

Specifický vysokoškolský výzkum

Podmínky open science



UNIVERZITA KARLOVA
Centrum pro podporu open science

Dagmar Hanzlíková
Centrum pro podporu open science
19. 6. 2023, 15:30
openscience@cuni.cz



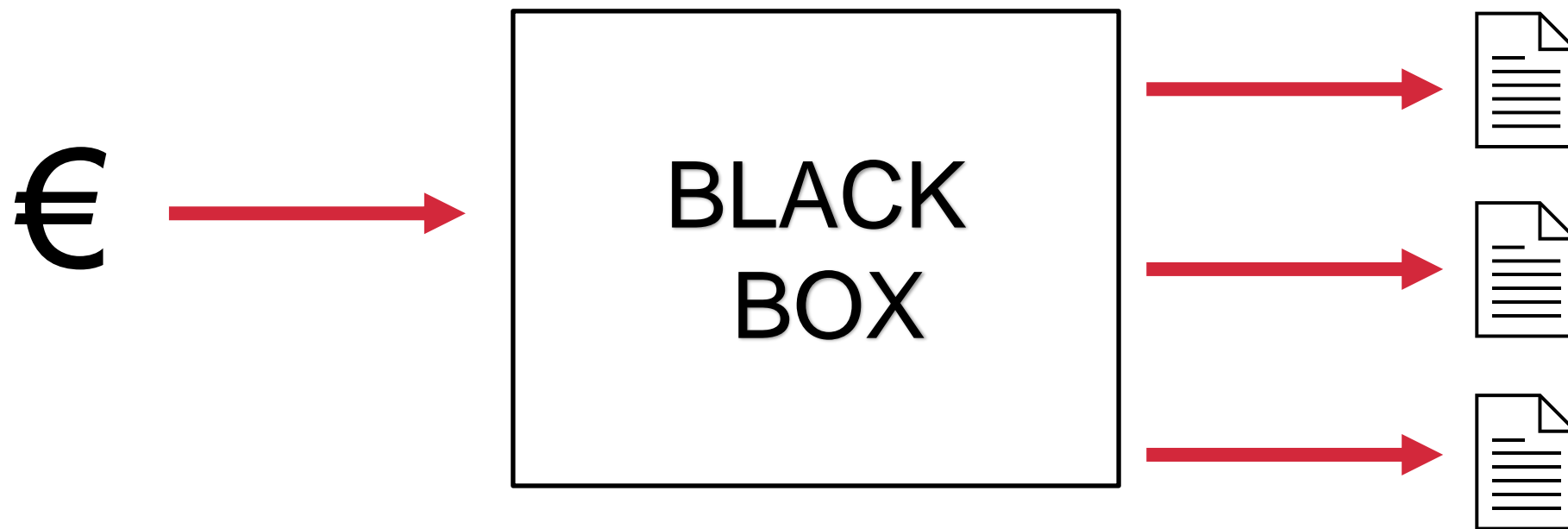


Open science



Open Science: Proč?

