



Univerzita
Karlova

Jak na data management plan (nejen) pro GAČR



Dagmar Hanzlíková & Lucie Hošková
openscience@cuni.cz
Centrum pro podporu open science
Univerzita Karlova





Výzkumná data





Výzkumná data

= údaje, které byly vytvořeny či získány za účelem zjištění či reprodukování vašich výsledků výzkumu

Výzkumná data mohou mít různou podobu (digitální i nedigitální)

- Tabulky, dokumenty
- Audio a video nahrávky, obrázky, fotografie
- Dotazníky, odpovědi na otázky, přepisy rozhovorů
- Laboratorní deníky, terénní poznámky
- Software, skript
- Vzorke, exempláře

Správa výzkumných dat

- Je třeba v průběhu celého výzkumného procesu
- Usnadňuje psaní článků a jejich revize
- Snížení rizika nutnosti stažení článku z důvodu pomíchání dat nebo jejich špatného označení
- V případě pochybností o validitě výsledků se můžete o svá data opřít





Replikační krize

= mnoho studií není možné zopakovat a snahy dospět ke stejným výsledkům končí neúspěšně

Psychology's Replication Crisis Is Running Out of Excuses

Another big project has found that only half of studies can be repeated. And this time, the usual explanations fall flat.

Half of top cancer studies fail high-profile reproducibility effort

Barriers to reproducing preclinical results included unhelpful author communication, but critics argue that one-time replication attempts don't tell the whole story.

More than 10,000 research papers were retracted in 2023 – a new record

The number of articles being retracted rose sharply this year, but this is only the tip of the iceberg.

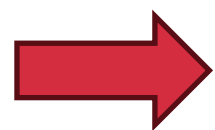
A wave of retractions is shaking physics

Grappling with problematic papers and poorly documented data, researchers and journal editors gathered in Pittsburgh to hash out the best way forward.

Nový rektor Mendelovy univerzity má problém. Jeho laboratoř falšovala data



Replikační krize: Co s tím?



Snaha o robustní a replikovatelný výzkum

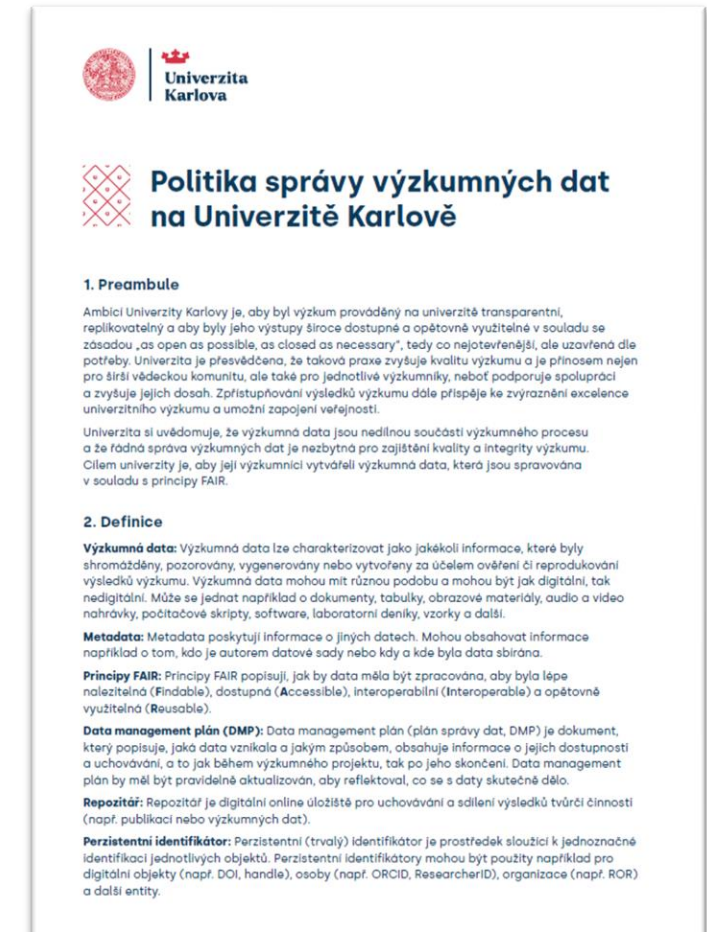
- Časopisy: Povinnost zveřejňovat podkladová data
- Poskytovatelé financí: Povinnost správy (a zveřejnění) dat
 - Popis v projektové žádosti
 - Data management plan
- Instituce: Datové politiky



Datová politika UK

Cíl: Specifikovat **základní principy** správy výzkumných dat a vymezit **odpovědnosti** univerzity a jejích výzkumníků.

