

# Úvod do FAIR dat

Michal Růžička <[ruzicka@ics.muni.cz](mailto:ruzicka@ics.muni.cz)>



Spolufinancováno  
Evropskou unií



# Anotace

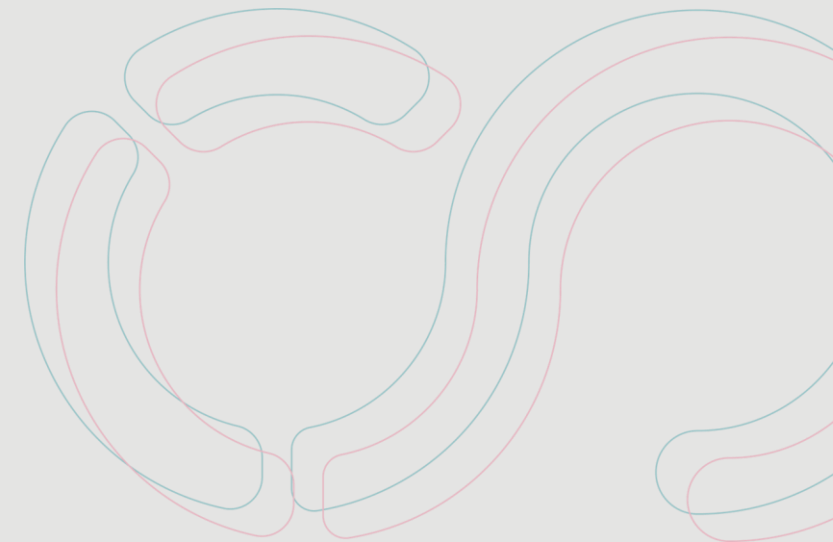
V rámci této přednášky si představíme **životní cyklus výzkumných dat** zasazený do **širšího kontextu Open Science** a soudobých **požadavků donátorů** výzkumu na **práci s daty**.

Představíme si **FAIR principy** a **nástroje pro jejich naplňování**, včetně **perzistentních identifikátorů**. Ukážeme si **možnosti pro ukládání a sdílení dat** v akademických institucích a české e-infrastruktúře.

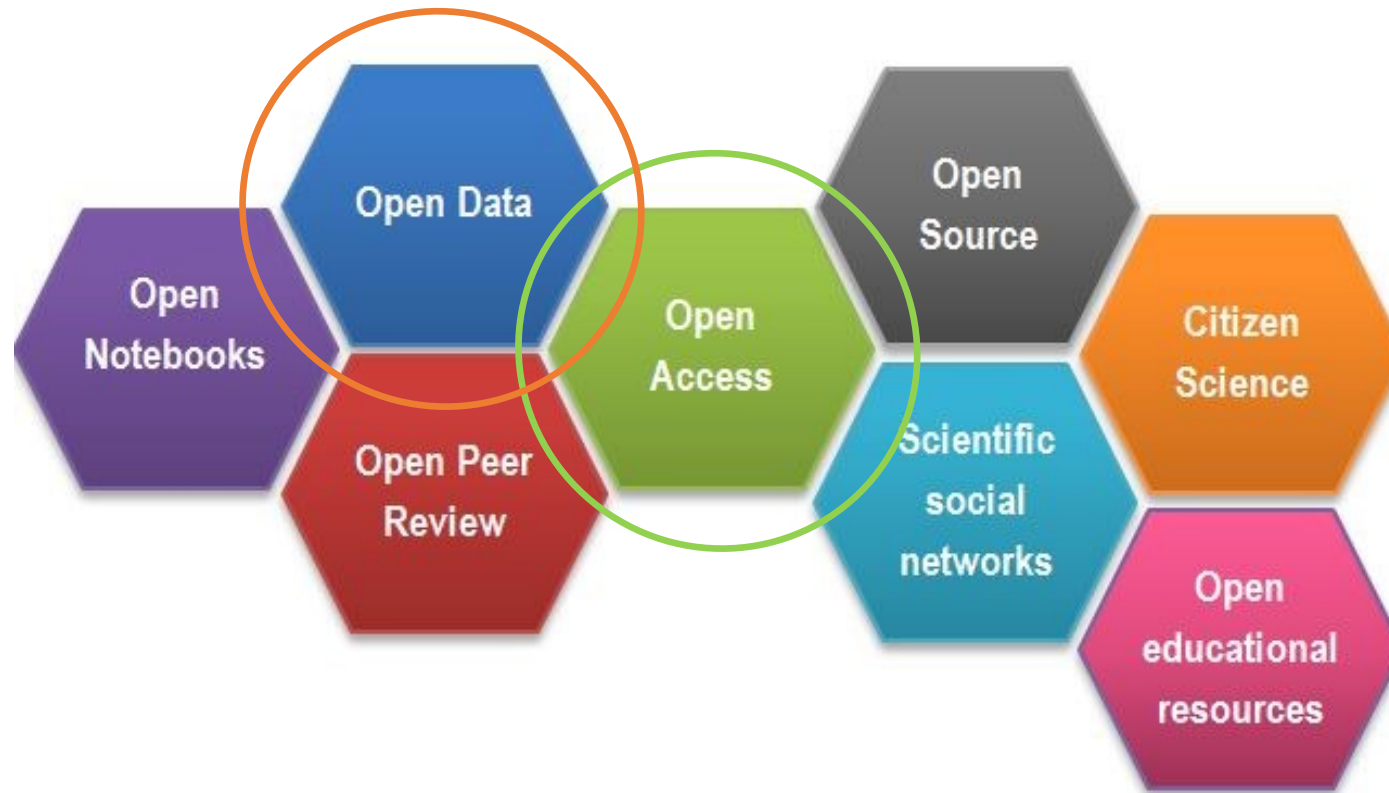
# Co si z přednášky odnesete

- Seznámíte se s jednotlivými **prvky životního cyklu výzkumných dat**.
- Seznámíte se s jednotlivými **FAIR principy**.
  - A **nástroji**, které jim umožní je naplňovat v jejich **každodenní akademické praxi**.
- Poznají **služby pro práci s daty** nabízené **českou e-infrastrukturou**.
- Seznámí se s **druhy perzistentních identifikátorů** a jejich významu a použití v kontextu práce **s výzkumnými daty**.
- Dozví se o **aktivitách implementace EOSC v ČR** a možnostech, které jim to přinese pro práci s výzkumnými daty.

# Kontext



# Schéma Open Science



Zdroj: <https://www.fosteropenscience.eu/node/1420>

**Politika Open Access**

- **OA1:** Uchovávání vědeckých publikací MU
- **OA2:** Zvýšení dostupnosti vědeckých publikací MU

**Specializovaná podpora pro rozvoj Open Access**

- **OA3:** Repozitář MU
- **OA4:** Vydávání otevřených publikací

**Financování Open Access**

- **OA5:** Fond OA MU
- **OA6:** Adaptace EIZ na transformační model

# Příklad cílů Open Access (Masarykova univerzita)

**Politika FAIR Data**

- **FD1:** Strategie pro správu a zpřístupňování výzkumných dat MU
- **FD2:** Evidence výzkumných dat

**Specializovaná podpora pro rozvoj FAIR Data**

- **FD3:** Datový repozitář MU
- **FD4:** Podpora pro správu a zpřístupňování výzkumných dat

**Infrastruktura FAIR Data**

- **FD5:** Infrastruktura pro výzkumná data

**Podpora pro vědce**

- **OS1:** DMP+
- **OS2:** Lidské zdroje
- **OS3:** Vzdělávání a informovanost
- **OS4:** Služby Open Science

**Interní procesy**

- **OS5:** Politiky na jednotlivých HS
- **OS6:** Open Science v kontextu evaluace vědy
- **OS7:** Open Science a grantová schémata MU

**Standardy**

- **OS8:** DOI
- **OS9:** ORCID
- **OS10:** OpenAIRE
- **OS11:** AOAP, ACAN

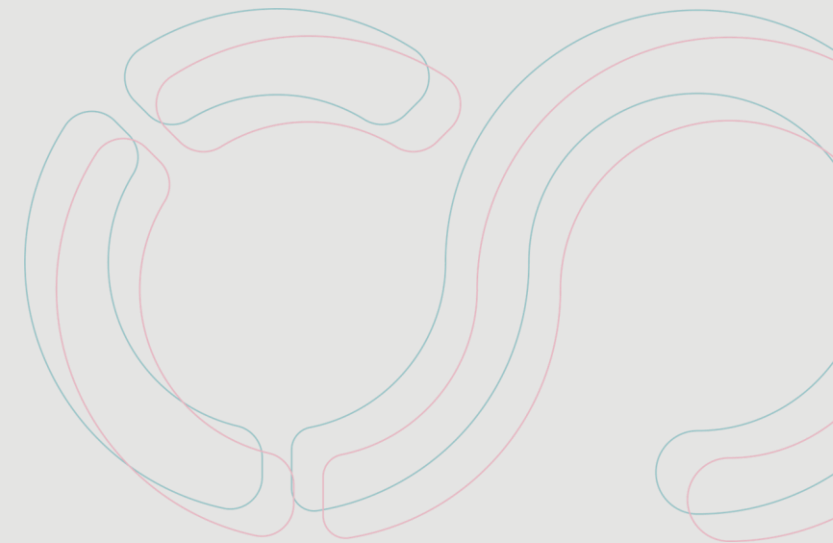
**Spolupráce**

- **OS12:** Národní iniciativy
- **OS13:** Mezinárodní iniciativy

# Open Science a projekty

- **Podmínky:**
  - Horizon Europe
  - Exceles
  - OP JAK – Špičkový výzkum
- Povinné/nepovinné **principy**
  - **Povinné:** Open Access, Data Management (DMP)
  - **Nepovinné:** Citizen Science, Open Source, Preprints, ...