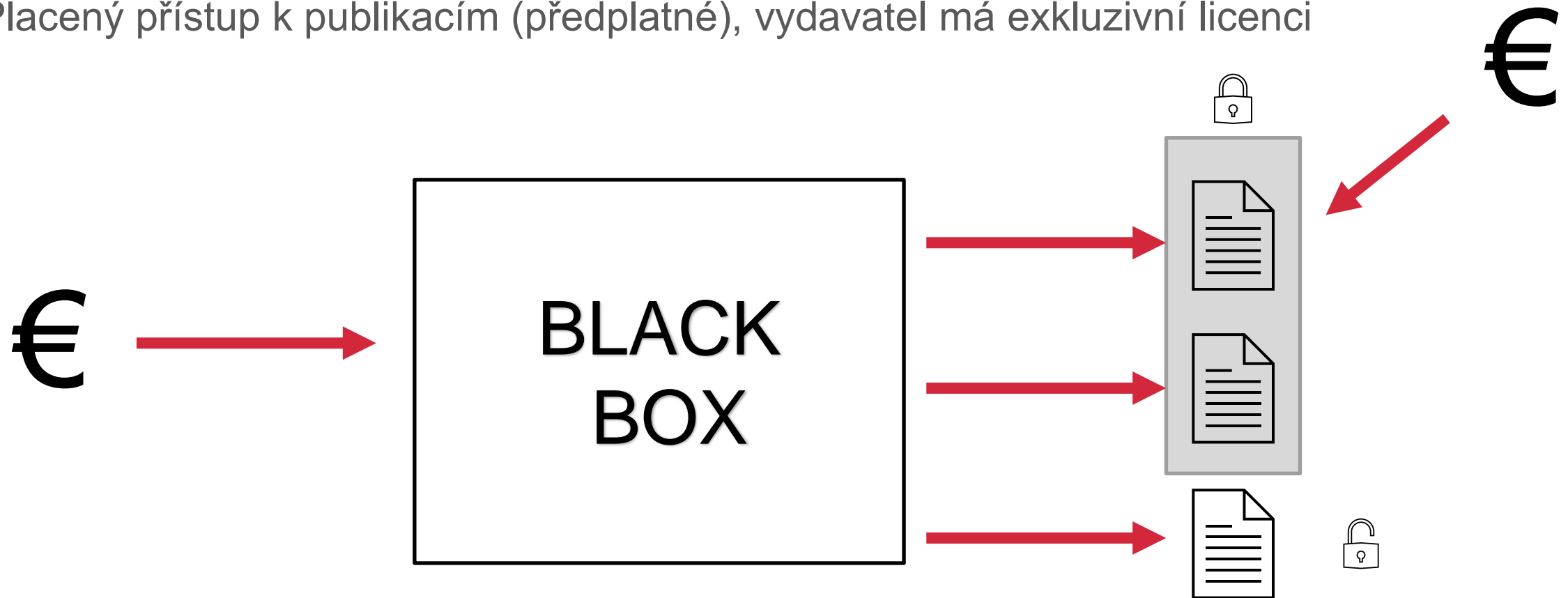
Jak vypadá věda nyní



Open Science: Proč?

...a často je to ještě horší...

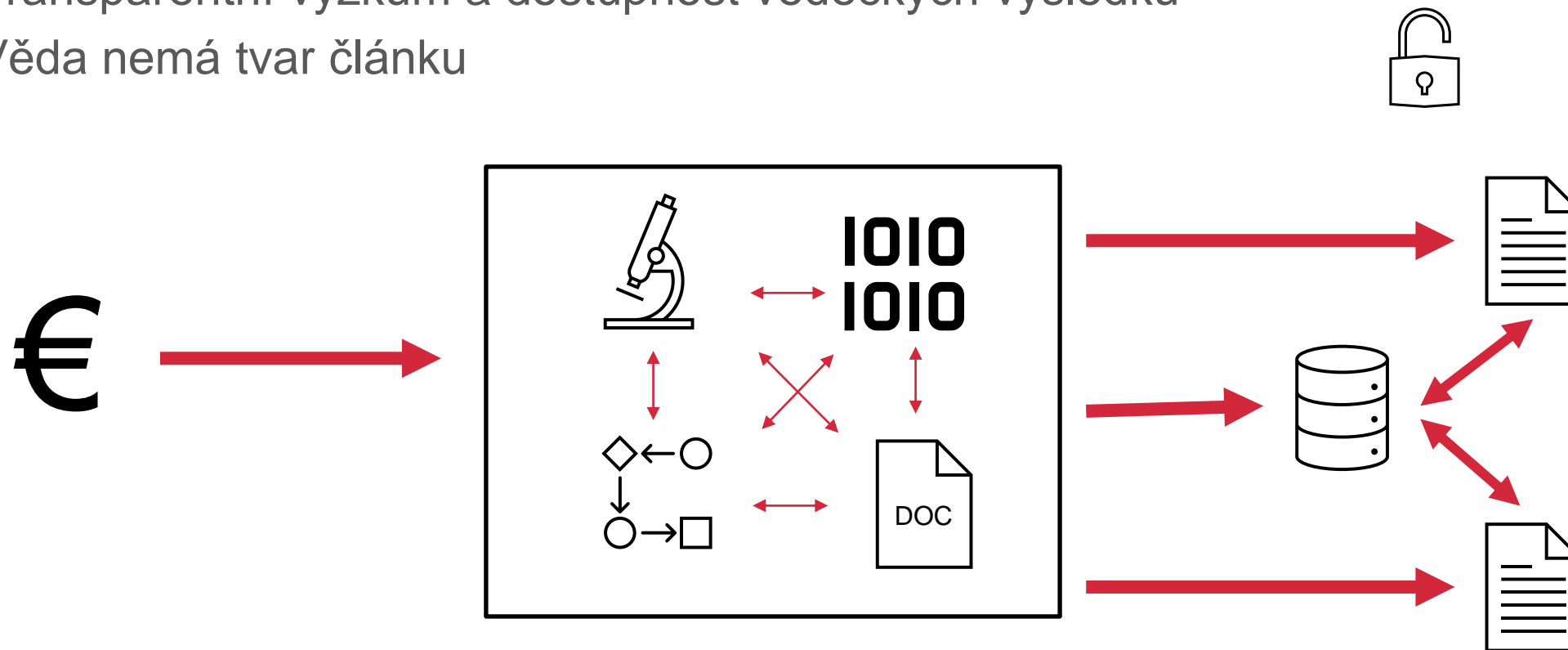
- Placený přístup k publikacím (předplatné), vydavatel má exkluzivní licenci



Open Science: Proč?