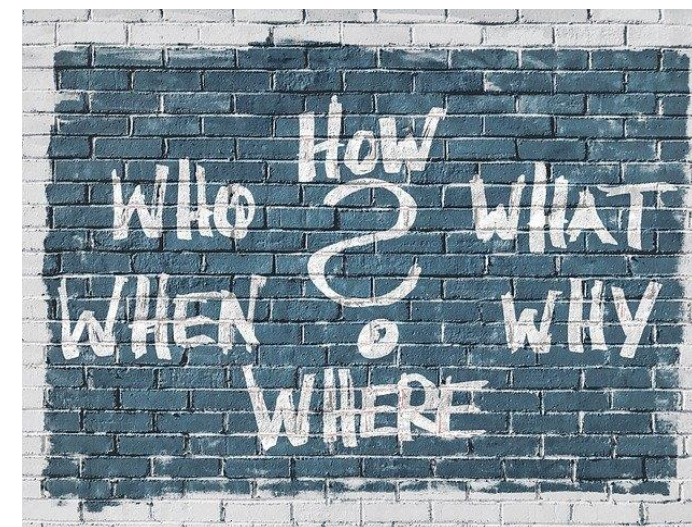
1. Preambule (Preamble)
2. Definice (Definitions)
3. Cíl (Aim)
4. Oblast působnosti (Scope)
5. Základní principy (Basic principles)
6. Odpovědnosti (Responsibilities)
7. Podpora (Available Support)
8. Revize dokumentu (Review period)
9. Související dokumenty (Related documents)



Data management plan (DMP)

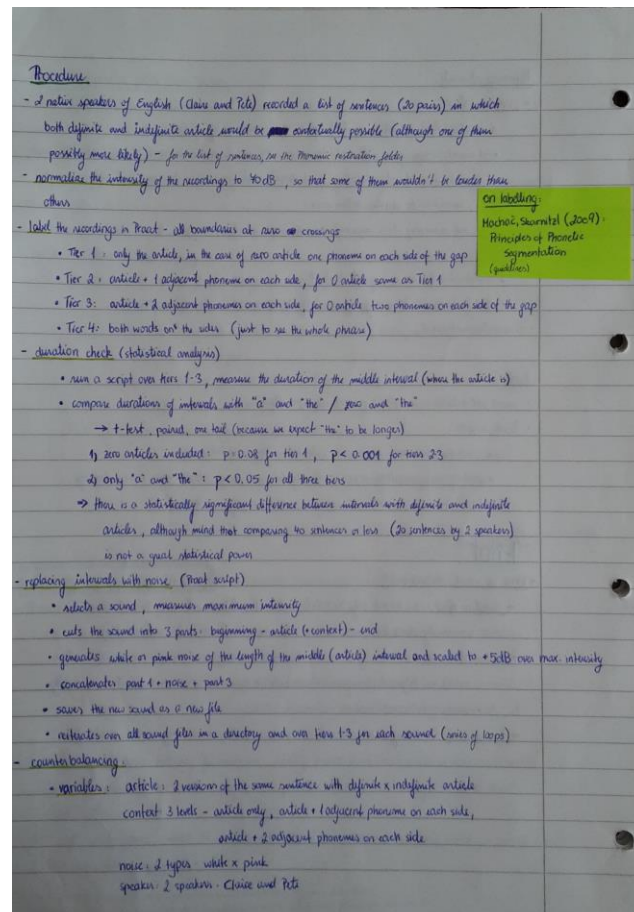
= dokument, který specifikuje, *jaká data a jakým způsobem* budou během výzkumu *vytvářena*, a obsahuje informace o jejich *dostupnosti a využití*

- **Živý dokument** aktualizovaný dle potřeby
- Pomáhá předvídat **potenciální problémy**
- Slouží k **průběžné reflexi** práce s daty
- Zajištění, že data jsou **přesná, úplná a spolehlivá**
- **Kontinuita a konzistence** dlouhodobých projektů
- Pomoc při **sdílení dat**