# Životní cyklus výzkumných dat





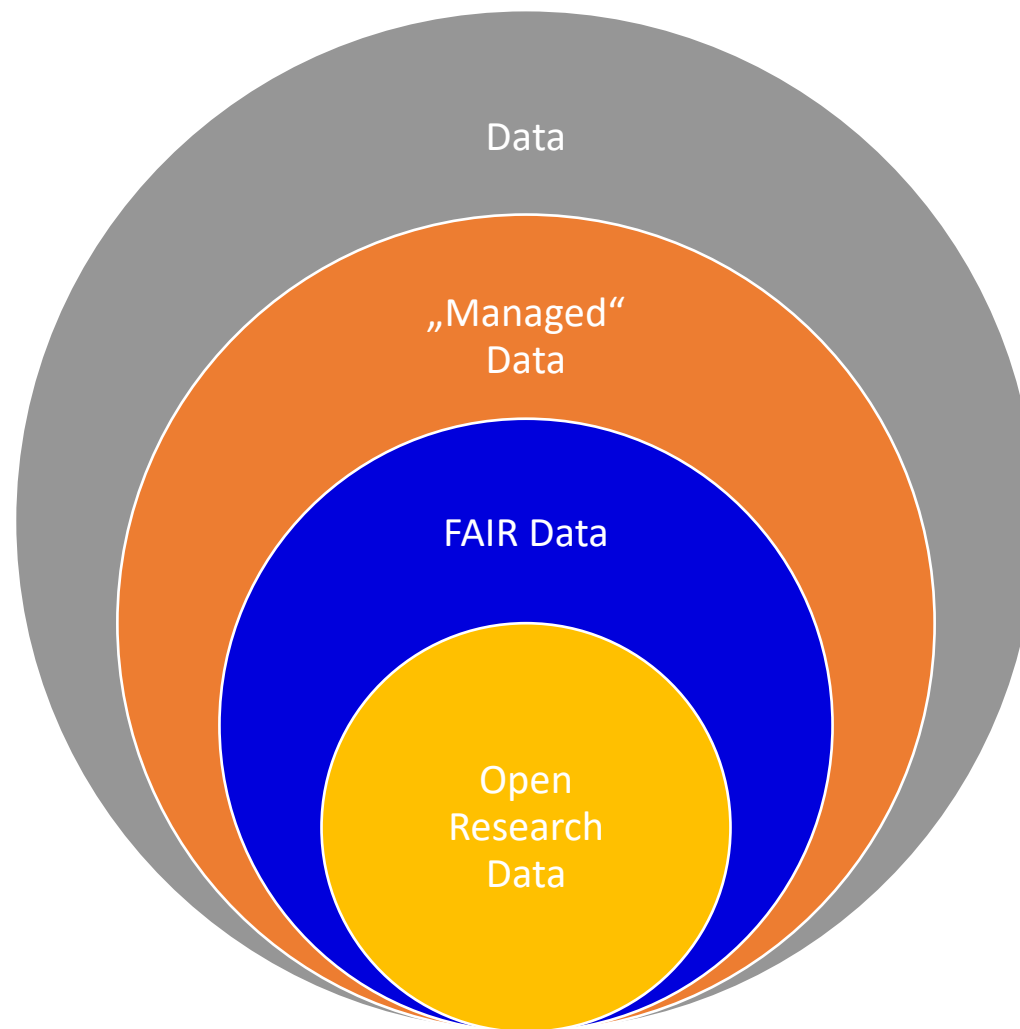
# Životní cyklus výzkumných dat



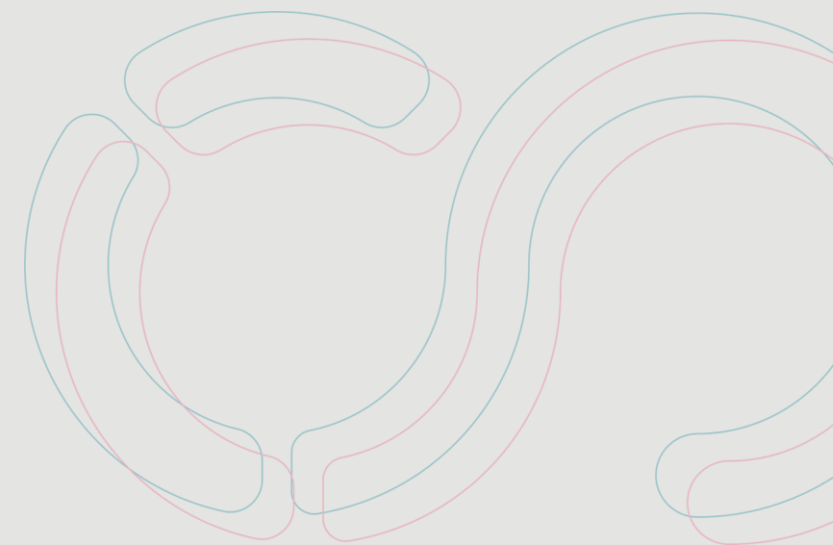
- Jaká **data (znovu) používáte**
  - včetně **licencí**, které vám to umožňují,
- jaká **data generujete** a jakým způsobem,
- **kde je ukládáte, zálohujete, dlouhodobě uchováváte**,
- jak je **trvale a jedinečně identifikujete**,
- jak je **zpracováváte**,
- jak je **analyzujete**,
- kde je **zveřejňujete a sdílíte**,
- kdo tohle všechno **zaplatí**;
- **o čem data skutečně jsou**,
- **k čemu jsou data (ne)vhodná**,
- **kdo může data znovu použít**,
- jaká konkrétní data podporují vaše výsledky,
- jak je použít pro **opakování** vašich **experimentů** atd.

Zdroj: ELIXIR RDMkit, <https://rdmkit.elixir-europe.org/>

# Úrovně dat



# FAIR principy



# FAIR principy – Findable

The first step in (re)using data is to find them. Metadata and data should be **easy to find** for both **humans** and **computers**. **Machine-readable metadata** are essential for automatic discovery of datasets and services, so this is an essential component of the [FAIRification process](#).

- F1. (Meta)data are assigned a **globally unique** and **persistent identifier**.
- F2. Data are described with **rich metadata** (defined by R1 below).
- F3. **Metadata** clearly and **explicitly include** the **identifier** of the **data they describe**.
- F4. (Meta)data are **registered** or **indexed** in a **searchable resource**.

Zdroj: GO FAIR, <https://www.go-fair.org/fair-principles/>

# FAIR principy – Accessible

Once the user finds the required data, she/he/they need to know **how can they be accessed**, possibly **including authentication and authorisation**.

- A1. (Meta)data are **retrievable by their identifier** using a **standardised communications protocol**.
  - A1.1 The **protocol is open, free, and universally implementable**.
  - A1.2 The protocol **allows for an authentication and authorisation** procedure, where necessary.
- A2. **Metadata are accessible, even when the data are no longer available**.

Zdroj: GO FAIR, <https://www.go-fair.org/fair-principles/>

# FAIR principy – Interoperable

The data usually need to be **integrated with other data**. In addition, the data need to **interoperate with applications** or **workflows** for analysis, storage, and processing.

- I1. (Meta)data use a **formal, accessible, shared, and broadly applicable** language for **knowledge representation**.
- I2. (Meta)data use **vocabularies** that follow FAIR principles.
- I3. (Meta)data **include** qualified **references to other** (meta)data.

Zdroj: GO FAIR, <https://www.go-fair.org/fair-principles/>

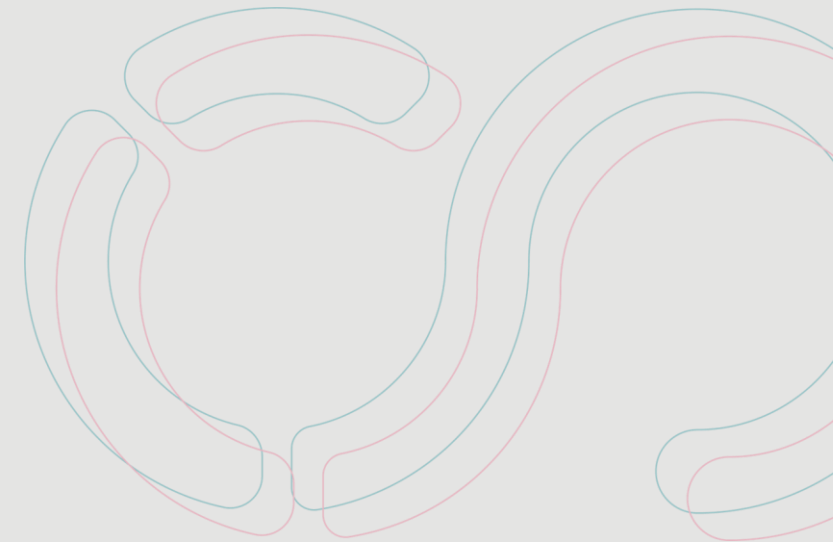
# FAIR principy – Reusable

The ultimate goal of FAIR is to optimise the **reuse of data**. To achieve this, metadata and data should be **well-described** so that they **can be replicated and/or combined** in different settings.

- R1. (Meta)data are richly **described with** a plurality of **accurate** and **relevant attributes**.
  - R1.1. (Meta)data are released with a clear and accessible **data usage license**.
  - R1.2. (Meta)data are associated with **detailed provenance**.
  - R1.3. (Meta)data meet **domain-relevant** community **standards**.

Zdroj: GO FAIR, <https://www.go-fair.org/fair-principles/>

# Nástroje a zdroje pro podporu FAIR principů





# Nástroje a zdroje pro podporu FAIR principů

- **The ELIXIR Research Data Management Kit (RDMkit)**

- <https://rdmkit.elixir-europe.org/>
- Online **průvodce dobrou praxí** správy dat v přírodních vědách.
- **Rozcestník** na řadu externích **zdrojů**.
  - **Tematické dělení** dle **fází životního cyklu dat**,
  - **role osoby** v procesu práce s daty,
  - **úlohy s daty** (analýza, publikování, kontrola kvality, ...).
  - Včetně odkazů na **národní zdroje**, relevantní **software** a **vzdělávací materiály**.



- Data management
  - Data life cycle
  - Your role
  - Your domain
  - Your tasks**
    - Compliance monitoring
    - Data analysis
    - Data management plan
    - Data organisation
    - Data protection
    - Data publication
    - Data quality
    - Data storage
    - Data transfer
    - Documentation and metadata
    - Existing data
    - Identifiers
    - Licensing
    - Machine actionability
    - Sensitive data
- Tool assembly**
  - National resources
  - All tools and resources
  - All training resources

# Nástroje a zdroje pro podporu FAIR principů (2)

- **UK Data Service – Research data management**

- <https://www.ukdataservice.ac.uk/manage-data.aspx>
- Stručný průvodce s odkazy na **externí zdroje** k různým úlohám při správě výzkumných dat.
- Obsahuje např.
  - **checklist pro DMP**,
  - odkazy na **legislativu** relevantní pro správu výzkumných dat,
  - **doporučení** k určování nákladů na správu dat apod.

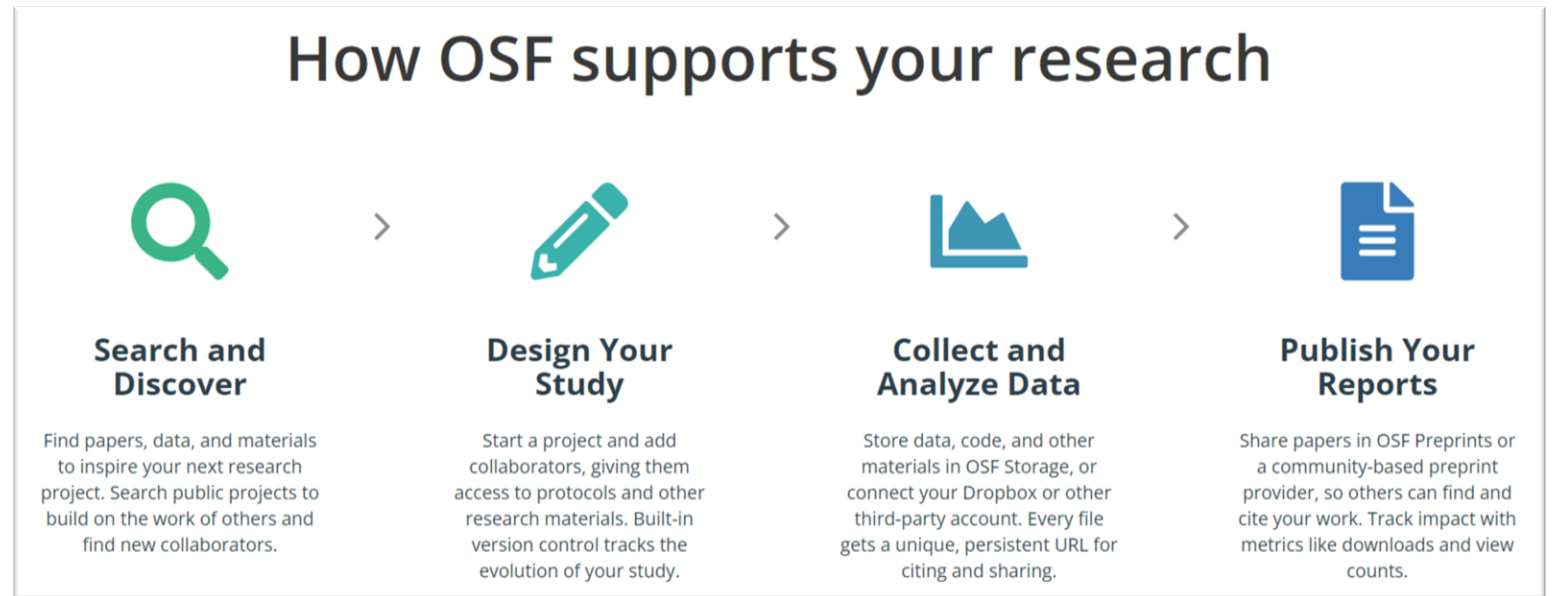


Rights in data
Document your data
Format your data
Store your data
Collaborative research
Data lifecycle
Plan to share
Data Protection
Ethical issues
Anonymisation

