

ELIXIR CZ Annual Conference 2024

A Few Glimpses into Data Management at Masaryk University

Michal Růžička, Tomáš Svoboda, Adrián Rošinec Masaryk University | CERIT-SC openscience@ics.muni.cz

2024-10-10

Annotation

In this talk, we would like to provide a few glimpses into some aspects of data management at Masaryk University, which is representative of a mid-size university with faculties covering topics from arts through computer science to medicine.

We will quickly look at preparations for the university's data strategy, centralized support of projects data management consultations and Data Management Plans preparations, and particular use cases of data management technical support on selected university centers.



About Masaryk University (MU)

Masaryk University, Brno, Czech Republic



Masaryk University, Brno, Czech Republic

- Since January 1919.
- The second largest university in the Czech Republic now, many building across
 Brno, Telč, Antarctic, …
- 10 faculties and other specialized centers (Institute of Computer Science, Central European Institute of Technology [CEITEC], Mendel's Museum, ...).
- 7,000 employees (out of it 2,900 pedagogical staff and 2,800 specialists and technical staff).
- 30,500 students.
- 400 study programmes, 7,000 graduates per year.

Source: About us | Masaryk University, https://www.muni.cz/en/about-us



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.?



Open Science Support at MU

Research Data Life-Cycle Support at MU

- At Masaryk University, we are trying
 to support all aspects of data management within
 research data life-cycle.
- MU Open Science web is a gateway
- to the centralized support of
- Open Science at our university
 - <u>https://openscience.muni.cz/en</u>





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.



Open Science Support at MU

– Main topics are:



- Open Access and electronic information resources.
- Preparation and Open Science-related support of research projects.

Data Management Plans and Planning.

Research/FAIR data management.

– Activities in national and international projects.

 MU is important partner in the implementation of EOSC in the Czech Republic.

https://eosc.cz/en

National Repository Platform for Research Data (NRP) is a key project to the effective management, long-term preservation and sharing of scientific data in the Czech Republic, thus promoting open access and collaboration in the scientific community.

Open Science Team at MU

– Open Science Core Team.

- Four people at Institute of Computer Science at MU.

Open Science manager

Open/FAIR Data specialist

Open Access specialist

Open Science coordinator

- We are listening to you at <u>openscience@ics.muni.cz</u>.
- Provides centralized support to Open Science Methodologists and researchers, main contact point on Open Science outside MU.

- Open Science Methodologists.

- Dedicated methodologists at 13 faculties and specialized centers.
- Deep knowledge of the faculty/center.

Fist line of support able to quickly help or escalate the topic to the core team. Core team provides regular training sessions to them.

<u>https://openscience.muni.cz/en/contacts/</u>



Open Science Strategy at MU and Preparations of Data Strategy

MU Open Science Strategy 2022–2028

- Strategic framework for effective implementation of Open Science at MU
- Action plans containing the steps necessary to apply the whole strategy effectively.
 - Two periods: 2022–2024 and 2025–2028.

 The main goal: The infrastructure of the advanced research services supporting Open Science.



MU Open Science Strategy 2022–2028

Open Access policy	Specialized support for Open Access development	Open Access funding
 • OA1: Storage of MU scientific publications • OA2: Increasing accessibility of MU scientific publications 	 OA3: MU Repository OA4: Open publishing 	 OA5: MU OA fund OA6: EIR adaptation to a transformation model
FAIR Data policy	Specialized support for FAIR Data development	FAIR Data infrastructure
 FD1: Strategy for MU research data management and access FD2: Research data records 	 FD3: MU data repository FD4: Support for research data management and access 	• FD5: Research data infrastructure

Support for scientists	Internal processes	Standards and principles	Cooperation
 OS1: DMP+ OS2: Human resources OS3: Education and information OS4: Open Science services 	 OS5: Policies at individual ECUs OS6: Open Science in the science evaluation context OS7: Open Science and MU grant schemes 	 • OS8: DOI: • OS9: ORCID • OS10: OpenAIRE • OS11: AOAP, ACAN 	 OS12: National initiatives OS13: International initiatives

MUNI



Data Management Planning

DMPlanning vs. DMPlan

1. Data Management Planning

"The **process** of planning, describing, and communicating the life cycle of data and the activities associated with its management during research."

2. Data Management Plan (DMP)

"A **document** that describes these activities (documents are often required by grant or grant providers)."