Data management plan (DMP)

- Různá forma – šablona poskytovatele, oborové šablony (CESSDA)



- administrative data
- data collection
- documentation and metadata
- ethics and legal compliance
- storage and backup
- selection and preservation
- data sharing
- responsibilities and resources



- summary
- making data findable
- making data openly accessible
- making data interoperable
- increase data reuse
- allocation of resources and data security



DMP v GAČR

- DMP musí obsahovat minimálně informace o tom:
 - i. jaká výzkumná data bude příjemce v rámci projektu **vytvářet**, **zpracovávat** nebo **sbírat**;
 - ii. jaké **metody a zásady** jejich správy použije;
 - iii. zda a jakým způsobem budou data **sdílet**, **publikovat** nebo **uveřejňovat**, případně **odůvodnění**, **proč nemohou** být výzkumná data šířena a zveřejňována (příjemce nezveřejní informace o výzkumných datech, jejichž zveřejněním by došlo k nepřiměřenému zásahu do práva na ochranu duševního vlastnictví, obchodního tajemství, bezpečnosti státu nebo oprávněných obchodních zájmů příjemce nebo třetí osoby);
 - iv. a jakým způsobem budou data **ukládat** během řešení projektu a **uchovávat** po skončení projektu.



DMP v GAČR

Jako **vhodný nástroj** pro vytváření a správu plánu správy dat GA ČR doporučuje využít následujících, ale i jiných vhodných nástrojů, jako jsou například:

- [Šablona DMP](#) vycházející ze vzoru pro program Horizont Evropa ([CZ překlad](#))
- [Data Stewardship Wizard](#)



TRAINING: USING THE DATA STEWARDSHIP WIZARD FOR YOUR SCIENTIFIC DATA MANAGEMENT

The training will be introduced by GACR President, **Prof. Petr Baldrian**, who will explain GACR's approach to scientific data management. Next, Jana Martinková from DSW will take over with practical demonstrations of how to use this tool from the perspective of scientists.

Online

5 Dec 2024
9:30-11:30 a.m.
Language: English

Registration:
www.gacr.cz/en/trainingDSW/





DMP příklady

- Horizon Europe
 - DMP jsou Deliverables – jsou v systému CORDIS [tady](#)
 - Projekty jsou velké, konsorciální → DMP jsou komplikovanější
- Zveřejněné plány v DMPonline
 - Najdete v záložce [Public DMPs](#) nástroje DMPonline
 - Není to spravovaná databáze! (může tam zveřejnit kdokoli cokoli)
- Zenodo
 - Search: [data management plan](#)



Jak na DMP





Popis výzkumných dat

= *Co budete sbírat*

- Budete využívat již **existující data**?
- Jaké **typy** a **formáty** dat budete používat / vytvářet?
- Jaký **objem** dat budete generovat?



FAIR data

- Základní principy, jak by měla data vypadat, aby odpovídala současným požadavkům
 - co je potřeba pro budování služeb, která budou propojovat data z různých zdrojů
 - co je potřeba pro sdílení a znovuvyužití dat
 - co je potřeba pro strojové zpracování a data mining
 - co je potřeba k ověření výsledků výzkumu (reproducibilita)
- nestačí vzít svoje soubory a někam je nahrát, **kontext** je stejně důležitý jako data samotná!

FAIR data

Způsob zpracování výzkumných dat tak, aby byla v souladu se základními principy:

- **F**indable (nalezitelná)
- **A**ccessible (přístupná)
- **I**nteroperable (interoperabilní)
- **R**eusable (opětovně využitelná)

What is FAIR DATA?



Data and supplementary materials have sufficiently rich metadata and a unique and persistent identifier.

FINDABLE



Metadata and data are understandable to humans and machines. Data is deposited in a trusted repository.

ACCESSIBLE



Metadata use a formal, accessible, shared, and broadly applicable language for knowledge representation.

INTEROPERABLE



Data and collections have a clear usage licenses and provide accurate information on provenance.