# Nástroje a zdroje pro podporu FAIR principů (3)

- **OSF**

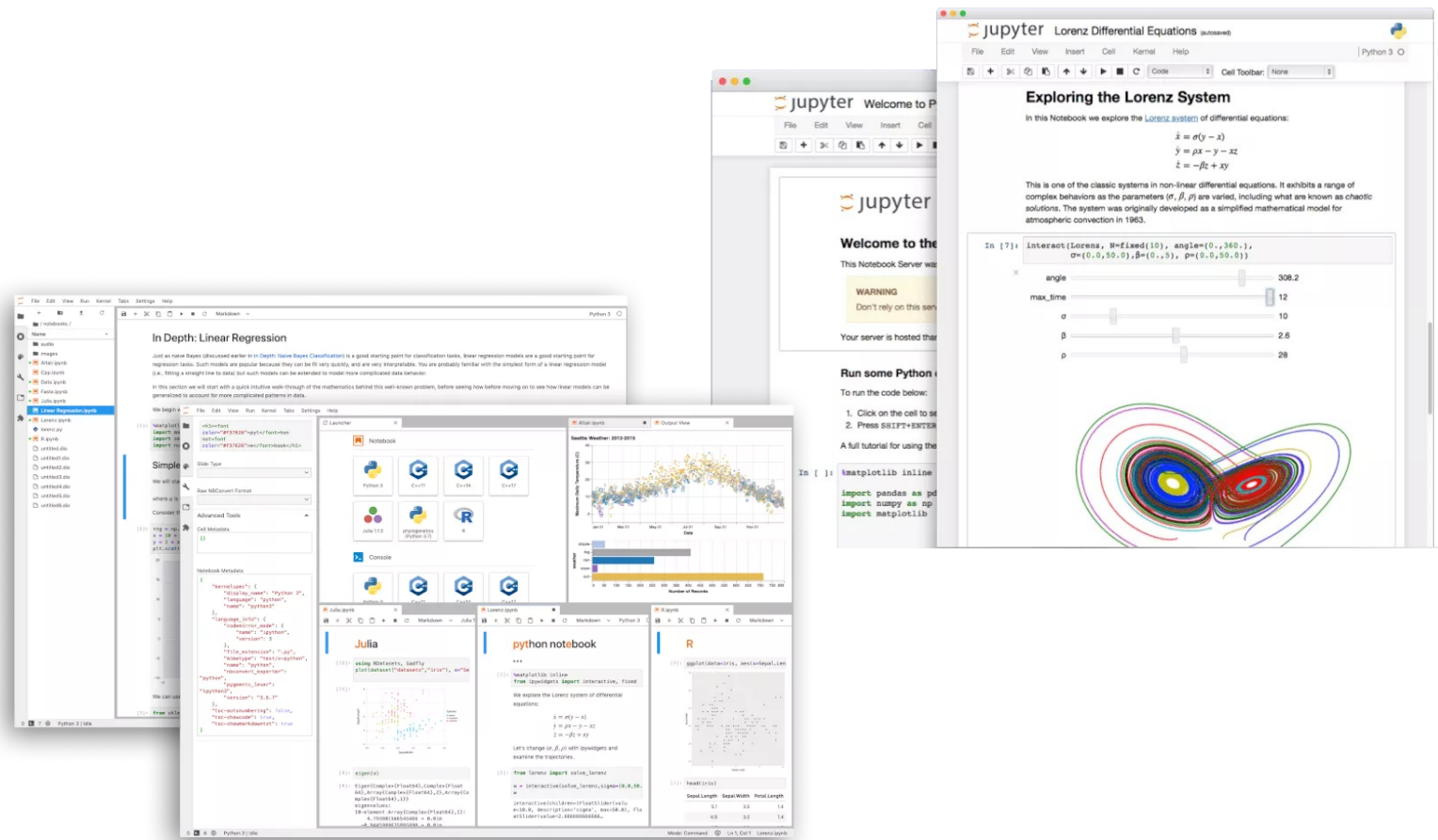
- <https://osf.io/>
- **Platforma pro podporu workflow vědeckého výzkumu.**



# Nástroje a zdroje pro podporu FAIR principů (4)

## • JupyterLab

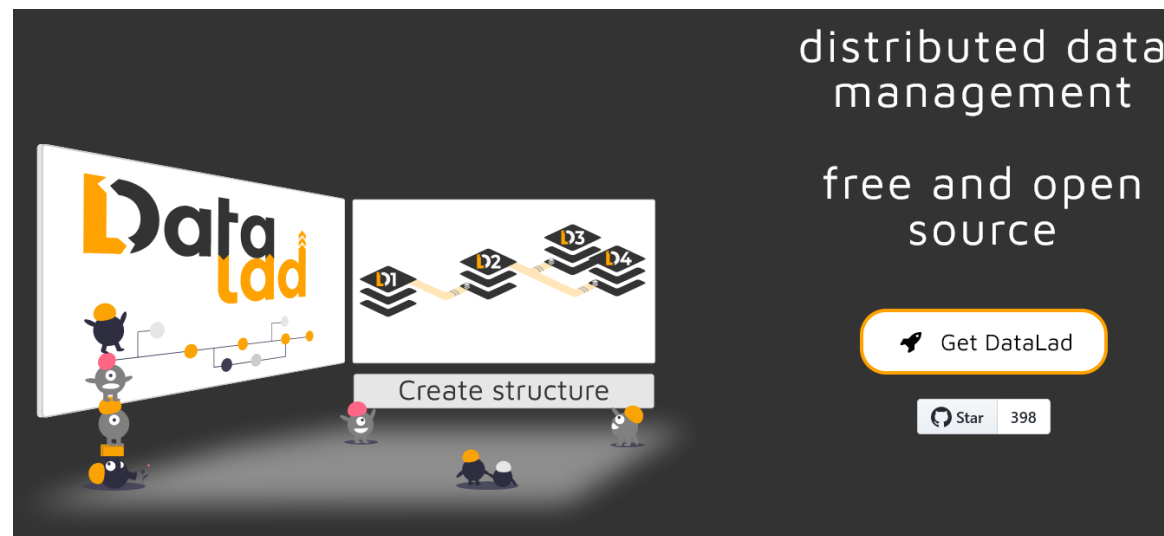
- <https://jupyter.org/>
- Dokumentované zpracování dat umožňující sdílení.



# Nástroje a zdroje pro podporu FAIR principů (5)

## • DataLad

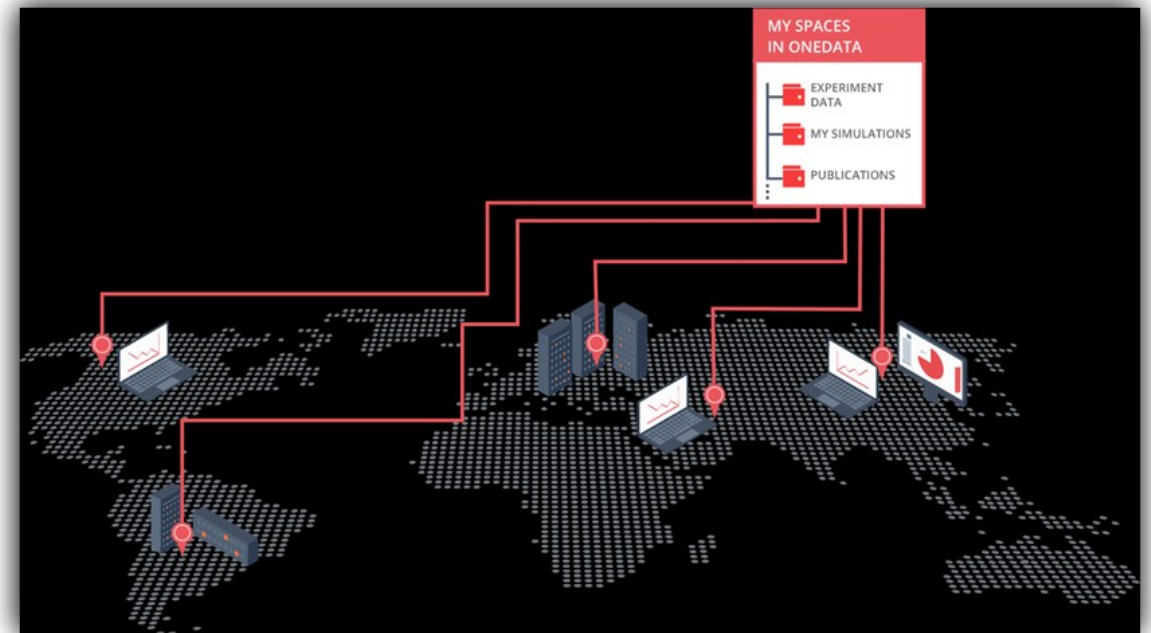
- <https://www.datalad.org/>
- DataLad is a free and open-source **distributed data management** system that **keeps track** of your data, creates structure, **ensures reproducibility**, supports collaboration, and integrates with widely used data infrastructure.



# Nástroje a zdroje pro podporu FAIR principů (6)

## • Onedata

- <https://onedata.org/>
- Perform **heavy computations** on **huge datasets**. **Access** your data in a dropbox-like fashion **regardless of its location**. **Publish** and **share** your results with public or closed communities.



# Úložiště pro výzkumná data v instituci a české eInfrastruktuře

Příklad: Masarykova univerzita, CESNET

# Kategorizace úložišť na MU a doporučení pro jejich využívání

- Datová **úložiště na MU**: <https://it.muni.cz/kategorie/datova-uloziste>
- **Doporučení pro užívání úložišť**:  
<https://it.muni.cz/prehledy/doporuceni-pro-uzivani-ulozist>
  - **Kategorizace dat.**
  - Kategorizace **úložišť**.
  - **Vhodnost úložišť pro různé typy dat.**



# Úložiště na MU – kategorizace dat

KATEGORIE	POPIS	PŘÍKLADY
<b>Veřejná data</b>	<ul style="list-style-type: none"><li>• Data zpřístupnitelná komukoliv bez jakýchkoliv omezení, např. veřejně vystavená na internetu.</li><li>• Jejich zveřejnění nepředstavuje žádné ohrožení pro MUNI nebo jiné instituce či osoby.</li></ul>	<ul style="list-style-type: none"><li>• Prezentace z veřejných přednášek;</li><li>• veřejně přístupné výzkumné zprávy;</li><li>• open-source software;</li><li>• veřejná výzkumná data;</li><li>• propagace, veřejné informace o službách.</li></ul>
<b>Interní data</b>	<ul style="list-style-type: none"><li>• Data určená jen pro vnitřní potřebu obecně definované skupiny osob (např. spolupracovníci projektu, pracovníci instituce apod.).</li><li>• Nevyžadují však zvláštní regulaci nebo ochranu (ze zákona, dle smlouvy apod.).</li><li>• Zpřístupnění mimo danou skupinu nezpůsobí přímou škodu (finanční, morální, právní apod.).</li></ul>	<ul style="list-style-type: none"><li>• Interní korespondence;</li><li>• zápisy z jednání;</li><li>• vnitřní regulace a předpisy;</li><li>• vnitřní plány práce, poznámky apod.;</li><li>• nedokončené/nepublikované výzkumné zprávy.</li></ul>