Source: Petra Dědičová, Data management a jak psát data management plan, <u>https://www.slideshare.net/butlibrary/data-management-a-jak-pst-data-management-plan-41441697</u>

Tools for DMP Preparation

- Shared document (Google Docs, Office 365 Word, Overleaf, ...).
 - In particular, it supports the preparation of a DMP document.
- Specialized software, typically a web application that enables collaboration.
 - Depending on the level of sophistication, it can support not only the creation of a DMP document, but also the research data management process itself.

Research Funding Agencies Requirements

- Data Management Plan is a common requirement of projects now.
 - For example, Horizon Europe wants an initial DMP,
 - the DMP in the middle of the project,
 - the final DMP.
 - DMP is a living document should be kept up-to-date during the project.
- Obligation to prepare DMP is often the trigger for the researchers to contact Open Science support team.

- In the next steps, we are often also helping with the data management itself.
- To satisfy the requirements we are utilizing specialized departments at faculties and institutes, like

CEITEC MU Core Facilities,

- Centers of Information Technologies at faculties,
- centralized storage and computation services provided by ICS MU etc.
- We are trying to support the data managing process.

Data Stewardship Wizard (DSW, <u>https://ds-wizard.org/</u>) is a nice tool.

DSW MUNI: MU institutional instance of DSW



DSW MUNI: API for Integrations

~

 \sim

Try it out

MUNI

					Poi /document-templates/{documentlemplateid}
		Uloz	it Kopirovat Sba	liit vse Rozbalit vse 🗑 Filtr JSON	GET /document-templates/{documentTemplateId}/bundle
		bu	iltAt:	"2024-03-05T08:28:50Z"	POST /document-templates/{documentTemplateId}/pull
		- co	mponents:		Document
			0:		GET /documents
		1	builtAt:	"2024-03-05T07:57:32Z"	POST /documents
) aplikaci	i Kopí	r	createdAt:	"2023-04-19T14:38:40.289504Z"	DELETE /documents/{docUuid}
			name.	"Mailer"	GET /documents/{docUuid}/available-submission-services
			undatedAt:	"2024 02 00T14+20+EE 000E277"	GET /documents/{docUuid}/download
Klient			upuaceuxc.	2024-03-08114.39.35.8095572	Parameters
Verze	v4.4.0~f9d9c7a		version:	V4.4.0~aa08/a1	Name Description
Sestaveno	5 3 2024 10:28	· · ·	1:		string Authorization (header)
bestarento	5. 5. 202 1, 10.20		builtAt:	"2024-03-05T07:57:42Z"	Host string (herefor)
			createdAt:	"2023-04-19T14:38:40.346442Z"	docUuid * neamd ststeer(fund) docUuid
Server			name:	"Document Worker"	(path)
Verze	v4.4.0~0f8e9e5e		updatedAt:	"2024-03-08T14:39:56.319526Z"	
c .	5 2 2024 40 20		version:	"v4.4.0~aa087a1"	Responses Respon
Sestaveno	5. 3. 2024, 10:28	na	me:	"Engine Wizard"	Code Description
JRL API	https://dsw.muni.cz:3443/wizard-api	Ve	ersion:	"v4.4.0~0f8e9e5e"	Example Value Model
Dokumentace	https://dsw.muni.cz:3443/wizard-api/swagger-ui	,			<pre>contestiype ' rear(plain y ''uni': "http://example.com/temporary-file-1" }</pre>
API					Headers: Name Description
					x-trace-unid 400 Invalid doctorid or Host or Authorization
Document Wor	rker				
Verze	v4.4.0~aa087a1				Documnet Submission
					GET /documents/{docUuid}/submissions
Sestaveno	5. 3. 2024, 9:57				POST /documents/{docUuid}/submissions
					Domain
Mailer					Ciel / domanns
Verze	v4.4.0~aa087a1				
Sestaveno	5 3 2024 9.57				
Sestaveno	5. 5. 202 (, 5.57				
	Ok				