REUSABLE



FAIR: Findable

= vaše data by měla být nalezitelná, jak lidmi, tak stroji. Pro tento účel jsou klíčová *strojově čitelná metadata* a *perzistentní identifikátory*

- Metadata
 - Špatná metadata = nenalezitelná data
- Perzistentní identifikátory (PID)
 - DOI, ORCID...
- Odkazy na další zdroje
 - Pomocí PID

DOI:

DOI [10.5281/zenodo.3978090](https://doi.org/10.5281/zenodo.3978090)

Keyword(s):

[pine wood nematode](#) [Bursaphelenchus xylophilus](#)
[quarantine pest](#) [survey](#) [Finland](#)

Subject(s):

[Bursaphelenchus](#) [surveillance](#) [data](#)

Related identifiers:

Supplement to
[10.3897/neobiota.58.38313](https://doi.org/10.3897/neobiota.58.38313) (Journal article)
[10.5281/zenodo.3842358](https://doi.org/10.5281/zenodo.3842358) (Software)



Findable v šabloně HE

- Budou data opatřena **perzistentním identifikátorem**?
- Budou poskytnuta **bohatá metadata** umožňující vyhledávání?
 - Budete využívat oborové standardy? Jaké?
 - Pokud neexistují oborové standardy, jaká metadata budete vytvářet?
- Bude možné metadata **automaticky zpracovávat** (harvestovat) a indexovat?



Tyto otázky typicky vyřeší vybraný **repozitář**



FAIR: Accessible

= zajištění *otevřeného přístupu* k datům nebo alespoň k jejich metadatům

- Přístup ideálně prostřednictvím **repozitáře**
 - Možnost stažení
- Omezení přístupu může být v souladu s FAIR
 - Autentizace & autorizace možné
- Volně přístupná alespoň metadata
 - Získatelná pomocí identifikátoru
 - Pokud by data byla odstraněna, metadata by měla zůstat



Accessible v šabloně HE

- Repozitář
 - Uložíte data do důvěryhodného repozitáře?
- Data
 - Budou data **volně dostupná**?
 - Bude uplatněné **embargo**? Proč a na jak dlouho?
 - Pokud bude **omezený přístup** k datům, jak jej budete řešit (e.g., přístupová komise)?
- Metadata
 - Budou metadata volně **přístupná** a **nalezitelná**? Jak dlouho?



FAIR: Interoperable

= možnost *integrovat* data s jinými datasey a zajištění *kompatibility* s nástroji pro jejich zpracování

- Používání **standardizovaných výrazů** pro popis
- Otevřené formáty
- Oborové standardy
 - Výzkumné infrastruktury



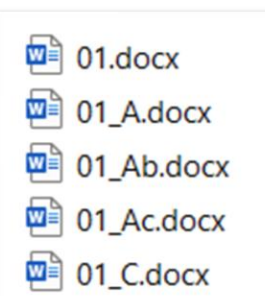
Interoperable v šabloně HE

- Jaké **slovníky, standardy, formáty** a **metodologie** budete používat, aby byla vaše data interoperabilní?
 - Mohou pomoci velké výzkumné infrastruktury
- Pokud budete vytvářet vlastní ontologie, budete je sdílet?

FAIR: Reusable

= optimalizace dat pro co nejvyšší možnost *opětovného využití*

- Důkladný popis dat – dokumentace
 - Popisná metadata
 - Provenance a kontext
- Použití licencí



	File names			
01	original text with no changes, along with title, source and word count			
01_O	original text with no changes, Task 1 format			
01_R	all articles removed from the text, Task 2 format			
01_Ac	articles manipulated - the same as in Task 4; update after 01/02/2018 meeting: no highlighting			
01_C	articles manipulated and highlighted along with their respective nouns, Task 4 format			





Reusable v šabloně HE

- Jakým způsobem poskytnete **dokumentaci**?
 - Např. readme file, číselníky, definice proměnných apod.
- Budou data volně dostupná? Pod jakou **licencí**?
 - Licence specifikuje, za jakých podmínek mohou být data dále využívána
- Bude původ dat (provenance) dostatečně zdokumentován?
 - Jak jste data sbírali a jak jste je upravovali?
- Jakými procesy zajistíte kvalitu dat?



