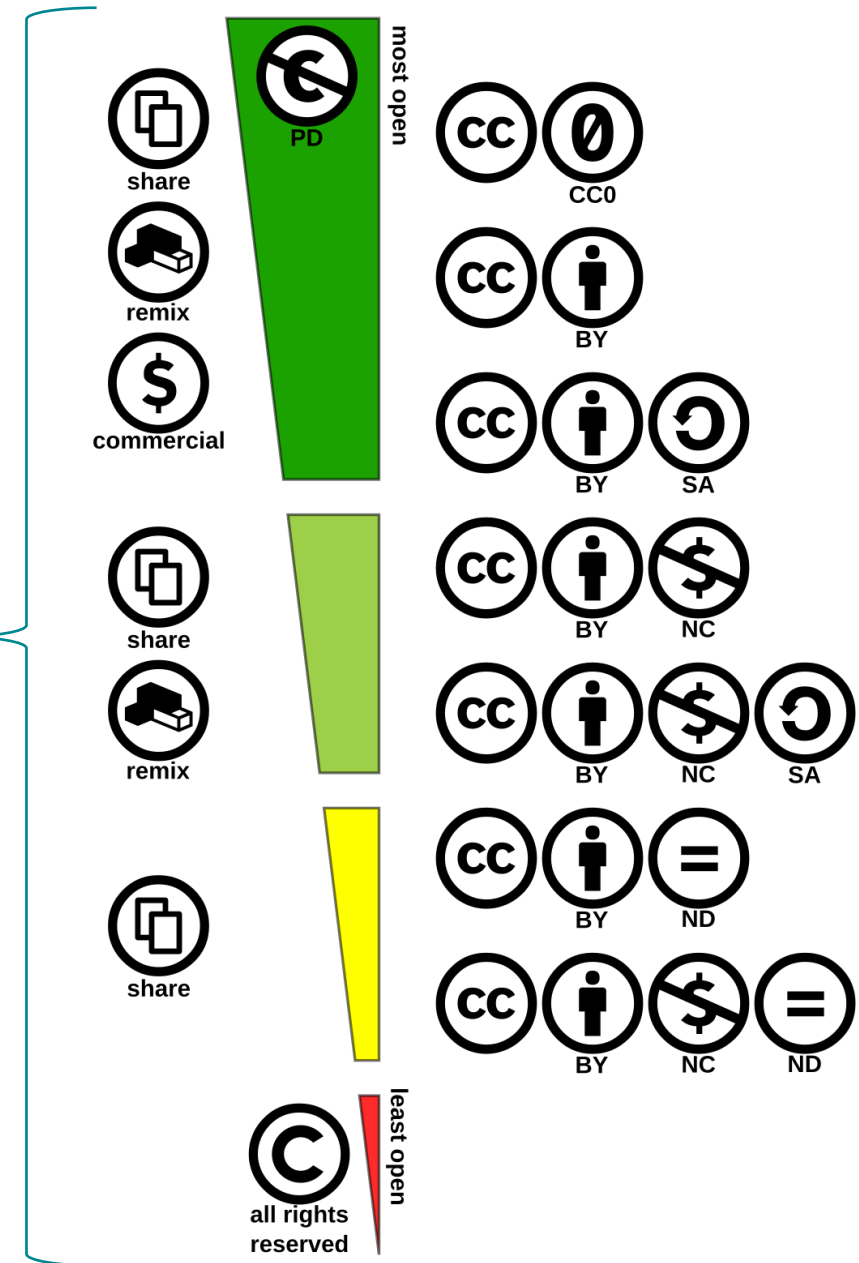
Separate **license, PID and metadata** for each data set

Separate **license** for data and software

Data **linked** to related open-source software

Licenses

- Several options exist
- **Creative Commons** are some of the most used open licenses today for research
- Public domain: free use, no attribution needed
- Software and data have different licenses
- Data in same project can have different license
- A publication and the relevant data can and should have separate license



Licenses for data

Share your work

Use Creative Commons tools to help share your work. Our free, easy-to-use copyright licenses provide [a simple, standardized way to give your permission to share and use your creative work](#) — on conditions of your choice. You can adopt one of our licenses by [sharing on a platform](#), sharing your work with [an open license](#), or dedicating your work to [the public domain](#).