Hlavní cíle

- Transparentní výzkum a dostupnost vědeckých výsledků
- Věda nemá tvar článku



Open science

- Zpřístupnění výsledků výzkumu (publikace, data a další)
- Opětovná využitelnost výsledků výzkumu
- Transparentní a replikovatelný výzkum
- Efektivní využití zdrojů
- Nové spolupráce





Open access



Open access

7. Hlavní řešitel se zavazuje, že prostředky projektu nebudou použity na hrazení *Article Processing Charge* („APC“) a *Book Processing Charge* („BPC“). Výjimku z první věty tvoří APC u výstupů projektu, které budou publikovány v časopisech s *Article Influence Score* („AIS“) v prvním kvartilu v sekci/kategorii *Journal Citation Report* („JCR“), který je pěstován na fakultě³. Dále je doporučeno zveřejnit výsledky výzkumu v režimu open access prostřednictvím trvalé archivace⁴.

- APC uznatelné jen v prvním kvartilu
- Doporučené zveřejnění v režimu OA prostřednictvím **autoarchivace**



Hrazení publikačních poplatků (APC)

APC = article processing charge, publikační poplatek

- Karger Publishers
- Taylor & Francis
- Springer Nature
- Wiley
- A další...



Uznatelné náklady jen v **prvním kvartilu!**



Autoarchivace: Repozitář publikační činnosti UK

- Ukládání publikačních výsledků vzniklých na UK
 - Forma výsledku: abstrakt, článek, kapitola v knize, kniha...
 - Verze: rukopis, preprint, postprint, vydavatelská verze
- Varianty dostupnosti souboru
 - Open access
 - Open access s embargem
 - Restricted access
- Vkládání přes IS Věda – OBD (FK publikační činnosti)
- Další informace o repozitáři



publications.cuni.cz



Správa výzkumných dat



Správa výzkumných dat

8. Hlavní řešitel se zavazuje, že zajistí správu výzkumných dat v souladu s principy FAIR⁵ a k vedení záznamů o výzkumných datech získaných během projektu, známé rovněž jako „data management plan“. Dále se zavazuje k poskytnutí informací prostřednictvím data management plánu o dostupnosti a způsobu šíření výsledků výzkumu a výzkumných dat v souladu s fakultními a univerzitními předpisy a se zásadou, že výsledky výzkumu a výzkumná data nejsou zveřejňovány pouze v odůvodněných případech⁶.

- FAIR principy
- Data management plan
- As open as possible, as closed as necessary

Co jsou výzkumná data?

= informace, které byly vytvořeny 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
 - Vzorčky, exempláře





FAIR principy

FAIR data

- **F**indable (naležitelná)
- **A**ccessible (přístupná)
- **I**nteroperable (interoperabilní)
- **R**eusable (znovuvyužitelná)

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**

- Perzistentní identifikátory (PID)
- Metadata
- 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)

Perzistentní identifikátory

Kdo/co může mít PID?

- Osoby (ORCID, ResearcherID)
- Publikace a časopisy (DOI, ISSN, ISBN)
- Digitální objekty (DOI, handle)
- Organizace (ROR, Ringgold ID, ISNI)
- A další...

Identifikátory.cz



Výzkumným datům přiděluje perzistentní identifikátor typicky **repozitář**

Metadata

= *data popisující jiná data*

- Strojově čitelná
- Podle metadat záznamu se vyhledává
 - špatná metadata => nenalezitelná data!
- Řízené slovníky (= ontologie)
- Užívání PID v metadatech (autor, organizace, ontologie...)
- Popisujte data takovými termíny, které byste zadali do vyhledávače, kdybyste je sami chtěli najít

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



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



FAIR: Reusable

= optimalizace dat pro co nejvyšší možnost *opětovného využití*

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

Teplota

- 36,6
- 36,1
- 37,8
- 41,5

Jak vytvořit FAIR data?