FAIR data

- Velkou část část požadavků na FAIR data lze vyřešit uložením do **vhodného repozitáře** (F, A)
- Naplnění dalších požadavků (I, R) je ale závislé na způsobu **práce s daty v průběhu výzkumu** a těžko se dohání v momentě publikování dat
- Principy FAIR nejsou „binární metrika“ – ne všechno může být 100% FAIR → naplňování principů



Repozitáře

- „Důvěryhodný“ repozitář
 - Otevřený přístup, identifikátory, metadata, licence, udržitelnost...
 - Existují certifikace pro repozitáře (např. CTS), ale pro některé typy repozitářů jsou z principu těžko dosažitelné
- Výběr repozitáře:
 1. Oborový
 2. Institucionální
 3. Obecný ([Zenodo](#), [figshare](#))



Alokace zdrojů

- **Kolik** bude FAIRifikace dat stát peněz?
 - Např. náklady na úložiště, zabezpečení dat, data steward v projektu
- **Z čeho** se to bude platit?
 - Typicky to bývají uznatelné náklady v grantech
- Kdo bude v projektu **odpovědný** za správu dat?
- Jak bude zajištěno **dlouhodobé uchování** dat?
 - Kdo bude rozhodovat o tom, jaká data se budou uchovávat a jak dlouho?
 - Jaké náklady jsou spojené s archivací?



Zabezpečení dat

- Jakým způsobem zabezpečíte svá data?
 - Úložiště, zálohování, přenos (citlivých) dat
- Uložíte data do důvěryhodného repozitáře pro dlouhodobé uchování?



Metodický materiál ÚVT a pověření pro GDPR



Služby CESNET: ownCloud, FileSender, Datová úložiště



Etika

- Je třeba řešit nějaké **etické** či **právní** otázky, které by mohly mít dopad na sdílení dat?
 - Např. sběr osobních údajů, práva třetích stran
 - Je možné odkázat např. na rozhodnutí etické komise
- Budou dotazníky sbírající osobní data obsahovat **informovaný souhlas** s dlouhodobým uchováním a sdílením dat?



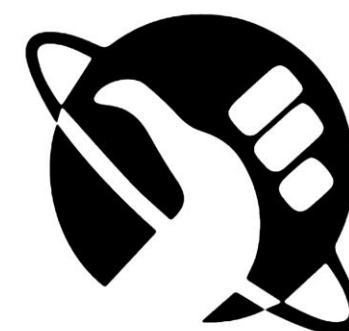
Další otázky

- Budete využívat nějaké národní, institucionální či oborové **postupy pro správu dat**? Jaké?
 - Národní: zákonné povinnosti
 - Institucionální: [Datová politika UK](#)



Take home messages

- DMP je dokumentace vašich dat – bude tak dobrá, jak dobrou si ji uděláte
- DMP je živý dokument, aktualizujte jej
- Buďte konkrétní
- Nemusíte vypisovat slohová cvičení
- Pokud nerozumíte některým termínům, zeptejte se



**DON'T
PANIC**



FAIR Wizard





FAIR Wizard: Co je zač

- Online služba ke správě dat
- Určené pro studenty a zaměstnance UK
- Možnost vytváření DMP

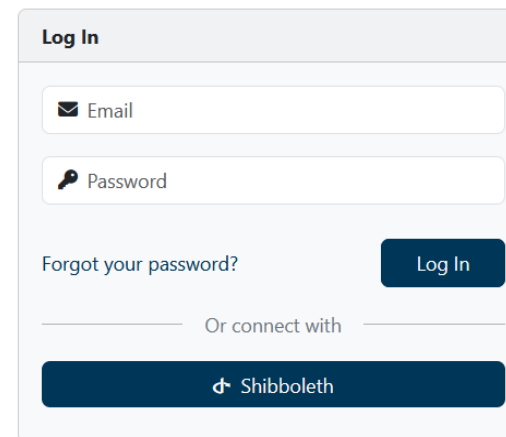


FAIR Wizard CUNI je nástroj usnadňující tvorbu DMP, zprostředkovaný pro zaměstnance a studenty Univerzity Karlovy.

V případě dotazů se obraťte na [podporu](#) či konzultujte [pomocné materiály](#).

FAIR Wizard CUNI is a tool simplifying the DMP creation process made available to the staff and students of Charles University.

If you have any questions, please contact [support](#) or access [the help materials](#).



The screenshot shows a login interface with the following elements:

- Log In** header
- Email input field with an envelope icon
- Password input field with a key icon
- Link: "Forgot your password?"
- Button: "Log In"
- Text: "Or connect with"
- Large button: "Shibboleth" with a key icon

Přihlašte se prosím pomocí systému Shibboleth.