DSW MUNI: Integrations with INET

- Accepted projects editor in INET:

https://inet.muni.cz/app/proj/projektlist

 \rightarrow select project \rightarrow tab *Results and DMP*

Basic data Description Beneficiaries/Partners People B	udget @Econom	y 🖹 Documents	Approval	⊟Events/tasks	A Results and D	MP Audits	Authorization
Data Management Plan (DMP) ③ Create the new DMP ③ Link the existing DMP							
Name Description DMP UUID							
3b8ec4f0-311e-4b53-8	3b12-b326f64bd093				×		
Create the DMP Name Creating a Robust Accessible Federated Technology for Open Access Description Save	Link the existing DMP DMP (selection or UUID) Description Close	Select Item (total 12) Select Item (total 12) Creating a Robust Accessible Fed Cyber-security Excellence Hub in DMP Example: Alliance for Life Sc DMP Example: Ashy layers: Monji DMP Example: Ashy layers: Monji DMP Example: Integration of RNA DMP Example: Investigating the tr DMP Example: Neolithic Settleme DMP Example: V\$CHT DMP Exam National Institute of Virology and E National Institute of Virology and E National Institute of Virology and E	erated Technology for Estonia and South Mo iences: From Strategi ikli Depe (UUID: a31c nunity Diversity for Pla Biology for Next-Gen anscriptional regulatio nts Analysis (UUID: 2a nple – DSW-Horizon F lacteriology (NIVB) - F UUID: e2ec8e7f-23db	Open Access (UUID: 3b8ec4f0-31 avia (CHESS) (UUID: 6d8b30cb-fn bd52-76f6-45dd-9e9f-8c3cb-f107 tStress Resilience (GENCOVER ration Scientists (INTEG-RNA) (U n of auxin biosynthesis in Arabidop 12429e-0ad9-470d-atbe-49365c90 urope (UUID: a3cdc700-78fd-4de/ avid \$majs (UUID: b7268f42-e590 man Pantúček (UUID: 740af880- 4631-a4cc-55a017d411a8)	1e-4b53-8b12-b326l64bd093) ifb-42c0-a054-caf787432eac) n Europe (A4L_ACTIONS) (UUID: ib)) (UUID: 6e7bed40-b7f2-4691-9dai UUID: 13d299cb-482a-4cb8-aaaf-8e sis embryo (ITRABAE) (UUID: ea3 de33) 2-b63b-b6aaaf47b4ca) -4284-83c7-e311ddbfb4aa) 8046-4415-8df8-ce94dc9d1a36)	6e1d7ddb-a462-4c53-a4 3-5ae2407d7(18) (57b7e4bc2) (0803-0c2a-4f31-ac04-7	* Id4-17b75c69908c) ISE



Data Management Technical Support

Selected Services

Storages at MU and recommendations for their use

Centrally managed by Institute of Computer Science

<u>https://it.muni.cz/en/categories/data-storage</u>

– Recommendations their use:

- Data categorization.
- Storage technology categorization.
- Suitability of storage for different types of data.

Storages at MU – data categorization

CATEGORY	DESCRIPTION	EXAMPLE
Public data	 Data that can be made available to anyone without any restrictions, e.g. publicly displayed on the Internet. Their publication does not pose any threat to MUNI or other institutions or persons. 	 Presentations from public lectures; publicly available research reports; open-source software; public research data; Promotion, public information about services.
Internal data	 Data intended only for the internal use of a generally defined group of persons (e.g. project collaborators, institutional staff, etc.). However, they do not require special regulation or protection (by law, contract, etc.). Disclosure outside the group will not cause direct damage (financial, moral, legal, etc.). 	 Internal correspondence; minutes of meetings; internal regulations and regulations; internal work plans, notes, etc.; unfinished/unpublished research reports.

MUNI

Storages at MU – data categorization (2)

CATEGORY	DESCRIPTION	EXAMPLE
Discrete data	 Data intended exclusively for the internal use of a precisely defined group of persons (e.g. an employee and his/her direct superior, an employee of the HR department and a job applicant, a group of IT system administrators with administrator rights to it). By their nature, they require regulation or protection, typically data are protected by law or on the basis of a contract/license (e.g. personal data of individuals, data falling under trade secrets, etc.). Disclosure outside the given group of people is very likely to cause damage (financial, moral, legal, etc.). 	 Economic and personal data of a personal nature; personal data of students/employees/co-workers; identification card numbers, birth numbers, etc.; credit card numbers; valuable research data (e.g. providing a competitive advantage) or data containing otherwise sensitive information; large collections of internal data; access data (e.g. passwords or encryption keys) to minor systems and internal data.