# Úložiště na MU – kategorizace dat (2)

KATEGORIE	POPIS	PŘÍKLADY
<b>Diskrétní data</b>	<ul style="list-style-type: none"><li>• Data určená výhradně pro vnitřní potřebu přesně definované skupiny osob (např. zaměstnanec a jeho přímý nadřízený, pracovník HR oddělení a uchazeč o zaměstnání, skupina správců IT systému s administrátorskými právy k němu).</li><li>• Vyžadují ze své povahy regulaci nebo ochranu, typicky jsou data chráněná ze zákona nebo na základě nějaké smlouvy/licence (jedná se např. o osobní údaje osob, data spadající pod obchodní tajemství apod.).</li><li>• Zpřístupnění mimo danou skupinu osob velmi pravděpodobně způsobí škodu (finanční, morální, právní apod.).</li></ul>	<ul style="list-style-type: none"><li>• Ekonomické a personální údaje osobní povahy;</li><li>• osobní údaje studentů / zaměstnanců / spolupracovníků...;</li><li>• čísla identifikačních průkazů, rodná čísla apod.;</li><li>• čísla kreditních karet;</li><li>• cenná výzkumná data (poskytující např. konkurenční výhodu) nebo data obsahující jinak citlivé informace;</li><li>• rozsáhlé kolekce interních dat;</li><li>• přístupové údaje (např. hesla či šifrovací klíče) k málo významným systémům a interním datům.</li></ul>

# Úložiště na MU – kategorizace dat (3)

KATEGORIE	POPIS	PŘÍKLADY
<b>Citlivá data</b>	<ul style="list-style-type: none"><li>• Data určená striktně jen pro vnitřní potřebu přesně definované skupiny osob (např. zdravotník a jeho pacient, řešitelé projektu s bezpečnostním prověřením určité úrovně apod.).</li><li>• Vyžadují ze své povahy zvláštní regulaci nebo obzvláštní ochranu, typicky jsou data přísně chráněná ze zákona nebo na základě smlouvy/licence (jedná se např. o velmi cenná data spadající pod obchodní tajemství, citlivé osobní údaje apod.).</li><li>• Zpřístupnění mimo danou skupinu oprávněných osob velmi pravděpodobně způsobí škodu (finanční, morální, právní apod.) velkého rozsahu se závažnými/nevratnými následky.</li><li>• <i>V praxi bude do této kategorie spadat jen málo dat, většina bude spadat nejvýše do kategorie diskrétních dat.</i></li></ul>	<ul style="list-style-type: none"><li>• Zdravotní data, citlivé osobní údaje;</li><li>• velmi cenná výzkumná data (poskytující např. unikátní a těžko opakovatelnou konkurenční výhodu) nebo výzkumná data obsahující vysoce důvěrné údaje;</li><li>• rozsáhlé kolekce diskrétních dat;</li><li>• přístupové údaje (např. hesla či šifrovací klíče) k důležitým systémům a datům kategorie diskrétní nebo citlivá.</li></ul>

# Úložiště na MU – kategorizace úložišť

TYP ÚLOŽIŠTĚ	POPIS
<b>Přenosná média</b>	Např. flash disky, paměťové karty, externí HDD/SSD, CD, DVD, ... Tj. externí paměťová média, která nejsou pevnou součástí žádného zařízení a uživatelé je používají k přenášení informací mezi zařízeními nebo pro off-line uložení dat.
<b>Lokální úložiště</b>	<p><b>V počítačích</b> Datová úložiště pevně zabudovaná ve stolních počítačích/noteboocích (typicky interní HDD/SSD apod.) v kancelářích zaměstnanců, ve studovnách apod.</p> <p><b>V mobilních zařízeních</b> Datová úložiště pevně zabudovaná v mobilních zařízeních, tj. mobilních telefonech, tabletech apod. (typicky interní nevyjímatelná paměť, v zařízení instalovaná paměťová karta apod.) v použití zaměstnanců/studentů.</p>
<b>Síťová a cloudová úložiště ÚVT</b>	<a href="#">Datová úložiště provozovaná ÚVT</a> a zpřístupněná koncovým uživatelům přes datovou síť – tzv. <a href="#">standardní</a> a <a href="#">střední</a> síťová úložiště. Do této kategorie spadají také <a href="#">datová úložiště CERIT-SC</a> pro velkoobjemová výzkumná data.
<b>Úložiště IS MUNI</b>	Dokumentový server, Úschovna a podobné úložné kapacity integrované v systému <a href="#">IS MUNI</a> .
<b>Úložiště CESNET</b>	Datová úložiště provozována <a href="#">Oddělením datových úložišť sdružení CESNET</a> . Do této kategorie spadají i služby, které tato úložiště využívají pro fyzické uložení dat, např. <a href="#">CESNET OwnCloud</a> , <a href="#">CESNET FileSender</a> apod.

# Úložiště na MU – kategorizace úložišť (2)

TYP ÚLOŽIŠTĚ	POPIS
<p style="text-align: center;"><b>Externí úložiště</b></p> <p style="text-align: center;"><i>Datová úložiště provozovaná externími subjekty, tj. mimo MUNI a CESNET.</i></p>	<p><b>Se smlouvou s MUNI</b></p> <p><b>MUNI Microsoft O365</b> Cloudová datová úložiště poskytovaná v rámci služby <a href="#">Microsoft Office 365 pro Masarykovu univerzitu</a>. Zejména se jedná o osobní úložiště OneDrive a dokumentové knihovny služby SharePoint a Skupin O365. Patří sem ale také další data uložená v MUNI O365 cloudu, jako např. elektronická pošta v MUNI O365 Outlook, soubory sdílené v sociální síti MUNI Yammer apod.</p> <p><b>MUNI Google Workspace for Education</b> Cloudová datová úložiště poskytovaná v rámci služby <a href="#">Google Workspace for Education pro Masarykovu univerzitu</a>. Zejména se jedná o datové kapacity MUNI Disk Google, patří sem ale i další data uložená v MUNI Workspace for Education cloudu, např. elektronická pošta v MUNI Google Mail, poznámky v MUNI Google Keep, kalendářová data v MUNI Google Calendar apod.</p> <hr/> <p><b>Bez smlouvy s MUNI</b></p> <p><b>Veřejná Google/Microsoft/Dropbox/... úložiště</b> Do této kategorie spadají zejména veřejné cloudové služby (zřízené typicky zdarma soukromým koncovým uživatelem jen proti elektronické registraci přes web) jako Disk Google, Microsoft OneDrive, Dropbox, Amazon úložiště, repositáře na GitHub apod. Zásadním rozdílem a „poznávacím znamením“ této kategorie cloudových úložišť oproti cloudovým službám uvedeným výše je, že MUNI nemá žádný (právní) vztah s provozovateli těchto externích služeb, a proto není schopna garantovat jakékoliv záruky ohledně bezpečnosti/důvěrnosti uložených dat nebo politiky nakládání s nimi.</p>

# Úložiště na MU – Vhodnost úložišť pro různé typy dat

TYP ÚLOŽIŠTĚ	POUŽITÍ			
	ZELENÁ: VEŘEJNÁ DATA	MODRÁ: INTERNÍ DATA	ORANŽOVÁ: DISKRÉTNÍ DATA	ČERVENÁ: CITLIVÁ DATA
PŘENOSNÁ MÉDIA (FLASH DISKY, EXTERNÍ HDD, CD, DVD, ...)	Vhodné	Možné doporučeno použít šifrování	Nevhodné možné při použití šifrování	Nevhodné
<b>LOKÁLNÍ ÚLOŽIŠTĚ</b>				
V POČÍTAČÍCH (STOLNÍ, NOTEBOOKY)	Vhodné	Vhodné	Vhodné doporučeno použití šifrování	Nevhodné možné v dobře odůvodněných případech, při provedení individuální analýzy, použití šifrování a aplikaci dalších bezpečnostních opatření, která z analýzy vyplynou
V MOBILNÍCH ZAŘÍZENÍCH (MOBILNÍ TELEFONY, TABLETY, ...)	Vhodné	Vhodné nutný zámek obrazovky (vzor, čtečka otisků prstů, PIN, heslo)	Možné nutné použití šifrování nutný silný zámek obrazovky (čtečka otisků prstů, PIN, heslo)	Nevhodné možné v dobře odůvodněných případech, při provedení individuální analýzy, použití šifrování a aplikaci dalších bezpečnostních opatření, která z analýzy vyplynou
SÍŤOVÁ A CLOUDOVÁ ÚLOŽIŠTĚ ÚVT (TZV. STANDARDNÍ A STŘEDNÍ SÍŤOVÉ ÚLOŽITĚ, VIZ <a href="#">KATALOG IT</a> , <a href="#">ÚLOŽIŠTĚ CERIT-SC</a> )	Vhodné	Vhodné	Vhodné	Vhodné, doporučeno provedení individuální analýzy, použití šifrování a aplikaci dalších bezpečnostních opatření, která z analýzy vyplynou

# Úložiště na MU – Vhodnost úložišť pro různé typy dat (2)

TYP ÚLOŽIŠTĚ	POUŽITÍ			
	ZELENÁ: VEŘEJNÁ DATA	MODRÁ: INTERNÍ DATA	ORANŽOVÁ: DISKRÉTNÍ DATA	ČERVENÁ: CITLIVÁ DATA
ÚLOŽIŠTĚ IS MUNI (NAPŘ. DOKUMENTOVÝ SERVER, ÚSCHOVNA APOD.)	Vhodné	Vhodné	Vhodné	Vhodné, doporučeno provedení individuální analýzy, použití šifrování a aplikaci dalších bezpečnostních opatření, která z analýzy vyplynou
ÚLOŽIŠTĚ CESNET (NAPŘ. CESNET ARCHIVNÍ ÚLOŽIŠTĚ, OWNCLOUD, FILESENDER, ..., VIZ <a href="#">ODDĚLENÍ DATOVÝCH ÚLOŽIŠŤ CESNET</a> )	Vhodné	Vhodné	Vhodné	Vhodné, doporučeno provedení individuální analýzy, použití šifrování a aplikaci dalších bezpečnostních opatření, která z analýzy vyplynou
EXTERNÍ ÚLOŽIŠTĚ				
SE SMLOUVOU S MUNI				
MUNI MICROSOFT O365 (MUNI O365 ONEDRIVE, SHAREPOINT, ..., VIZ <a href="#">MUNI O365</a> )	Vhodné	Vhodné	Vhodné doporučeno použití šifrování	Možné  výhradně s adekvátním procesním pokrytí dané situace na základě individuální analýzy a aplikaci bezpečnostních opatření, která z analýzy vyplynou
MUNI GOOGLE G SUITE FOR EDUCATION (VIZ <a href="#">MUNI GOOGLE APPS</a> )	Vhodné	Vhodné	Nevhodné možné při použití šifrování	Nevhodné
GRAMMARLY	Vhodné	Vhodné	Nevhodné	Nevhodné
BEZ SMLOUVY S MUNI				
VEŘEJNÁ GOOGLE, MICROSOFT, DROPBOX, ... ÚLOŽIŠTĚ	Vhodné	Nevhodné	Nevhodné	Nevhodné

# Oddělení datových úložišť CESNET

- <https://du.cesnet.cz/>
- Mohou být využívána pro **účely výzkumu**, a to **akademickými pracovníky a studenty výzkumných institucí v ČR**.
- Datová úložiště jsou **certifikována** podle normy pro **system managementu bezpečnosti informací ČSN EN ISO/IEC 27001:2014**.
- Případy použití:
  - Datový prostor pro **zálohování**.
  - Datový prostor pro **archivaci**.
  - Datový prostor pro **sdílení dat**.