Ochrana osobních údajů

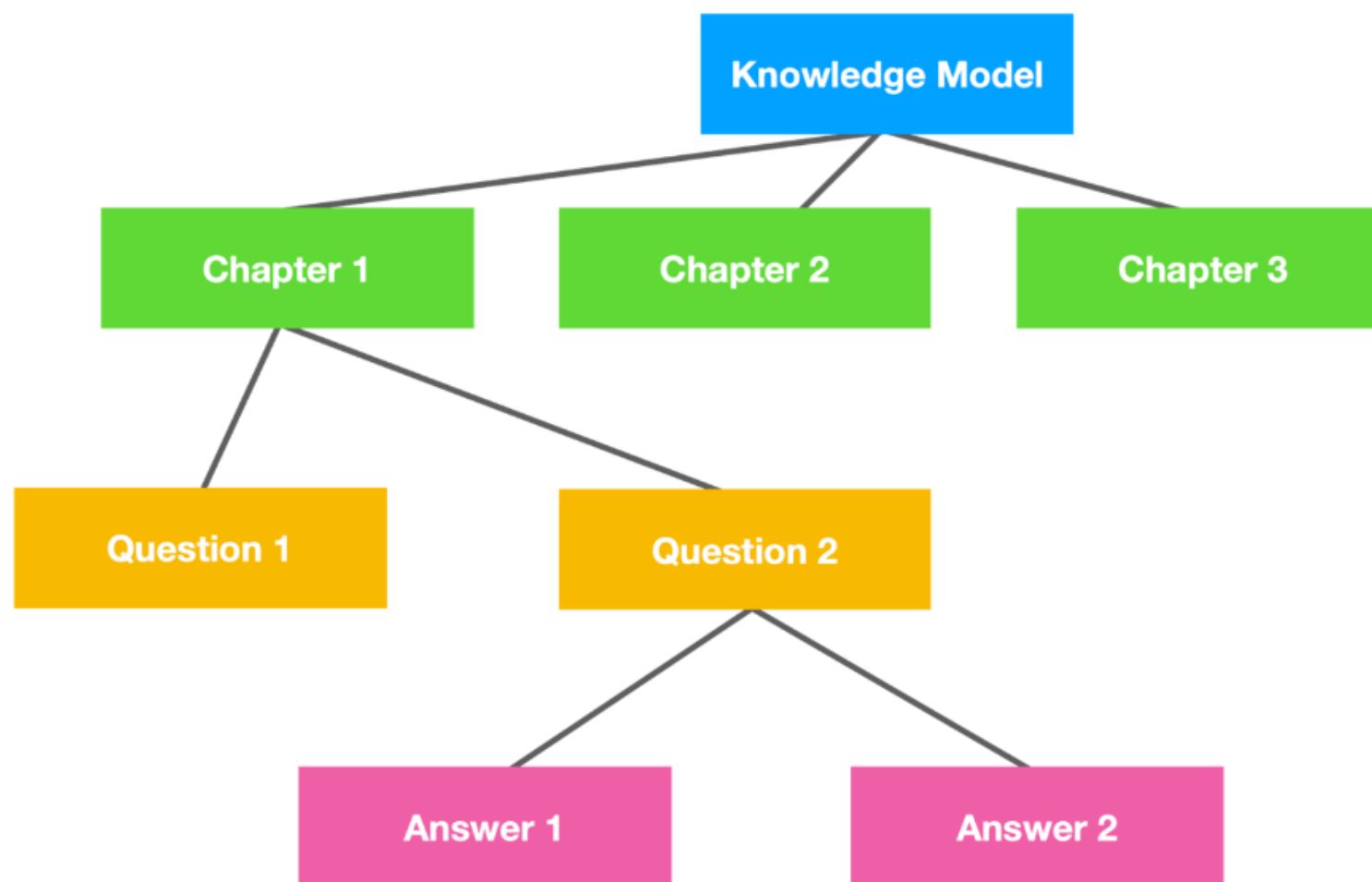
Please log in using the Shibboleth system.