- Dokumentace a metadata
- Standardizované výrazy
- Perzistentní identifikátory
- Otevřené souborové formáty
- Přístup k datům
- Licence





Data management plan

Data management plan (DMP)

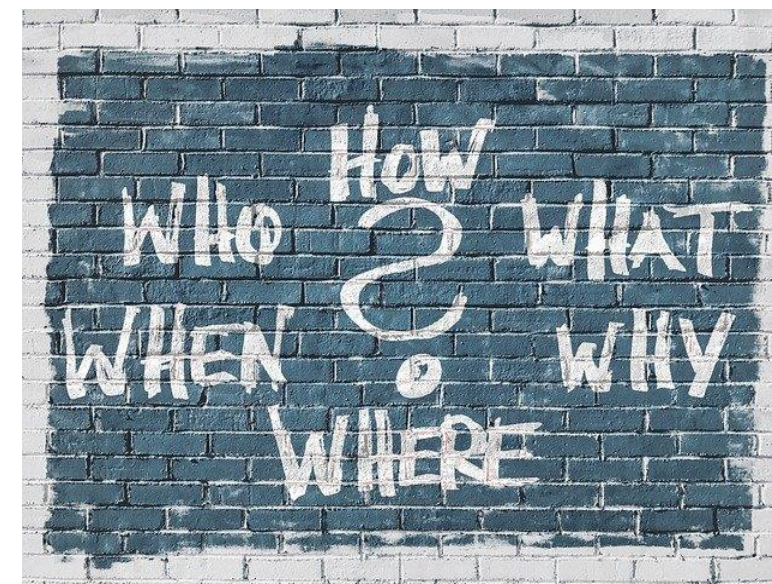
= dokument, který specifikuje, *jaká data a jakým způsobem budou během výzkumu vytvářena*, a obsahuje informace o jejich *dostupnosti a využití*



- 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



Horizon Europe
DMP template



Proč vytvářet DMP?

- Pomáhá předvídat **potenciální problémy**
- Kontrola, že máte všechno potřebné **vybavení a podporu**
- Zajištění, že data jsou **přesná, úplná a spolehlivá**
- **Kontinuita** a **konzistence** dlouhodobých projektů
- Pomoc při **sdílení dat**

• version B: only changing the article, eg. OIP_a-WN1 → OIP_the-WN1

↓ sentence ↓ speaker ↓ article ↓ white noise → context 1

Nástroje pro tvorbu DMP

A role of striatal M4 muscarinic acetylcholine receptors in learning of stereotypic motor sequences

The screenshot shows the DMPonline interface for a project titled "A role of striatal M4 muscarinic acetylcholine receptors in learning of stereotypic motor sequences". The top navigation bar includes "Project Details", "Plan overview", "Write Plan", "Share", and "Download". A progress indicator shows "13/13 answered". The main content area is titled "Data Collection (2 / 2)" and asks "What data will you collect or create?". It lists several data types: "master sheet", "operant conditioning box program", and "operant conditioning box logs". A sidebar on the right provides "Guidance" and "Comments" for the selected data type, with "DCC" and "FAIRFAIR" highlighted. Red arrows point to the "Write Plan" button, the progress bar, the "DCC" and "FAIRFAIR" options, and the "Guidance" section.

DMPonline



The screenshot shows the Data Stewardship Wizard interface for a project titled "Vliv ochyťovacího zaklínadla na výsledky testů". The top navigation bar includes "Questionnaire", "Metrics", "Preview", "Documents", and "Settings". A sidebar on the left shows "Knowledge Models" and "Projects". The main content area is titled "Current Phase" and "Chapters". The "Chapters" list includes: I. Administrative information (1), II. Re-using data (1), III. Creating and collecting data (6), IV. Processing data (3), V. Interpreting data (2), VI. Preserving data (5), and VII. Giving access to data (2). The "Creating and collecting data" chapter is highlighted. The questionnaire questions are: "2. What existing encodings/terminologies/vocabularies/ontologies will you be using?", "3. Will you be using new types of data?", and "4. How will you be collecting and keeping your metadata?". Red arrows point to the "Horizon 2020 DMP" button, the "Interoperability" option, and the "Science Europe DMP" button.

Data Stewardship Wizard





Sdílení výzkumných dat

Sdílení výzkumných dat

plan“. Dále se zavazuje k poskytnutí informací prostřednictvím data management plánu o dostupnosti a způsobu šíření výsledků výzkumu a výzkumných dat v souladu s fakultními a univerzitními předpisy a se zásadou, že výsledky výzkumu a výzkumná data nejsou zveřejňovány pouze v odůvodněných případech⁶.