Storages at MU – data categorization (3)

CATEGORY	DESCRIPTION	EXAMPLE
	 Data intended strictly for the internal use of a precisely defined group of people (e.g. a healthcare professional and his patient, project investigators with security clearance of a certain level, etc.). By their nature, they require special regulation or special protection, typically the data are strictly. 	 Health data, sensitive personal data; very valuable research data (e.g. providing a unique and difficult to repeat competitive)
Sensitive data	protected by law or on the basis of a contract/license (e.g. very valuable data falling under trade secrets, sensitive personal data, etc.).	 advantage) or research data containing highly confidential data; extensive collections of discrete data;
	Disclosure outside the given group of entitled persons is very likely to cause damage (financial, moral, legal, etc.) of a large extent with serious/irreversible consequences.	 access data (e.g. passwords or encryption keys) to important systems and data categorized as discrete or sensitive.
	 In practice, few data will fall into this category, most will fall into the discrete data category at most. 	

Storages at MU – categorization by technology

STORAGE TYPE	DESCRIPTION
Portable media	E.g. flash drives, memory cards, external HDD/SSD, CD, DVD, That is, external storage media that are not fixed to any device and are used by users to transfer information between devices or to store data offline.
Local	In desktops/notebooks Data storage devices permanently built into desktop computers/laptops (typically internal HDD/SSD, etc.) in employee offices, study rooms, etc.
storage	In portable devices (smartphones, tablets) Data storage devices permanently built into mobile devices, i.e. mobile phones, tablets, etc. (typically internal non-removable memory, memory card installed in the device, etc.) for use by employees/students.
Network and cloud ICS storages	Data storage operated by ICS and made available to end users via a data network – so- called standard and medium network storage. This category also includes CERIT-SC data repositories for large-volume research data.
IS MUNI storage	Document Server, File Depository, and similar storage capacities integrated in the IS MUNI system.
CESNET storages	Data storage is operated by the CESNET Data Storage Department. This category also includes services that use these services for physical data storage, e.g. CESNET OwnCloud, CESNET FileSender, Object storage (S3), etc.

MUNI

Storages at MU – categorization of repositories (2)

STORAGE TYPE	DESCRIPTION
	With the contract with MUNI
	MUNI Microsoft O365 Cloud data storage provided as part of the Microsoft Office 365 service for Masaryk University. In particular, these are personal storage OneDrive and document libraries of SharePoint and O365 Groups. However, this also includes other data stored in the MUNI O365 cloud, such as e-mail in MUNI O365 Outlook, files shared in the MUNI Yammer social network, etc.
External storages Data repositories operated by external entities, i.e. outside MUNI and CESNET.	MUNI Google Workspace for Education Cloud data storage provided within the Google Workspace for Education service for Masaryk University. In particular, these are the data capacities of MUNI Google Drive, but it also includes other data stored in the MUNI Workspace for Education cloud, e.g. e-mail in MUNI Google Mail, notes in MUNI Google Keep, calendar data in MUNI Google Calendar, etc.
	Without a contract with MUNI
	Public Google/Microsoft/Dropbox/ storages This category includes in particular public cloud services (typically set up free of charge by a private end user only against electronic registration via the web) such as Google Drive, Microsoft OneDrive, Dropbox, Amazon storage, GitHub repositories, etc. The fundamental difference and "hallmark" of this category of cloud storage compared to the cloud services listed above is that MUNI has no (legal) relationship with the operators of these external services and is therefore unable to guarantee any guarantees regarding the security/confidentiality of stored data or the policy of handling them.

MUNI

Storage at MU – Suitability of storage for different types of data

STORAGE 1	ГҮРЕ	USAGE			
		GREEN: PUBLIC DATA	BLUE: INTERNAL DATA	ORANGE: DISCRETE DATA	RED: SENSITIVE DATA
PORTABLE EXTERNAL	MEDIA (FLASH DRIVES, HDD, CD, DVD,)	Appropriate	Possible Encryption recommended	Inappropriate Possible when using encryption	Inappropriate
LOCAL STO	DRAGE				
	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 NETWO STORAGE (SO-CALLE MEDIUM NE IT CATALOO STORAGE)	RK AND CLOUD D STANDARD AND TWORK STORAGE, SEE GUE, CERIT-SC	Appropriate	Appropriate	Appropriate	Appropriate It is recommended to perform an individual analysis, use encryption and apply other security measures that result from the analysis