Choose a license

[Our chooser](#) helps you determine which Creative Commons License is right for you in a few easy steps. If you are new to Creative Commons, you may also want to read [licensing considerations](#) before you [get started with the chooser](#) >

Share your work on a CC platform

Over 2.5 billion CC-licensed works exist across millions of websites. The majority are hosted on content platforms that provide CC license options for their users.

Licenses and Tools

[About CC Licenses](#)

[Use & remix](#)

[FAQ](#)

[Technology Platforms](#)

[Made with Creative Commons](#)

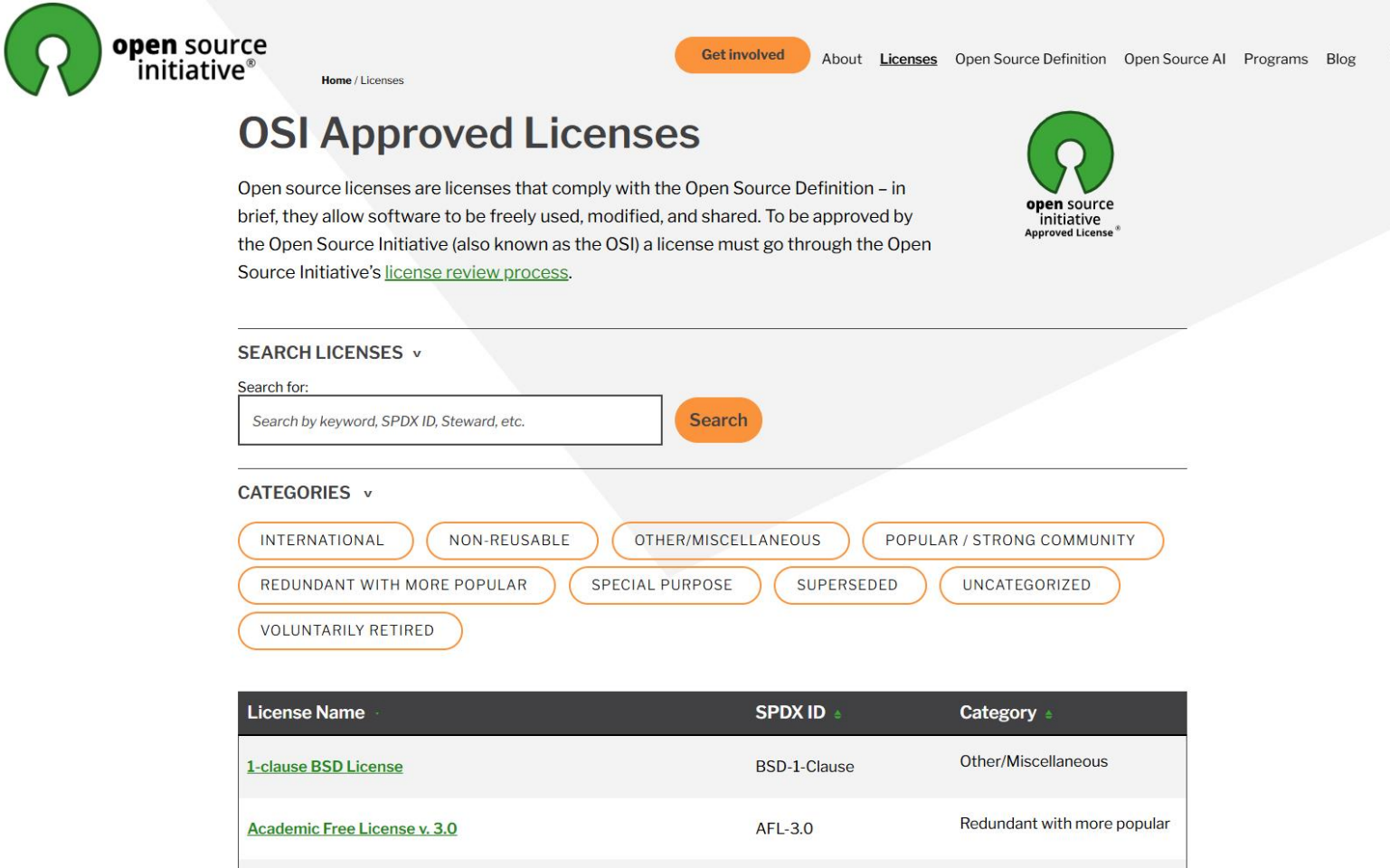
[Made with CC Community](#)