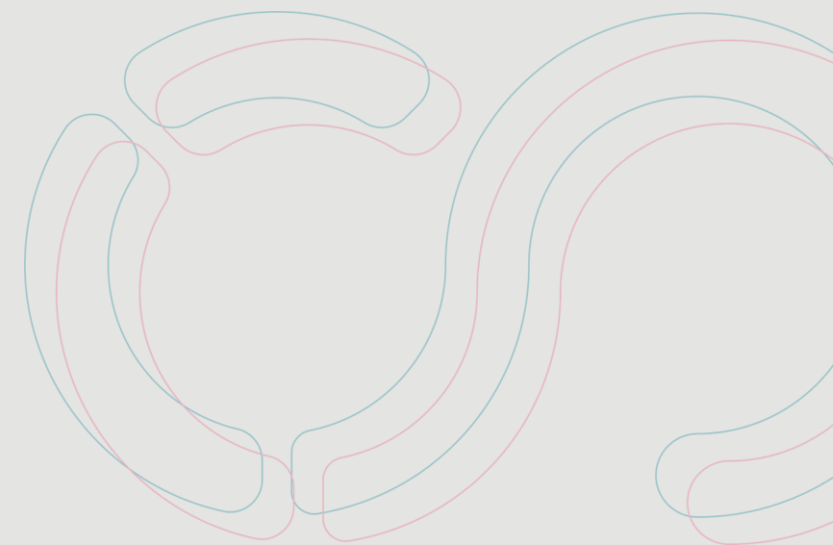
# Datová úložiště CESNET – FileSender

- Webová služba pro **jednorázové předávání** (dočasné uschování) **velkých** souborů, které poskytuje sdružení CESNET.
- Uživatel **přes webové rozhraní nahraje soubor** a připojí e-mail adresáta nebo adresátů, případně zprávu.
- Adresáti **dostanou e-mailem odkaz** ke stažení souboru.
- Pro nahrání souboru je třeba se přihlásit **přes federaci identit [eduID.cz](https://eduID.cz)**, služba je tedy dostupná **všem zaměstnancům a studentům většiny akademických institucí v ČR**.
- Pokud potřebujete dostat velký soubor od někoho, kdo **identitu** ve federaci **nemá**, je možné ve FileSenderu po přihlášení **vytvořit pozvánku**. Druhá strana komunikace pak obdrží webový odkaz, na který může nahrát soubor. Mailem pak dostanete odkaz, kde si tento soubor můžete stáhnout.
- Soubor je v systému **uložen maximálně měsíc**, poté je automaticky smazán. Lze nastavit i kratší dobu.
- Velikost jednoho souboru může být **až 500 GB**.

# Datová úložiště CESNET – OwnCloud

- Nástroj pro **synchronizaci souborů** a podporuje synchronizaci dat **na platformách Windows, Mac a Linux** a na mobilních **zařízeních s Android a iOS**.
- Obdoba Google Drive, Microsoft OneDrive apod., ale provozovaná na CESNET infrastruktuře s vyšší standardně dostupnou kapacitou pro „neplaticí“ uživatele.
- **Přihlášení přes federaci identit [eduID.cz](https://eduID.cz)**.
- Pro vytvoření účtu postačuje první přihlášení do systému.
- Standardně poskytuje **100 GB kapacity na uživatele**.
- Umožňuje sdílení dat ve skupinách, které si uživatel může nastavit
- Dovoluje soubory **sdílet webovým odkazem komukoli**, komu odkaz sdělíte
- **Data zůstávají v ČR**, sdružení CESNET nezískává žádná práva na uložený obsah, ten stále náleží vlastníkům dat.

# Repozitáře pro výzkumná data

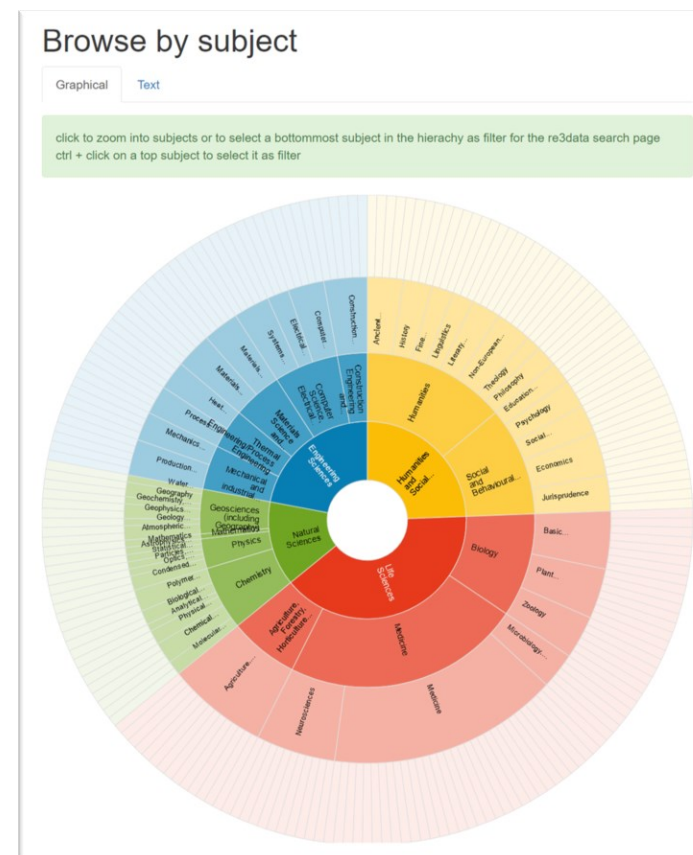


# Typy repozitářů

- **Oborové, národní, institucionální, catch-the-all, ...**
- Běžný postup práce s výzkumnými daty je **dělení po oborech, nikoliv po státech či institucích**.
  - Nezohledňujeme národní/institucionální/... specifika, ale oborová specifika.
- Pokud řešíme projektový/institucionální/národní **standard**, musíme být **kompatibilní** s **mezinárodními oborovými** zvyklostmi.
  - Oborové standardy řeší např. EOSC Task Forces: <https://eosc.eu/eosc-task-forces/>
- Pokud chceme shromažďovat **informace napříč obory**, musí to být **formát dostatečně obecný**, který možná přesáhne rámec výzkumných dat.
- Např. aplikační profil formátu DCAT ([DCAT-AP](#)), který se používá pro popis otevřených dat např. ze státní správy.

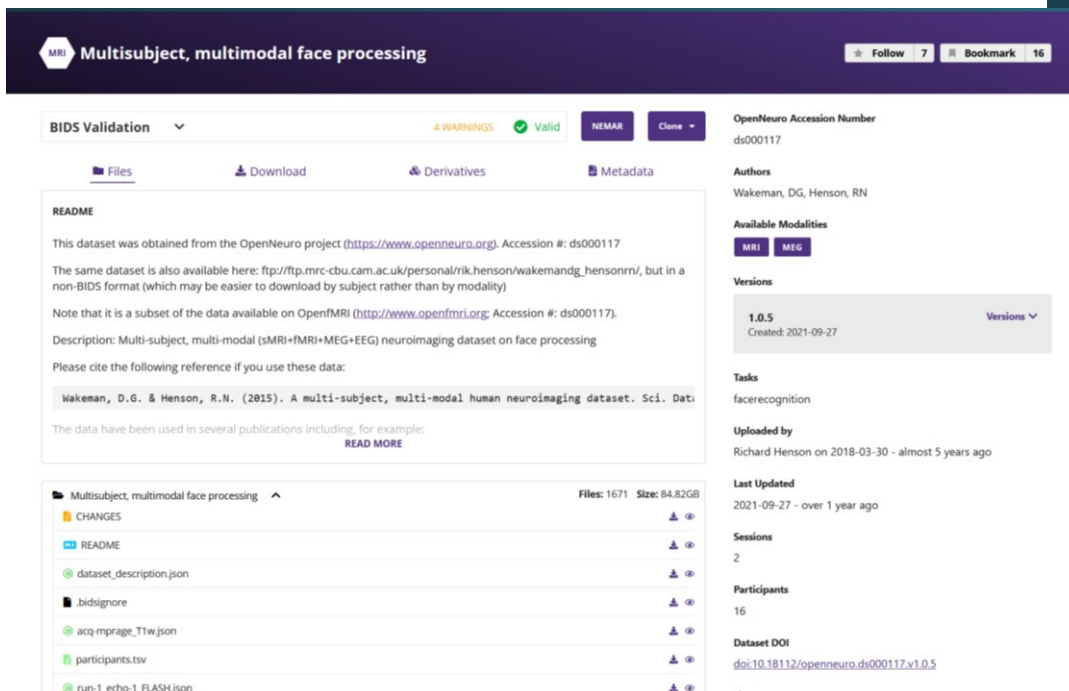
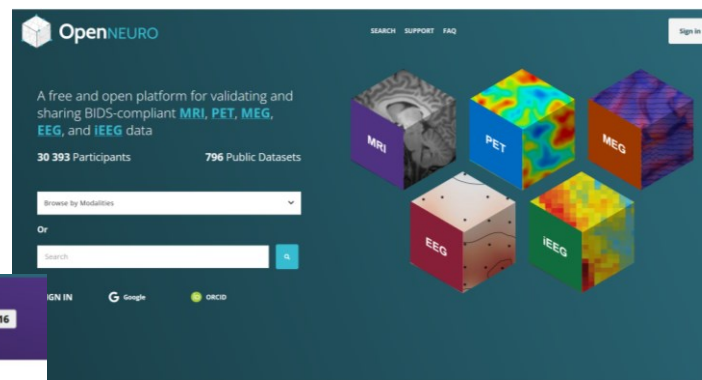
# Jak najít/vybrat datový repozitář

- OpenAIRE: [Jak najít důvěryhodný repozitář pro vaše data](#)
  - **Preferovány** jsou **důvěryhodné** certifikované **repozitáře**.
  - [CoreTrustSeal](#) (seznam certifikovaná repozitářů).
  - [Nestor Seal](#) (verifikace dle DIN 31644).
  - [ISO 16363](#).
  - Např. ale známé [Zenodo](#) žádnou certifikaci nemá...
- Mezi **nejpoužívanější obecné repozitáře** patří
  - [Zenodo](#),
  - [Figshare](#) nebo
  - [Dryad](#).
- Adresáře repozitářů:
  - Open Access repozitáře: [OpenDOAR](#)
  - **Datové repozitáře**: [re3data.org](#)
- [ELIXIR RDMkit | Data publication](#)
  - [EMBL-EBI's data submission wizard](#).
  - [ELIXIR Deposition Databases](#).
  - [Scientific Data journal's recommended repositories](#).



# Oborový datový repozitář prakticky

- OpenNeuro  
(<https://openneuro.org/>)
- Repozitář pro neurovizuální data.



- Viz *Workshop pro neurovědní uživatele MAFIL CEITEC MU* (22. 2. 2023):

- Slajdy: <https://openscience.muni.cz/openscience-na-mu/realizovane-akceprednasky/workshop-pro-neurovedni-uzivatele-cf-mafil-2023-02-22>
- Videozáznam na YouTube [Open Science kanálu MU](https://www.youtube.com/channel/UCj6ifh88D3h0): <https://youtu.be/j6ifh88D3h0>

# Oborový datový repozitář prakticky (2)

- Upload do repozitáře.

**Upload Dataset** ×

Step 1: Select Files | Step 2: Validation | Step 3: Metadata | Step 4: Accept Terms

To protect the privacy of the individuals who have been scanned, we require that all scan data be defaced before publishing a dataset.

Select a [BIDS dataset](#) to upload

**Select folder**

**close**

## Upload Dataset

×

Step 1: Select Files | Step 2: Validation | **Step 3: Metadata** | Step 4: Accept Terms

Incomplete fields in this form will make it more difficult for users to search for your dataset.

We recommend completing the applicable fields to improve your search results.