Privacy notice



<https://fair-wizard.cuni.cz>

FAIR Wizard: Princip fungování



- Core DS Knowledge Model
 - Administrative information
 - Contributors
 - Name
 - E-mail address
 - ORCID Identifier
 - Affiliation
 - Role
 - Research Project(s)
 - To execute the DMP, is additional specialist expertise required?
 - Do you require hardware or software in addition to what is usually available?
 - Re-using data
 - Is there any pre-existing data?
 - No
 - Yes
 - Will you be using any pre-existing data (including other people's data)?
 - No
 - Yes
 - ezi
 - Do you need to harmonize different sources of existing data?
 - Will you be using data that needs to be (re-)made computer readable?
 - atq
 - Google dataset search
 - Datacite Search
 - Creating and collecting data
 - Processing data
 - Interpreting data
 - Preserving data



FAIR Wizard: Výsledný DMP



Example Horizon Europe Data Management Plan



Horizon Europe Data Management Plan



03 May 2024

Data Management Plan created in FAIR Wizard «fair-wizard.com»
using Common DSW Knowledge Model v2.6.5 (dsw:root:2.6.5).

1. Data Summary

Re-used datasets

We have found the following non-reference datasets that we have considered for re-use:

- Previous in-house Arsenic and Selenium analysis
Owner of this dataset: Carl Brown, carlbrown@example.com.
The dataset can be used in the provided format without any conversion needed.
We will use its online version without downloading it.
It is a fixed dataset, changes will not influence reproducibility of our results.
Only part of the dataset will be used; any filtering or selection will be well documented.
We will use the dataset as follows: To compare past results.

We will need to harmonize different sources of existing data.

We will be using data that needs to be (re-)made computer readable first. We will provide machine readable, standardised metadata to others. The data itself will be made available in computer readable form to others through a standard repository.

Data formats and types

We will be using the following data formats and types:

- JPG, CSV
It is a standardized format. This is not a suitable format for long-term archiving; however, we plan to convert it to a suitable format before the end of the project. We expect to have 100 GB of data in this format.

2. FAIR Data

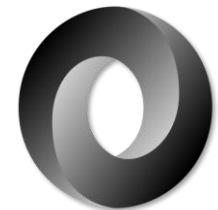
2.1. Making data findable, including provisions for metadata

- Levels of Arsenic in soil (published)
The dataset has the following identifiers:

We will distribute the dataset using:

- Our institutional repository.
A persistent identifier will be assigned by an institutional data steward.

L^AT_EX



{JSON}



FAIR Wizard: Tvorba DMP

- Předdefinované šablony

- Otázka: Budete používat cizí data?
- Odpověď - výběr z možností: ano/ne
- Výsledek: Nebudeme používat cizí data.

- Otázka: V jakém formátu budou vaše data?
- Odpověď - volná odpověď: CSV
- Výsledek: Budeme používat data ve formátu C

II.2 Is there any pre-existing data?

ELSI Horizon 2020 DMP Horizon Europe DMP Science Europe DMP maDMP

Are there any data sets available in the world that are relevant to your planned research?

Desirable: *Before Submitting the Proposal*

External links: [Google dataset search](#), [Datacite Search](#), [RDMkit on Reusing Data](#), [RDMkit on Existing Data](#)

a. No

b. Yes

III.4 What existing data formats/types will you be using?

Horizon 2020 DMP Horizon Europe DMP Science Europe DMP

Have you identified types of data that you will use that are used by others too? Some types of data (for example "images" or "tables") are used by many different projects. For such data, often common standards exist (in our example "JPG" and "CSV" [comma separated values]) that help to make these data reusable. Are you using such common data formats?

Please make sure you list all the data types that are important for your project. You should make sure also to list the formats used in any data sets that you are re-using.

Desirable: *Before Submitting the Proposal*



FAIR Wizard: Tvorba DMP

4. Data security

Project members will not store data or software on computers in the lab or external hard drives connected to those computers. They will not carry data with them (e.g. on laptops, USB sticks, or other external media). All data centers where project data is stored carry sufficient certifications. All project web services are addressed via secure HTTP (<https://...>). Project members have been instructed about both generic and specific risks to the project.

The possible impact to the project or organization if information is lost is small. The risk of information leak in the project or organization is acceptably low. The risk of information vandalism in the project or organization is acceptably low.

We are not using any personal information.

The archive will be stored in a remote location to protect the data against disasters. The archive need to be protected against loss or theft. It is clear who has physical access to the archives.

We are running the project in a collaboration between different groups and institutes. A collaboration agreement that describes who can have access to what data in the project is set.