[Search the Commons](#)

- <https://opendatacommons.org/licenses/>
- <https://data.europa.eu/en/academy/open-data-licensing>
- <https://creativecommons.org/share-your-work/>



Licenses for software



open source initiative®

Home / Licenses

Get involved About Licenses Open Source Definition Open Source AI Programs Blog

OSI Approved Licenses

Open source licenses are licenses that comply with the Open Source Definition – in brief, they allow software to be freely used, modified, and shared. To be approved by the Open Source Initiative (also known as the OSI) a license must go through the Open Source Initiative's [license review process](#).

SEARCH LICENSES ▾

Search for:

Search by keyword, SPDX ID, Steward, etc. Search

CATEGORIES ▾

INTERNATIONAL NON-REUSABLE OTHER/MISCELLANEOUS POPULAR / STRONG COMMUNITY

REDUNDANT WITH MORE POPULAR SPECIAL PURPOSE SUPERSEDED UNCATEGORIZED

VOLUNTARILY RETIRED

License Name	SPDX ID	Category
1-clause BSD License	BSD-1-Clause	Other/Miscellaneous
Academic Free License v. 3.0	AFL-3.0	Redundant with more popular

Open Source initiative
<https://opensource.org/licenses>

Open source licenses are licenses that comply with the Open Source Definition – in brief, they allow software to be freely used, modified, and shared.

To be approved by the Open Source Initiative a license must go through the Open Source Initiative's license review process.

Naming conventions

You wouldn't name your child finalfinal2, would you?

final/

final3

final55@jnew_final

final2

finalFinal4

Naming conventions

The power of file naming

Naming conventions

➤ Example 1:

YYYYMMDD_Concrete_strength_raw	-contains raw data of concrete samples strength in their original data format, in MPa SI units
YYYYMMDD_Concrete_strength_raw_convert	-contains raw data in a long-term preservation data format, in MPa SI units
YYYYMMDD_Concrete_strength_analysed	- contains analysed data in a long-term preservation format, in MPa SI units

➤ Example 2:

Spectra_XRPD_HW1(VZ_Y_XXX)_NS_analysed_vi(j)

Spectra = type of data

XRPD = X Ray Powder Diffraction

NS = NameSurname initials of data collector

vi(j) = version i.j

HW1 = Hardware1: e.g. Automated Multipurpose Powder X-Ray Diffractometer: XRDynamic 500 (Anton Paar GmbH, Graz, Austria).

(VZ_Y_XXX) = Software version Z.Y.XXX

How to stay up-to-date?



What to learn?



Which data can my team
reuse and under which
conditions?



“What data do
we produce?”



What data do we
need/want to submit?
Raw, analyzed, curated?

Possible trainings for the team



Mandatory workshop on data management for team members



3-month data check with all the team



Individual short DMPs for team members or sub-teams



A specialized person to deal with data management practices

How can a project data steward help?



Intermediate
between PI,
researchers,
stakeholders



Focused on
data



Create RDM
document for
the team



Naming
conventions



Check data
formats



**BREAK
TIME**

Help for Institutes



Help for institutes – Data Stewards' map

MAP OF DATA STEWARDS IN CZECH REPUBLIC

Stats Join us! About cz

Search for NAME or ORGANISATION

Adrian Rosinec (Masarykova univerzita)

Aleš Mučka (Vysoká škola báňská - Technická)

Aneta Pilátová (Masarykova univerzita)

Anna Soldánová (Knihovna AV ČR, v. v. i.)

Antonín Fejfar (Fyzikální ústav AV ČR, v. v. i.)

Arzuv Čaryjeva (Ústav experimentální botaniky)

Arzuv Čaryjeva (Mikrobiologický ústav AV ČR, v. v. i.)