- 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

As open as possible...
...as closed as necessary



Otevřený přístup k datům

As **open** as possible...

- Podkladová data k publikacím
- Samostatné datové sady

...as **closed** as necessary (= v odůvodněných případech)

- Ochrana osobních údajů
- Oprávněné obchodní zájmy a práva duševního vlastnictví třetích stran
- Bezpečnost státu
- Rozpor s oprávněnými zájmy příjemce, včetně komerčního využití dat



Omezení sdílení by mělo být **vysvětleno v DMP!**

Otevřený přístup k datům: Jak sdílet?

- Anonymizace dat
- Opatřit data **licencí Creative Commons Attribution** (CC-BY) nebo ekvivalentní
- Přidělit datům **trvalý identifikátor**
- Doporučený formát **citace**
- Zároveň s daty sdílet **dokumentaci**



Otevřený přístup k datům: Jak sdílet?

Způsoby sdílení dat

- ~~Supplementary materials~~
- Datový repozitář ←
- Datový časopis (data journal)



Uložení dat do repozitáře

1) Oborový repozitář

- Data zaměřená na konkrétní obor
- Registr datových repozitářů
re3data.org

2) Institucionální repozitář

- Spravuje instituce
- Např. [ASEP](#)
- Datový repozitář UK **bude!**

3) Obecný repozitář

- Bez ohledu na oborové zaměření
- Např. [Zenodo](#), [Figshare](#),
[Dryad](#)

Generalist Repository
Comparison Chart





Filter

Subjects

- Humanities and Social Sciences (4)
 - Humanities (4)
 - Linguistics (4)
- Engineering Sciences (3)
 - Computer Science, Electrical and System Engineering (3)
 - Systems Engineering (1)
 - Human Factors, Ergonomics, Human-Machine Systems (1)
 - Computer Science (3)
 - Artificial Intelligence, Image and Language Processing (2)

Content Types

Countries

API

Certificates

Data access

Data access restrictions

Database access

Database access restrictions

Data licenses

Data upload

Data upload restrictions

Enhanced publication

Institution responsibility type

Institution type

Keywords

Metadata standards

czech linguistics

← Previous 1 Next →

Found 4 result(s)

Czech National Corpus

CNK

Subject(s)

Humanities and Social Sciences Humanities Linguistics

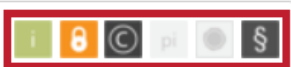
Content type(s)

Databases Audiovisual data Standard office documents

Country

European Union Czech Republic

The aim of the project is systematic mapping of Czech and other languages in comparison with Czech. CNC corpora are accessible to everybody interested in studying the language after free registration.



LINDAT/CLARIN repository

Subject(s)

Humanities Linguistics Artificial Intelligence, Image and Language Processing Humanities and Social Sciences Computer Science
Computer Science, Electrical and System Engineering Engineering Sciences

Content type(s)

Standard office documents Audiovisual data Plain text Structured text Archived data Source code

Country

Czech Republic European Union

LINDAT/CLARIN is designed as a Czech "node" of Clarin ERIC (Common Language Resources and Technology Infrastructure). It also supports the goals of the META-NET language technology network. Both networks aim at collection, annotation, development and free sharing of language data and basic technologies between institutions and individuals both in science and in all types of research. The Clarin ERIC infrastructural project is more focused on humanities, while META-NET aims at



Search

Browse

Suggest

Resources

Contact

Search

Toggle short help

Sort by



Uložení dat do repozitáře

Academia.edu

Dropbox

SharePoint

Osobní stránky autora

Google Drive

Stránky vydavatele

ResearchGate



Data journal

= recenzovaný časopis, který publikuje příspěvky *popisující datové sady* uložené v repozitáři

- Vědecky cenné datové soubory
- Data uložená v repozitáři
- Možnost snadného citování a vykazování
- Čistě datové časopisy nebo kombinované
- Např. *Journal of Open Humanities Data*, *Research Data Journal for the Humanities and Social Sciences*, *Journal of Open Research Software*

Fakultní a univerzitní předpisy

Etický kodex UK

*[p]ři své vědecké a tvůrčí práci reflektuje nejnovější vývoj ve svém oboru, odpovídá za přiměřenost, přesnost a objektivitu použitých metod a dbá na to, aby nedošlo k jejich zkreslování. **Pečuje o data či zkoumané materiály, o jejich přesný popis a standardní způsoby jejich ochrany a archivace. Dbá o to, aby postupy i výsledky vědecké a tvůrčí práce byly přezkoumatelné.** Dodržuje specifická pravidla profesionální etiky daného oboru. (Část III, bod 4.)*