Storage at MU – Suitability of storage for different types of data (2)

STORA	GE TYPE	USAGE			
		GREEN: PUBLIC DATA	BLUE: INTERNAL DATA	ORANGE: DISCRETE DATA	RED: SENSITIVE DATA
IS MUNI (E.G. DC DEPOSI	REPOSITORY DCUMENT SERVER, FILE TORY, 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 (E.G. CE OWNCL DATA S	T STORAGE ESNET ARCHIVE STORAGE, OUD, FILESENDER,, SEE CESNET TORAGE 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
EXTERN	IAL 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				
31	PUBLIC GOOGLE, MICROSOFT, DROPBOX, STORAGES	Appropriate	Inappropriate	Inappropriate	Inappropriate

Persistent Identifiers (PIDs)

- Persistent Identifiers (PIDs)
 important part of FAIR data
 principles.
 - MU member of DataCite, able to assing DOIs also for data.
 - DataCite DOI on MU
 - (<u>https://commons.datacite.org/ror.org/02j46</u> <u>qs45</u>), as of 2024-04:

967 Works	627 Citations ⑦	19,263 Views 🕜	2,277 Downloads	:?	
Founded 1919					
Links			Other Identifiers		
Homepage			GRID grid.10267.32		
Nikipedia			Crossref Funder ID 1	10.13039/50110001065	3
Twitter			ISNI 000000121940	0956	
Geolocation					
49° 11' 55.0" N, 16° 3 Czechia Educatio	on 🗣 DataCite Co	nsortium Organizatio	n		
49° 11' 55.0" N, 16° 3 Czechia Educatio Z https://ror.org/02 067 Works	86' 19.0" W on 🗣 DataCite Co 1;j46qs45	nsortium Organizatic	n		
49° 11' 55.0° N, 16° 3 Czechia) Educatii ≩ https://ror.org/02 267 Works	66 19.0" W On DataCite Col 1/46qs45	nsortium Organizatio		Liconcor	
49° 11' 55.0° N, 16° 3 Czechia) Educati ≩ https://ror.org/02 267 Works Publicatio	66 19.0" W On C DataCite Co 1/46qs45 On Year	nsortium Organizatio	n (7)	Licenses	0
49° 11' 55.0° N, 16° 3 Czechia) Educatio 2° https://ror.org/02 267 Works Publicatio	196 19.0" W on PataCite Co 1946qs45 on Year	nsortium Organizatio Work Types Dataset	n @ 39%	Licenses CC-BY-4.0	? 52%
49° 11' 55.0° N, 16° 3 Czechia) (Educatii 2ª https://ror.org/02 167 Works Publicatio 384 384 384 384 384 384	96° 19.0" W DataCite Co Pl46qs45 Don Year	nsortium Organizatio Work Types Dataset	m (9) 39%	Licenses CC-BY-4.0	⑦ 52%
99° 11' 55.0° N, 16° 3 Czechia) Educatii 22 https://ror.org/02 2667 Works Publicatio	is6' 19.0" W on (Work Types Dataset	m 39%	Licenses CC-BY-4.0	@ 52%
99° 11' 55.0° N, 16° 3 Czechia) Educatii Z'https://ror.org/02 26 TWorks Publicatio 286 296 296 296 296 296	on) (⊉ DataCite Co. }46qs45	Work Types Dataset	m 39% 6% 8% 1%%	Licenses CC-BY-4.0	(?) 52%
99° 11' 55.0° N, 16° 3 Czechia) Educati 2° https://ror.org/02 267 Works Publicatio 306 306 306 306 306 306 306 306	in Year	Work Types Dataset	© 39% 50% 40% 19%	Licenses CC-BY-4.0	(7) 52% 10% 80% 10% 10% 60% 10%

- I need a DOI for data what should
 I do?
 - Write to <u>openscience@ics.muni.cz</u>.
 - We will arrange a meeting and show you how DOIs are assigned.

By manually filling out forms. Machine via API.

- We will set up your own DOI prefix in the testing and production infrastructure.
- We will test the filling in of records in the test infrastructure.

- We transfer it to production.

Questions?



Source: <u>Communicate_communication_conference_2028004</u> by <u>OpenClipart-Vectors</u> from <u>Pixabay</u>

33 Data Management at MU | ELIXIR CZ 2024 | Růžička at al. | 2024-10-10

MUNI