Barbora Gurecká (Etnologický ústav AV ČR, v. v. i.)

Barbora Krutáková (Všeobecná fakultní)

Barbora Kubátová (Univerzita Hradec Králové)

Barbora Loskotová (Univerzita Tomáše Bati ve

Blanka Farkašová (Masarykova univerzita)

Dagmar Hanzlíková (Univerzita Karlova)

Dana Alamer (Fyzikální ústav AV ČR, v. v. i.)

Daniela Kristeková (Mendelova univerzita v

David Broul (Univerzita Palackého v Olomouci)

Eliška Pospíchalová (Knihovna AV ČR, v. v. i.)

Eliška Vavrečková (Ostravská univerzita)

František Vorel (Biologické centrum AV ČR, v. v. i.)

Georgia Koutentaki (České vysoké učení

Graciela Eguia Marquez (Ústav fyzikální chemie

Hana Kubelková (Masarykova univerzita)

Illyria Brejchová (Masarykova univerzita)

Ing. Iveta Drobníková (Ústav analytické chemie

Iva Hasíková (Mendelova univerzita v Brně)

Jakub Navařík (Univerzita Palackého v

Jakub Zach (Technická univerzita v Liberci)

Jan Aubrecht (Ústav fotoniky a elektroniky AV

Jan Skúpa (Vysoké učení technické v Brně)

Jan Vališ (Vysoká škola chemicko-technologická

DATASTEWARDS COUNT: 95
INSTITUTIONS INVOLVED: 41

Skillset - All -

Can help with - All -

Field of Science - All -



Who are the data stewards of Czech Republic?

MAP OF DATA STEWARDS IN CZECH REPUBLIC

Q KOUTENTAKI X ↕ A-Z

Georgia Koutentaki (České vysoké učení) ⓘ

Georgia Koutentaki X

- Data Steward for STEM
- České vysoké učení technické v Praze
- institution
- Technical Sciences, Engineering

Share the card of Georgia Koutentaki

<https://fit.cvut.cz/cs>

✉ georgia.koutentaki@cvut.cz

Can help with:

- Data Stewardship Wizard (FAIR Wizard)
- Excel
- GitHub / GitLab
- re3data
- Repozitáře (Zenodo, DSpace, oborové repozitáře)
- FAIRsharing

Agenda:

- Data management plan administration
- Data and metadata quality assurance
- Compliance support (policies and domain standards)
- Training and education of researchers and support staff
- Long-term data archiving support (e.g. repository selection)

Help for institutes – EOSC CZ

EOSC CZ trainings - <https://www.eosc.cz/en/training>

Training Aligned with the Data Lifecycle



Overview and Introductory
Courses



Data Management Planning



Data Collecting



Data Processing



Data Analysis



Data Storage



Data Sharing



Data Reuse



Other

Help for institutes – National Library of Technology



[Skill set map](#)
for mentors of
early career researchers



[Calendar](#)



Help for institutes – Library of the CAS

Open Science Department:

- OS support for the CAS institutes and others
- Consultations services, trainings
- OS Information website
- News and events from OS world

Webinar: Digital Data Collections (T) and the National Metadata Directory (cze)

March 11, 2026

When: March 16, 2026, 10:00–11:00 CET Where: online The EOSC CZ Training Centre will host a webinar on 16 March 2026, focused on digital data collections (research output type T) and their registration in the National Metadata Directory (NMA). [...]

[Read more](#)

Webinar: Legal Aspects of Research Data (cze)

March 9, 2026

When: March 24, 2026, 10:00–11:30 Where: online More information: can be found here. Registration: open until 20 March here. Open Science Working Group of the Association of Libraries of the Czech Universities (AKVŠ) is organizing a webinar about legal aspects of research [...]

[Read more](#)

PIDfest 2026 – Call for Proposals

March 9, 2026

When: October 27–29, 2026 Where: Leiden, the Netherlands Proposal submission deadline: March 22, 2026 PIDfest 2026, the second international conference on persistent identifiers (PIDs), will take place in Leiden, the Netherlands, from October 27 to 29, 2026. The three-day event [...]