DOI of papers from the source data lab

10.1002/hbm.25698

Species

Human

Study Type

Longitudinal

Domain Studied

Brain

Number of Trials (if applicable)

1

Study Design

Block visual-motor tasks with repetitions over various acquisition parameters

Papers published from this dataset

DX status(es)

Healthy / Control

Grant Funder Name

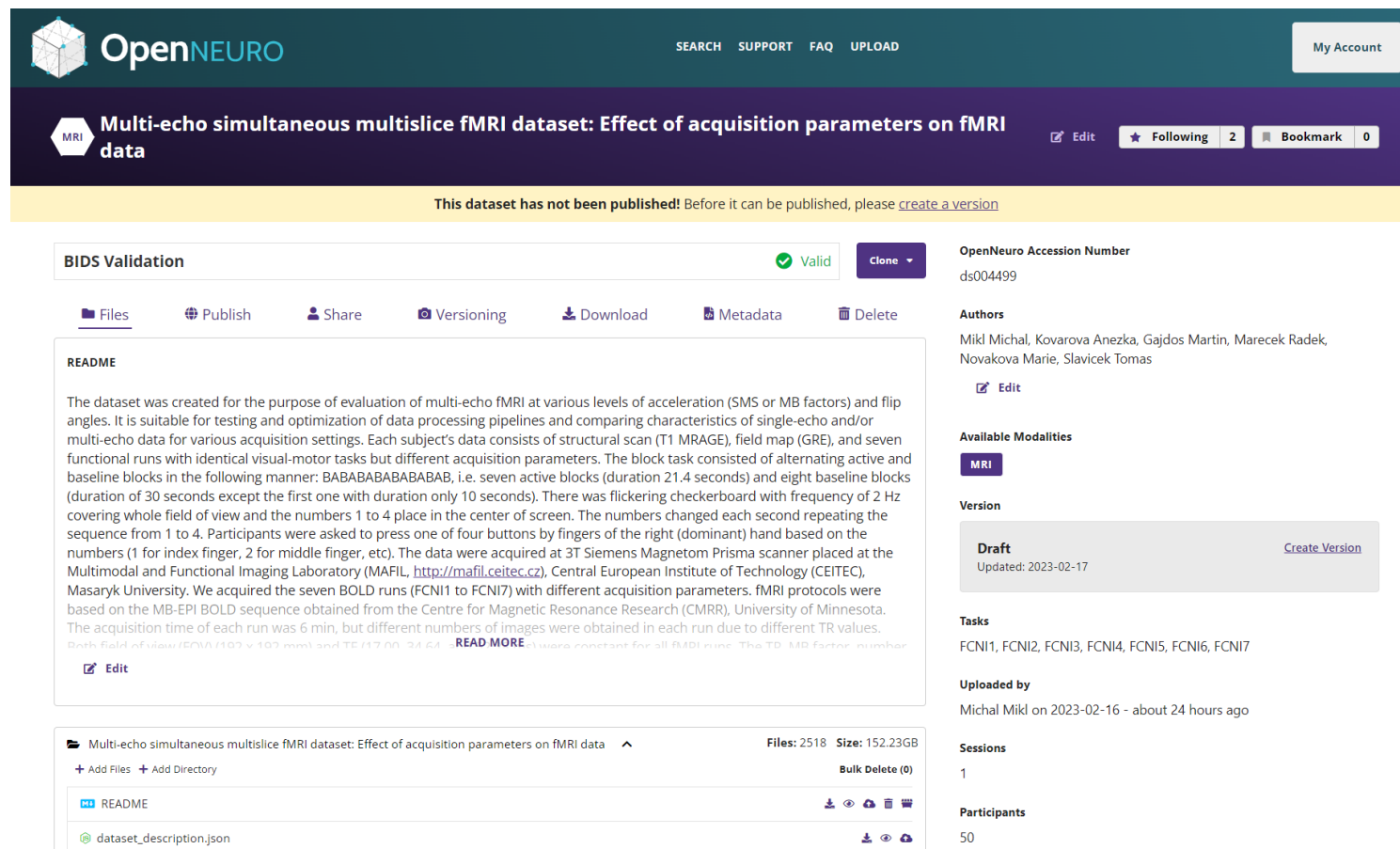
Ministry of Education, Youth, and Sports of the Czech Republic

Grant Identifier

LM2018129

# Oborový datový repozitář prakticky (3)

- Ověření záznamu DOI.



The screenshot shows the OpenNeuro dataset page for "Multi-echo simultaneous multislice fMRI dataset: Effect of acquisition parameters on fMRI data". The page is in a draft state, as indicated by the yellow banner: "This dataset has not been published! Before it can be published, please [create a version](#)".

**BIDS Validation:** Valid (with a green checkmark icon). A "Clone" button is visible.

**OpenNeuro Accession Number:** ds004499

**Authors:** Mikl Michal, Kovarova Anezka, Gajdos Martin, Marecek Radek, Novakova Marie, Slavicek Tomas. An "Edit" button is present.

**Available Modalities:** MRI

**Version:** Draft (Updated: 2023-02-17). A "Create Version" button is available.

**Tasks:** FCNI1, FCNI2, FCNI3, FCNI4, FCNI5, FCNI6, FCNI7

**Uploaded by:** Michal Mikl on 2023-02-16 - about 24 hours ago

**Sessions:** 1

**Participants:** 50

**Dataset Description (README):**  
The dataset was created for the purpose of evaluation of multi-echo fMRI at various levels of acceleration (SMS or MB factors) and flip angles. It is suitable for testing and optimization of data processing pipelines and comparing characteristics of single-echo and/or multi-echo data for various acquisition settings. Each subject's data consists of structural scan (T1 MRAGE), field map (GRE), and seven functional runs with identical visual-motor tasks but different acquisition parameters. The block task consisted of alternating active and baseline blocks in the following manner: BABABABABABAB, i.e. seven active blocks (duration 21.4 seconds) and eight baseline blocks (duration of 30 seconds except the first one with duration only 10 seconds). There was flickering checkerboard with frequency of 2 Hz covering whole field of view and the numbers 1 to 4 place in the center of screen. The numbers changed each second repeating the sequence from 1 to 4. Participants were asked to press one of four buttons by fingers of the right (dominant) hand based on the numbers (1 for index finger, 2 for middle finger, etc). The data were acquired at 3T Siemens Magnetom Prisma scanner placed at the Multimodal and Functional Imaging Laboratory (MAFIL, <http://mafili.ceitec.cz>), Central European Institute of Technology (CEITEC), Masaryk University. We acquired the seven BOLD runs (FCNI1 to FCNI7) with different acquisition parameters. fMRI protocols were based on the MB-EPI BOLD sequence obtained from the Centre for Magnetic Resonance Research (CMRR), University of Minnesota. The acquisition time of each run was 6 min, but different numbers of images were obtained in each run due to different TR values. Both field of view (FOV) (192 x 192 mm) and TE (17.00, 24.64, ...) were constant for all BOLD runs. The TR, MB factor, number... [READ MORE](#)

**Files:** 2518, Size: 152.23GB

**Files List:**  
+ Add Files + Add Directory  
Bulk Delete (0)  
- README  
dataset\_description.json



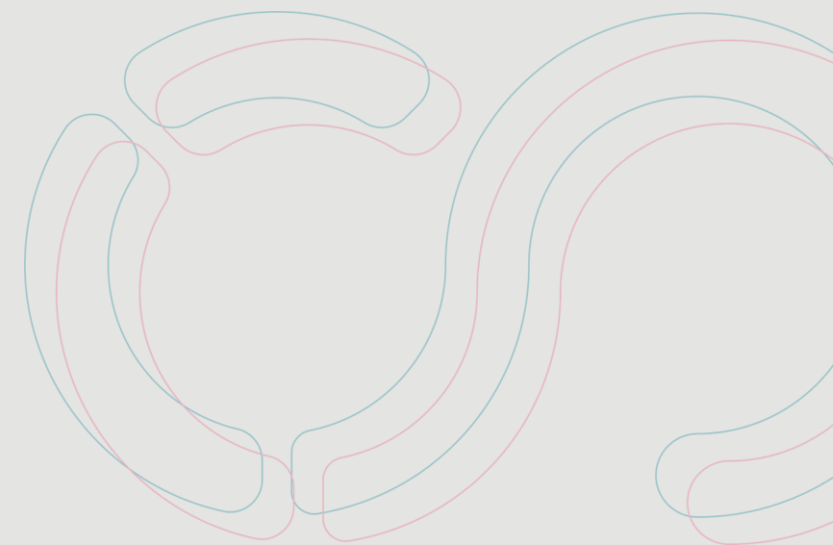
# Datový repozitář CESNET

- Pilotní provoz:  
<https://data.narodni-repozitar.cz/>
- Návody:  
[https://du.cesnet.cz/cs/navody/narodni\\_repozitar/start](https://du.cesnet.cz/cs/navody/narodni_repozitar/start)
- Pro ukládání a řízené zveřejňování datových sad.
  - Vytvoření metadatového záznamu, který popisuje datovou sadu,
  - připojení jednotlivých souborů, které tvoří vlastní datovou sadu.
- Každý veřejný záznam může být **na požádání opatřen** persistentním identifikátorem DOI.
- **Možnost vazby na publikace** založená na dané datové sadě přes DOI publikace v metadatech.
- Pro nahrávání dat nutná registrace.
  - Je schvalovaná, není okamžitý přístup.
- Možnost **odložené zveřejnění** dat.
  - Časové embargo.
  - Metadata jsou viditelná okamžitě pro dohledatelnost.

# Na co dát pozor při výběru repozitáře

- Přidělí repozitář vašim datům **perzistentní identifikátor** (např. DOI)?  
Díky perzistentním identifikátorům jsou vaše data snáze vyhledatelná a citovatelná.
  - Je repozitář **důvěryhodný/certifikovaný**?  
U certifikovaných repozitářů máte větší jistotu, že je o vaše data dobře postaráno.
  - Poskytuje vybraný repozitář **otevřený přístup** k uloženým datům?  
Pokud chcete svá data sdílet otevřeně, pak je toto klíčová informace.
  - Opatří repozitář vaše data **licencí**? Uvádí jasné podmínky, za jakých lze data uložená v repozitáři využívat?
- Je důležité, aby uživatelé vašich dat věděli, jak s nimi mohou nakládat.
- Poskytne repozitář pro vaše data **úvodní stránku s metadaty** (tzv. landing page)?  
Metadata pomohou ostatním data najít, zjistit, co je to za data, a také jak je citovat.
  - Umožňuje repozitář **verzování**?  
Pokud svůj dataset aktualizujete, můžete aktualizovanou verzi nahrát jako novou verzi k původnímu datasetu. Nový dataset dostane přidělený svůj vlastní perzistentní identifikátor a uživatelé snadno zjistí, jaká je nejnovější verze nebo jaká verze byla použita v původní studii.

# PID – perzistentní identifikátory



# Perzistentní identifikátory (PID)

- Mají zajistit **oddělení identifikace objektu** jako takového,
  - osoba,
  - instituce,
  - publikace,
  - dataset,
- **od** jeho momentálního **fyzického umístění**.
- **Příklad** – datová sada *https-set*
  - **Identifikátor** datové sady: <https://doi.org/10.48791/4mxx-r725>
  - Současné fyzické **umístění**:  
<https://ucnmuni.sharepoint.com/teams/mu-UVT-https-set/Shared%20Documents/Forms/AllItems.aspx?id=%2Fteams%2Fmu%2DUVT%2Dhttps%2Dset%2FShared%20Documents%2Fhttps%2Dset%2Dv1%2E0%2E0&p=true&ga=1>
  - Fyzické umístění se **bude pravděpodobně** v budoucnu **měnit** – zvažován přesun do [pilotně provozovaného datového repozitáře CESNET](#).
  - Změny nevadí – **uživatelům** je jako odkaz na data **vždy** prezentováno [DOI 10.48791/4mxx-r725](https://doi.org/10.48791/4mxx-r725), které je vždy **zavede** na **aktuální umístění**.