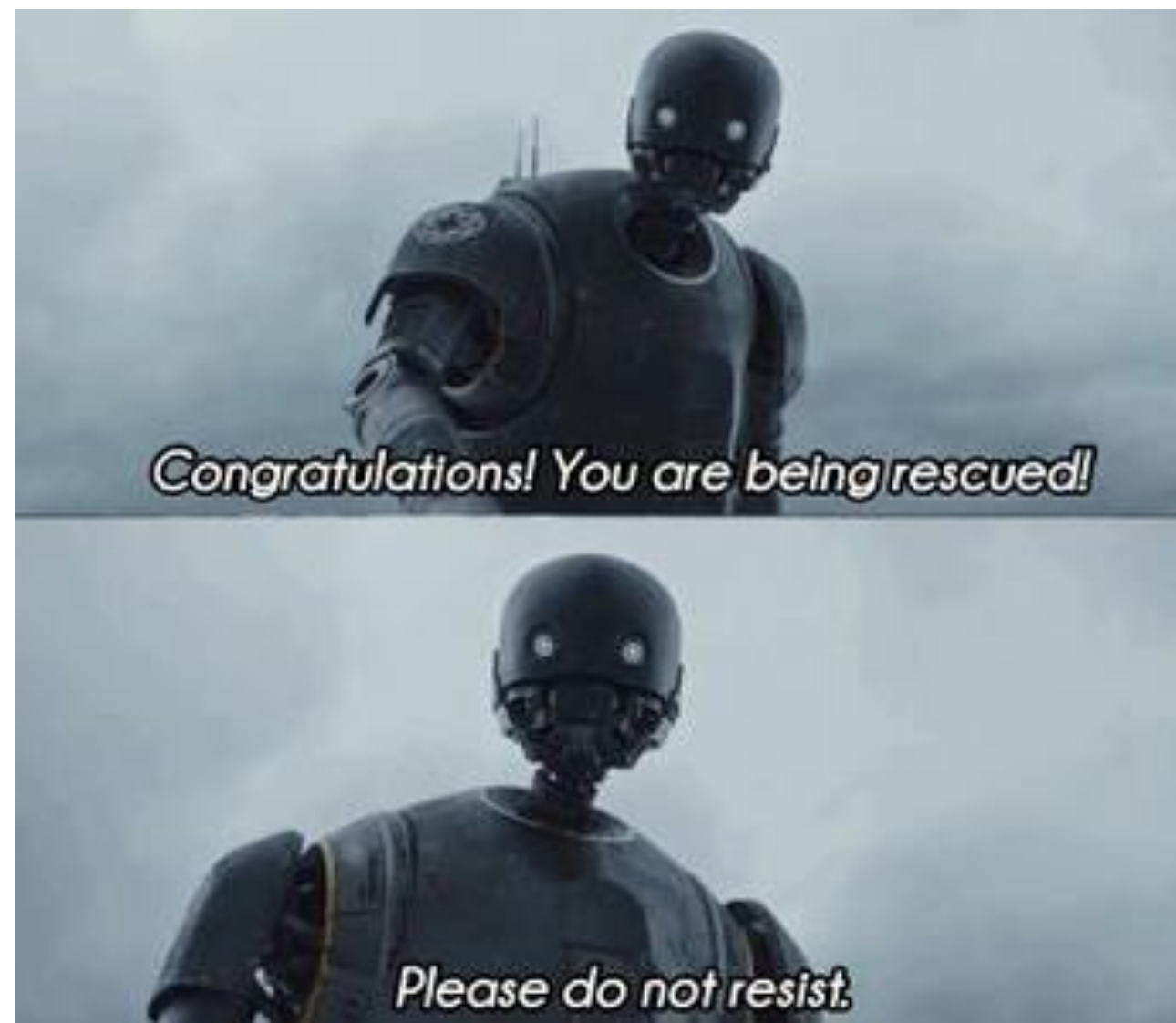
Politika



Podpora na UK

- [Webové stránky](#)
- Konzultace DMP
- Školení, přednášky
- Elektronická konference [oa-1](#)

openscience@cuni.cz



Prostor pro diskuzi





Děkujeme za pozornost.

Centrum pro podporu open science
Ústřední knihovna UK
openscience@cuni.cz