[Read more](#)

Prague EOSC CZ Networking

March 9, 2026

When: March 30, 2026, 9:30–18:00 Where: Czech University of Life Sciences Prague How can researchers work effectively with their data, and which tools can make everyday research easier? These questions, among others, will be addressed at the Prague EOSC CZ [...]

[Read more](#)

Open science – Library of the Czech Academy of Sciences (LCAS)

Our Services



Consultations

We offer consultations on Open Science, creating OS policies, DMPs, and repositories



Training

We offer training on creating DMPs using the FAIR Wizard



AV CR FAIR Wizard

We provide access to this tool for creating DMPs

Useful Information



Data Management



Funders requirements



FAQ

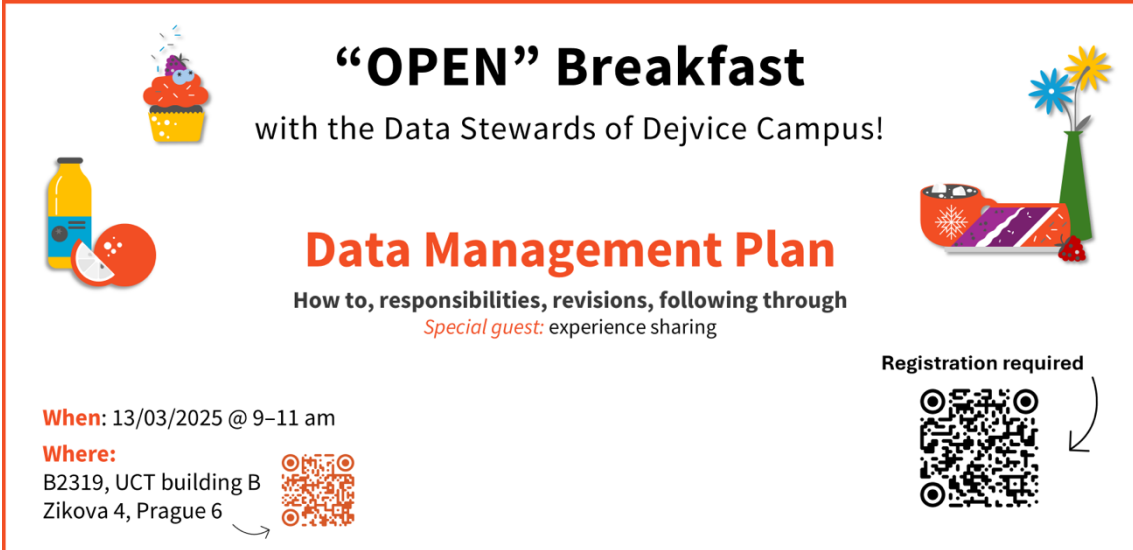
Help for institutes – Funders

- Ministry of Education, Youth and Sports of the Czech Republic (MŠMT). (2025). Open science in the Czech Republic. Retrieved November 4, 2025, from <https://msmt.gov.cz/vyzkum-a-vyvoj-2/otevrena-veda-v-cr>
- Czech Science Foundation (GA ČR). (2025). GA ČR and open science. Retrieved November 4, 2025, from <https://gacr.cz/en/gacr-and-open-science/>
- Technology Agency of the Czech Republic (TA ČR). (2025). Data management plan TA ČR. Retrieved November 4, 2025, from <https://tacr.gov.cz/dokumenty/data-management-plan-ta-cr-1>
- European Research Executive Agency (REA). (2025). Open science. European Commission. Retrieved November 4, 2025, from https://rea.ec.europa.eu/open-science_en

Help for institutes

Breakfasts with data stewards

- Open and free to everyone
- Informal setting
- Coffee served
- Knowledge served
- Collaborative




“OPEN” Breakfast
with the Data Stewards of Dejvice Campus!