# Perzistentní identifikátory (PID) (2)

- Mají zajistit **jednoznačnost**.
- **Příklad** – jména fyzických osob
  - **Více forem zápisu** jména **jedné fyzické osoby**.
  - **Více různých fyzických osob** se **stejným jménem**.
- Mají zajistit **trvalost** (perzistenci).
  - **Metadata** fyzicky umístěna **u třetí strany** nezávisle na fyzickém umístění odkazované entity.
  - Vlastník identifikátoru se stará o **aktualizaci metadat** a aktualizace **směrování** na aktuální umístění.
  - Třetí strana **pečuje o zachování** poslední verze a historie **metadat** a **existenci identifikátoru** jako takového, i pokud se vlastník identifikátoru o něj starat přestane. A dokonce i tehdy, pokud identifikovaná entita jako taková nebude zachována.

<input type="checkbox"/>	<a href="#">NovakD (1)</a>	Novák, David (1)	Join	Delete
<input type="checkbox"/>	<a href="#">NovakE (5)</a>	Nováková, Eva (5) Nováková, E. (0)	Join	Delete
<input type="checkbox"/>	<a href="#">NovakJ (52)</a>	Novák, Josef (38) Novák, Jos. (1) Novák, J. (13)	Join	Delete
<input type="checkbox"/>	<a href="#">NovakJ2 (7)</a>	Novák, Jiří (7) Novak, Jiri (0)	Join	Delete
<input type="checkbox"/>	<a href="#">NovakJ7 (19)</a>	Novák, Josef (16) Novák, J. (3)	Join	Delete
<input type="checkbox"/>	<a href="#">NovakK (2)</a>	Novák, Karel (2)	Join	Delete
<input type="checkbox"/>	<a href="#">NovakM (2)</a>	Novák, Mirko (2) Novak, Miroslav M. (0) Novák, M. M. (0) Novák, M. (0)	Join	Delete
<input type="checkbox"/>	<a href="#">NovakM2 (2)</a>	Nováková, Markéta (2)	Join	Delete
<input type="checkbox"/>	<a href="#">NovakM3 (1)</a>	Novák, Miroslav (1)	Join	Delete
<input type="checkbox"/>	<a href="#">NovakO (2)</a>	Novák, Ondřej (2)	Join	Delete
<input type="checkbox"/>	<a href="#">NovakP (1)</a>	Novák, Petr (1)	Join	Delete
<input type="checkbox"/>	<a href="#">NovakS (1)</a>	Novák, Stanislav (1)	Join	Delete
<input type="checkbox"/>	<a href="#">NovakV (57)</a>	Novák, Vítězslav (55) Novák, V. (2)	Join	Delete
<input type="checkbox"/>	<a href="#">NovakV2 (7)</a>	Novák, Vilém (7) Novák, V. (0)	Join	Delete
<input type="checkbox"/>	<a href="#">NovakV3 (53)</a>	Novák, Vladimír (53)	Join	Delete
<input type="checkbox"/>	<a href="#">NovakZ (1)</a>	Novák, Zdeněk (1)	Join	Delete

CODEVALUE	PREFLABEL_EN	HREF	Maturity	Globally resolvable	For which object type	Comments
ADSBibcode	Astrophysics Data System - Bibliographic Reference Code	<a href="https://ui.adsabs.harvard.edu/">https://ui.adsabs.harvard.edu/</a>	High	Needs token	Publication	
ARK	Archival Resource Key	<a href="https://arks.org/">https://arks.org/</a>	High	Local	Other (in comments)	Everything
arXiv	arXiv identifier scheme	<a href="https://arxiv.org/">https://arxiv.org/</a>	High	Yes	Publication	
ASIN	Amazon Standard Identification Number	<a href="https://sellercentral.amazon.ca/gp/help/external/200317470?language=en-CA&amp;ref=mpbc_200576730_cont_200317470">https://sellercentral.amazon.ca/gp/help/external/200317470?language=en-CA&amp;ref=mpbc_200576730_cont_200317470</a> <a href="https://indico.cern.ch/event/780651/attachments/1776614/2888642/Conference_PIDs_and_Crossmark.pdf">https://indico.cern.ch/event/780651/attachments/1776614/2888642/Conference_PIDs_and_Crossmark.pdf</a>	High		Other (in comments)	Things sell by Amazon
ConfID	Conference identifier			?	Event	Not clear if it is a Crossref service
Crossref DOI			High	Yes	Publication	
Crossref_funders	Crossref Funder Registry	<a href="https://www.crossref.org/services/content-registration/grants/">https://www.crossref.org/services/content-registration/grants/</a>	?	Yes	Organisation	
Crossref_grants	Registering research grants	<a href="https://www.crossref.org/community/grants/">https://www.crossref.org/community/grants/</a>			Other (in comments)	Grants
DataCite DOI						
DOI	Digital Object Identifier	<a href="https://www.doi.org/">https://www.doi.org/</a>	High	Yes	Other (in comments)	28 different resources and outputs
EAN13	The 13-digit International Article Number	<a href="https://www.gs1.org/standards/barcodes/ean-upc">https://www.gs1.org/standards/barcodes/ean-upc</a>	High		Publication	Services supporting PIDs and metadata for 40+ resource and output types
eiSNB	electronic International Standard Book Number	<a href="https://www.isbn-international.org/">https://www.isbn-international.org/</a>	High		Other (in comments)	Physical product identifier. A Whole famili of id: UPC-A, UPC-E, EAN13, EAN8
eISSN	Electronic International Standard Serial Number	<a href="http://portal.issn.org/">http://portal.issn.org/</a>	High		Publication	
GRID	Global Research Identifier Database	<a href="https://www.grid.ac/">https://www.grid.ac/</a>	Closed	?	Organisation	Identifies various types of serial publications (eg. journals, websites, blogs)
Handle	Handle	<a href="http://www.handle.net/">http://www.handle.net/</a>	High	Yes	Dataset	Transiton to ROR It is the base of DOI also
igsn	International Geo Sample Number	<a href="https://www.igsn.org/">https://www.igsn.org/</a>	High	Yes	Other (in comments)	Physical Samples and Sampling Features
ISAN	International Standard Audiovisual Number	<a href="https://www.isan.org/">https://www.isan.org/</a>	High	Yes?	Publication	
ISBN	International Standard Book Number	<a href="https://www.isbn-international.org/">https://www.isbn-international.org/</a>	High		Publication	
ISLI	Identifies the links between different entities	<a href="https://www.isbn-international.org/content/isli-introduction">https://www.isbn-international.org/content/isli-introduction</a>	?	Yes	Other (in comments)	Link bw. entities
ISMN	International Standard Music Number	<a href="https://www.ismn-international.org/">https://www.ismn-international.org/</a>		No	Publication	
ISNI	International Standard Name Identifier	<a href="https://isni.org/page/search-database/">https://isni.org/page/search-database/</a>			Person	Contributors to creative works and their distribution and organizations
ISSN	International Standard Serial Number	<a href="http://portal.issn.org/">http://portal.issn.org/</a>	High		Publication	Identifies various types of serial publications (eg. journals, websites, blogs)
ISTC	The International Standard Text Code	<a href="http://www.istc-international.org/">http://www.istc-international.org/</a>	Defunct	No	Publication	Ceased in 2017
LSID	Life Sciences Identifier	<a href="http://www.lsid.info/">http://www.lsid.info/</a>	?	Yes	Other (in comments)	Metadata for life science items
ORCID	Open Researcher and Contributor ID	<a href="https://orcid.org/">https://orcid.org/</a>	High	Yes	Person	
PIC	(EC) partner identity code	<a href="https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/support/faq/1055">https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/support/faq/1055</a>				
PMID	PubMed ID	<a href="https://www.ncbi.nlm.nih.gov/pmc/pmctopmid/">https://www.ncbi.nlm.nih.gov/pmc/pmctopmid/</a>			Publication	
PURL	persistent uniform resource locator	<a href="https://archive.org/services/purl/">https://archive.org/services/purl/</a>	High	Yes	Other (in comments)	Resources on the Web
QID	Wikidata identifier	<a href="https://www.wikidata.org/wiki/Wikidata:Identifiers">https://www.wikidata.org/wiki/Wikidata:Identifiers</a>			Other (in comments)	Knowledge item
RAID	Persistent Identifier for research projects	<a href="https://www.raid.org.au/">https://www.raid.org.au/</a>			Needs token	Research projects
Ringgold	Unique numerical identifier applied to organizations in the scholarly supply chain	<a href="https://www.ringgold.com/">https://www.ringgold.com/</a>			Organisation	
ROR	Research Organization Registry	<a href="https://ror.org/">https://ror.org/</a>	High	Yes	Organisation	
RRID	Research Resource Identifier	<a href="https://scicrunch.org/resources">https://scicrunch.org/resources</a>				
ScopusAuthorID	Scopus Author ID	<a href="https://service.elsevier.com/app/answers/detail/a_id/11212/supporthub/scopus/">https://service.elsevier.com/app/answers/detail/a_id/11212/supporthub/scopus/</a>	?	No?	Publication	
SWHID	Software Heritage persistent Identifiers	<a href="https://docs.softwareheritage.org/develop/swh-model/persistent-identifiers.html">https://docs.softwareheritage.org/develop/swh-model/persistent-identifiers.html</a>		Local	Source Code	
UPC	Universal Product Code	<a href="https://www.gs1.org/standards/barcodes/ean-upc">https://www.gs1.org/standards/barcodes/ean-upc</a>			Other (in comments)	Synonym of EAN13? Product identifier
URI	Uniform Resource Identifier					
URL	Uniform Resource Locator					
URN	Uniform Resource Name					
VAT-number	VAT number	<a href="http://ec.europa.eu/taxation_customs/vies/vatRequest.html">http://ec.europa.eu/taxation_customs/vies/vatRequest.html</a>			Organisation	

Zdroj: Interní pracovní materiály [EOSC Task Force PID Policy and Implementation](#)

# Populární PID

- Osoby

- **ORCID:** <https://orcid.org/>
  - Příklad: [0000-0001-6399-5453](https://orcid.org/0000-0001-6399-5453)

- Instituce

- **ROR:** <https://ror.org/>
  - Příklad: [02j46qs45](https://ror.org/02j46qs45)

- Publikace

- **DOI:** <https://www.crossref.org/>
  - Příklad: [10.5817/CP2022-3-1](https://doi.org/10.5817/CP2022-3-1)

- Datasetsy

- **DOI:** <https://datacite.org/>
  - Příklad: [10.48791/4mxx-r725](https://doi.org/10.48791/4mxx-r725)
- **Handle:** <https://handle.net/>
  - Příklad: [11222.digilib/130328](https://handle.net/11222.digilib/130328)

- Knihy

- **ISBN:** <https://www.isbn-international.org/>
  - Příklad: **978-3-16-148410-0**

- Časopisy

- **ISSN:** <http://portal.issn.org/>
  - Příklad: **0378-5955**

- Obchodní produkty

- **EAN13:** <https://www.gs1.org/standards/barcodes/ean-upc>
  - Příklad: **5901234123457**