FAIR Wizard: Tvorba DMP

5. Ethical aspects

Ethical approvals

The following project require ethical approval:

- **Arsenic and Selenium Speciation Using Hyphenated Technique (ASSUHT)**
 - case number: *1111-2222-3333-4444*, status: *granted*

Data we produce

For the data we produce, the ethical aspects are as follows:

- **Levels of Arsen in soil**
 - It does not contain personal data.
 - It does not contain sensitive data.

Data we collect

We will not collect any data connected to a person, i.e. "personal data".



FAIR Wizard: Na co si dát pozor

- FAIR Wizard není vševědoucí - pracuje pouze s tím, co mu dáte
 - We will save data in we want to save data in Zenodo.
 - We will save data in Zenodo.
- **Vždycky si vyexportované DMP pročtěte a zkontrolujte**
 - Možnost editace ve Wordu či jiných textových programech

FAIR Wizard: Proč ho používat

- **Uspodnění vyplňování**
 - Minimum ručního vypisování informací
 - Žádné formátování
 - Šablona podle podmínek GAČR
- **Spolupráce**
 - Podobná práce jako v Google dokumenty
 - Všichni vidí nejnovější verzi
- **Podpora**
 - Nápoředy k otázkám - jdete správným směrem
 - Data stewardi, admin, návody



vypisování
celého
DMP ručně



pár
kliknutí ve
FAIR Wizardu

FAIR Wizard: Sdílení mezi institucemi

- Jednotlivé instance mezi sebou nekomunikují
 - Účty z UK instance se nepřihlásí na instanci AV a naopak
- Je možné sdílet jednotlivým osobám
 - Stejný princip jako Google docs
 - Sdílení odkazu s různými úrovněmi oprávnění


UK




UK





FAIR Wizard: Podpora

- **Návody a postupy**
 - [Detailní návod](#) jak ve FAIR Wizardu krok po kroku vyplnit DMP
 - Rozsáhlá komunita uživatelů
- **Data stewardi – podpora obsahová**
 - "Jaké jsou možnosti pro můj obor?"
 - "Nevím jestli je toto platné pro náš výzkum, jdeme na to správně?"
 - "Neznám X, můžete mi to prosím vysvětlit?"
- **Administrátor FAIR Wizard UK – podpora technická**
 - "Pomoc, nemůžu najít svůj DMP!"
 - "Nefunguje mi/nejde mi nastavit/nevím co použít-"



FAIR Wizard: Jak začít

The screenshot shows the FAIR Wizard interface. On the left is a sidebar with navigation options: 'Přehled', 'Projekty', 'Návod / How-to', 'Podpora / Support', 'Data Stewardí / Data Stewards', 'AI Assistant', and a user profile for 'Výzkumník UK'. The main content area has a header 'Vítejte, Výzkumníku!' with a sub-header 'Jako výzkumník vytváříte a spolupracujete na projektech.' and an illustration of a person with a backpack. Below this is a 'Nedávné projekty' section listing three items: 'Výzkum abrazie dentinu u pastorálních populací', 'Projekt MI.25.5', and 'DMP projektu Rosnička'. To the right is a 'Vytvořit projekt' section with a description of the project creation process and a 'Vytvořit' button. At the bottom right of the interface, the logos for 'UK' and 'FAIRWizard' are displayed.



[Odkaz na návod k používání FAIR Wizardu](#)



Podpora na UK





Centrum pro podporu open science

- Webové stránky: openscience.cuni.cz
- Kontakt: openscience@cuni.cz
- Školení, přednášky a konzultace
- Metodická podpora
 - Prověřování pochybných časopisů, pomoc s výběrem vhodného OA časopisu...
 - Nakládání s výzkumnými daty, tvorba DMP...
 - Právní podpora v oblasti autorsko-právní problematiky
- Elektronická [konference oa-l](#)





Fakultní data stewardi

KTF	datasteward@ktf.cuni.cz
FHS	datasteward@fhs.cuni.cz
FaF	datasteward@faf.cuni.cz
LFP	datasteward@lfp.cuni.cz
FTVS	datasteward@ftvs.cuni.cz
FSV	datasteward@fsv.cuni.cz
PřF	openscience@natur.cuni.cz
COŽP	datasteward@czp.cuni.cz





Univerzita
Karlova

Děkujeme za pozornost!



Dagmar Hanzlíková & Lucie Hošková
openscience@cuni.cz
Centrum pro podporu open science
Univerzita Karlova