Data Management Plan
How to, responsibilities, revisions, following through
Special guest: experience sharing

When: 13/03/2025 @ 9–11 am
Where: B2319, UCT building B
Zikova 4, Prague 6

Registration required



Source: “OPEN” Breakfast with the Data Stewards of Dejvice Campus!,
G. Koutentaki, J. Vališ, M. Schätz, H. Thürlová , 2025, [CC-BY 4.0 International](https://creativecommons.org/licenses/by/4.0/)

Thank you for your attention

georgia.koutentaki@cvut.cz



Co-funded by
the European Union



Q&A



Literature

Ministry of Education, Youth and Sports of the Czech Republic (MŠMT). (2025). Open science in the Czech Republic. Retrieved November 4, 2025, from <https://msmt.gov.cz/vyzkum-a-vyvoj-2/otevrena-veda-v-cr>

Czech Science Foundation (GA ČR). (2025). GA ČR and open science. Retrieved November 4, 2025, from <https://gacr.cz/en/gacrand-open-science/>

Technology Agency of the Czech Republic (TA ČR). (2025). Data management plan TA ČR. Retrieved November 4, 2025, from <https://tacv.gov.cz/dokumenty/data-management-plan-ta-cr-1>

European Research Executive Agency (REA). (2025). Open science. European Commission. Retrieved November 4, 2025, from https://rea.ec.europa.eu/open-science_en

European Open Science Cloud (EOSC-CZ). (2025). Training. Retrieved November 4, 2025, from <https://www.eosc.cz/en/training>

European Open Science Cloud (EOSC-CZ). (2025). Data stewards in Czechia. Retrieved November 4, 2025, from <https://www.eosc.cz/en/communities/data-stewards/data-stewards-in-czechia>

National Library of Technology (NTK). (2025). Homepage. Retrieved November 4, 2025, from <https://www.techlib.cz/>

Data Stewardship Wizard. (2025). Data Stewardship Wizard. Retrieved November 4, 2025.

FAIR Wizard. (2025). FAIR Wizard. Retrieved November 4, 2025.

ARGOS. (2025). ARGOS: OpenAIRE data management planning tool. Retrieved November 4, 2025.

Digital Curation Centre (DCC). (2025). DMPOnline. Retrieved November 4, 2025.

ETH Zürich. (2025). File formats for archiving. Unlimited ETH Zürich. Retrieved November 4, 2025, from <https://unlimited.ethz.ch/spaces/DD/pages/194127879/File+formats+for+archiving>

Digital Preservation Coalition (DPC). (2025). File formats and standards. In Digital preservation handbook. Retrieved November 4, 2025, from <https://www.dpconline.org/handbook/technical-solutions-and-tools/file-formats-and-standards>

MIT Libraries. (2025). File formats for data storage and sharing. Retrieved November 4, 2025, from <https://libraries.mit.edu/data-management/store/formats/>

Literature

Cornell University Research Data Management Service Group. (2025). File formats for preservation. Retrieved November 4, 2025, from <https://data.research.cornell.edu/data-management/archiving-and-preservation/file-formats-for-preservation/>

Koutentaki, G., Vališ, J., Schätz, M., & Thürlová, H. (2025). "OPEN" Breakfast with the Data Stewards of Dejvice Campus! [Event]. Licensed under CC BY 4.0 International.

National Institute of Allergy and Infectious Diseases (NIAID). (2017). Pill Box (32636946385).jpg [Photograph]. Wikimedia Commons. Licensed under CC BY 2.0. [https://commons.wikimedia.org/wiki/File:Pill_Box_\(32636946385\).jpg](https://commons.wikimedia.org/wiki/File:Pill_Box_(32636946385).jpg)

Evelyn, J. (1699). Recipes in Acetaria: A Discourse of Sallets [Manuscript image]. Wikimedia Commons. Public Domain. https://commons.wikimedia.org/wiki/File:Recipes_in_Acetaria_-_A_Discourse_of_Sallets.jpg

Doyle, L. (2021). Fluorescent polarization data for Figure 7 [Data set]. Harvard Dataverse. <https://doi.org/10.7910/DVN/ROSFLC>

Library of the CAS <https://openscience.lib.cas.cz/en/>

DSW NRP - dmp.eosc.cz

Shaddim; original CC license symbols by Creative Commons - <https://creativecommons.org/about/downloads/https://creativecommons.org/policies/Original>

Opensource.org <https://opensource.org/>

Cornell University Library : <https://guides.library.cornell.edu/ecommons/formats>. "Cornell University Library's Copyright Services LibGuide is licensed under a Creative Commons Attribution 4.0 International License.