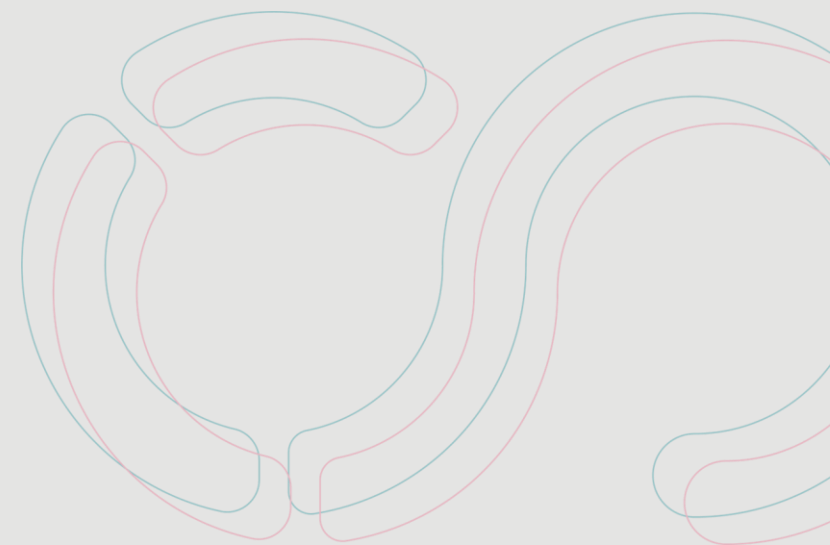
Zdroj: VaGla, CC BY-SA 3.0 - <http://creativecommons.org/licenses/by-sa/3.0/>, via Wikimedia Commons

- Obyvatelé ČR

- **Rodné číslo:** <https://www.zakonyprolidi.cz/cs/2004-302/>
  - Příklad: **736028/5163**

# DataCite

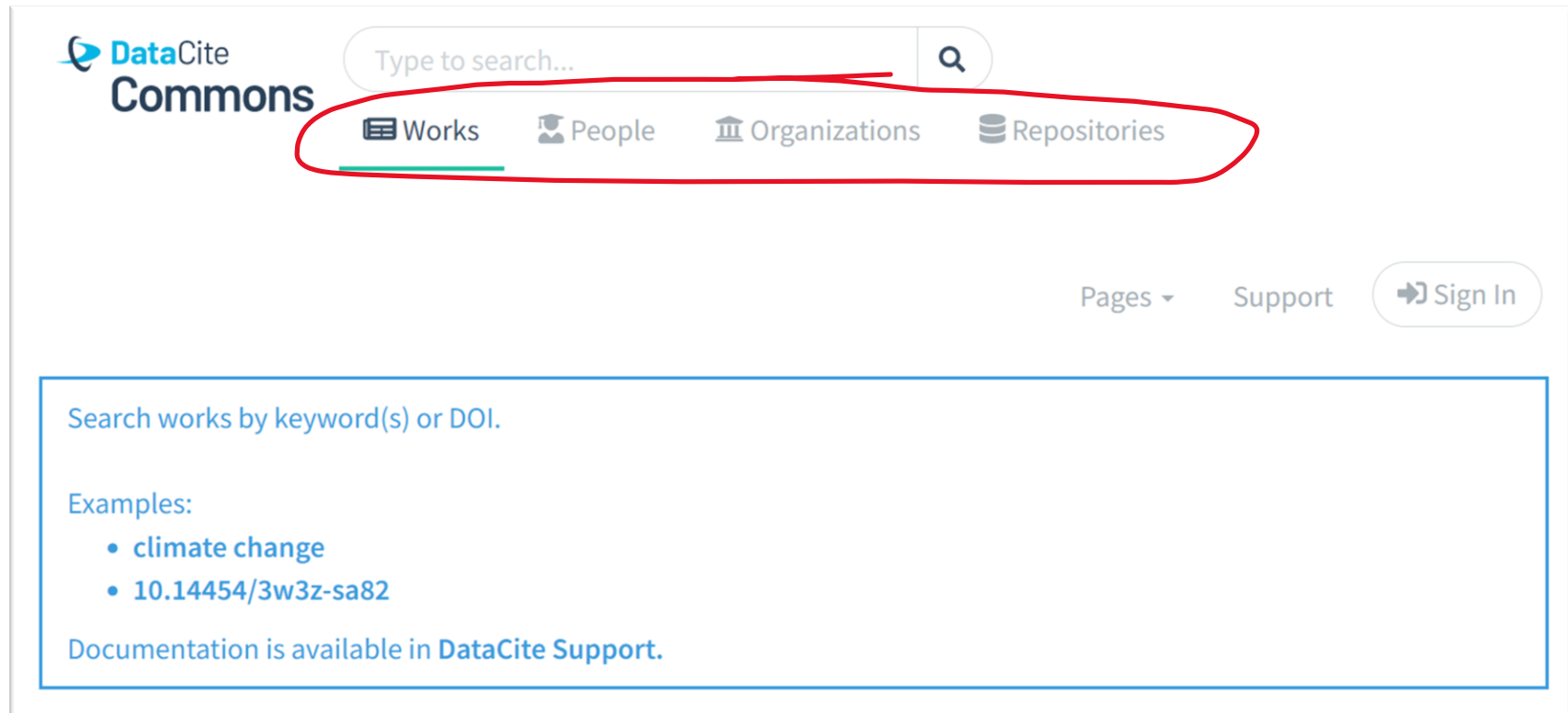
## DOI pro data





# Vyhledávání metadat z databází PID metadat

- Metadata evidovaná k PID bývají veřejně k dispozici.
- <https://commons.datacite.org/>



# Vyhledávání metadat z databází PID metadat (2)

- Přístup k metadatům – příklad *https-set*
- <https://commons.datacite.org/doi.org?query=10.48791%2F4mxp-r725>


**1 Work**

---


**Encrypted Web Traffic: Event Logs and Packet Traces**  
Stanislav Špaček, Petr Velan, Pavel Čeleda & Daniel Tovarňák  
Version 1.0.0 of HTTPS Traffic Dump With Corresponding Event Log From The Webserver published 2022 in [CSIRT-MU](#)

The dataset and supplementary materials for the following paper submitted to the Data in Brief journal: Stanislav Špaček, Petr Velan, Pavel Čeleda and Daniel Tovarňák. Encrypted Web Traffic: Event Logs and Packet Traces.

DOI registered March 11, 2022 via DataCite.



[Dataset](#) [Computer and information sciences](#) [English](#)

 <https://doi.org/10.48791/4mxp-r725>

# Vyhledávání metadat z databází PID metadat (3)

The screenshot shows the DataCite Commons interface for a dataset. At the top, there is a search bar and navigation tabs for Works, People, Organizations, and Repositories. The dataset title is "Encrypted Web Traffic: Event Logs and Packet Traces" with a DOI of <https://doi.org/10.48791/4mxxp-r725>. Below the title, there are two buttons: "Add to ORCID Record" and "Download Metadata", with the latter being circled in red. The "Cite as" section provides the citation: Špaček, S., Velan, P., Čeleda, P., & Tovarňák, D. (2022). *Encrypted Web Traffic: Event Logs and Packet Traces* (Version 1.0.0) [Data set]. Masaryk University. <https://doi.org/10.48791/4MXP-R725>. The "Share" section includes icons for Email, Twitter, and Facebook. The "Description" section is currently selected, showing the dataset's purpose and authors. At the bottom, there are tags for "Dataset", "Computer and information sciences", and "English", along with the DOI link.

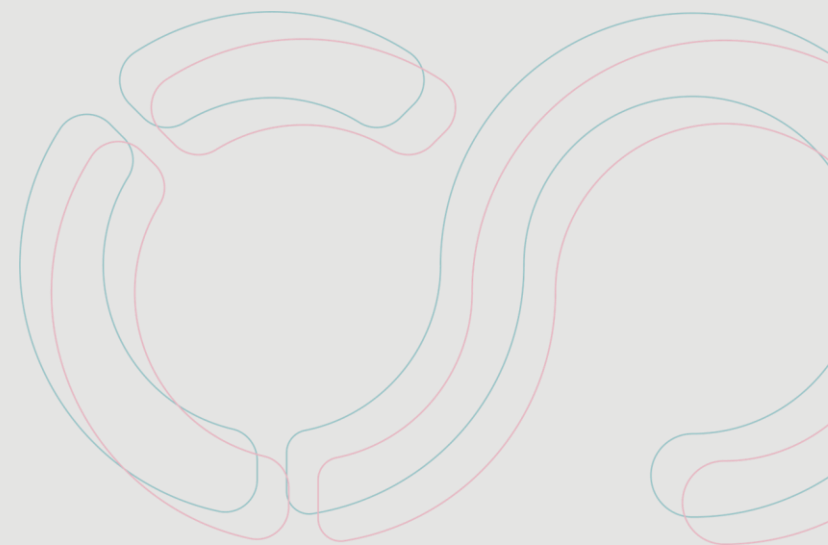
# Vyhledávání metadat z databází PID metadat (4)

```

<?xml version="1.0" encoding="UTF-8"?>
- <resource xsi:schemaLocation="http://datacite.org/schema/kernel-4 http://schema.datacite.org/meta/kernel-4/metadata.xsd" xmlns="http://datacite.org/schema/kernel-4"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <identifier identifierType="DOI">10.48791/4MXP-R725</identifier>
  - <creators>
    - <creator>
      <creatorName nameType="Personal">Špaček, Stanislav</creatorName>
      <givenName>Stanislav</givenName>
      <familyName>Špaček</familyName>
      <nameIdentifier schemeURI="" nameIdentifierScheme="Other">https://orcid.org/0000-0002-7187-5045</nameIdentifier>
    </creator>
    - <creator>
      <creatorName nameType="Personal">Velan, Petr</creatorName>
      <givenName>Petr</givenName>
      <familyName>Velan</familyName>
      <nameIdentifier schemeURI="https://orcid.org" nameIdentifierScheme="ORCID">https://orcid.org/0000-0002-2824-4299</nameIdentifier>
    </creator>
    - <creator>
      <creatorName nameType="Personal">Čeleda, Pavel</creatorName>
      <givenName>Pavel</givenName>
      <familyName>Čeleda</familyName>
      <nameIdentifier schemeURI="https://orcid.org" nameIdentifierScheme="ORCID">https://orcid.org/0000-0002-3338-2856</nameIdentifier>
    </creator>
    - <creator>
      <creatorName nameType="Personal">Tovarník, Daniel</creatorName>
      <givenName>Daniel</givenName>
      <familyName>Tovarník</familyName>
      <nameIdentifier schemeURI="https://orcid.org" nameIdentifierScheme="ORCID">https://orcid.org/0000-0002-7206-5167</nameIdentifier>
    </creator>
  </creators>
  - <titles>
    <title xml:lang="en">Encrypted Web Traffic: Event Logs and Packet Traces</title>
  </titles>
  <publisher>Masaryk University</publisher>
  <publicationYear>2022</publicationYear>
  <resourceType resourceTypeGeneral="Dataset">HTTPS Traffic Dump with Corresponding Event Log from the Webserver</resourceType>
  - <subjects>
    <subject schemeURI="http://www.oecd.org/science/inno" valueURI="http://www.oecd.org/science/inno/38235147.pdf" subjectScheme="Fields of Science and Technology (FOS)">FOS: Computer and information sciences</subject>
  </subjects>
  - <dates>
    <date dateType="Issued">2022</date>
  </dates>
  <language>en</language>
  - <sizes>
    <size>270 GiB</size>
  </sizes>
  - <formats>
    <format>Network traces in Libpcap file format (see https://wiki.wireshark.org/Development/LibpcapFileFormat) and event log from IIS web servers in a newline-delimited JSON (JSON Lines) file.</format>
  </formats>
  <version>1.0.0</version>
  - <rightsList>
    <rights rightsURI="https://creativecommons.org/licenses/by/4.0/legalcode">Creative Commons Attribution 4.0 International</rights>
  </rightsList>
  - <descriptions>
    <description xml:lang="en" descriptionType="Abstract">The dataset and supplementary materials for the following paper submitted to the Data in Brief journal: Stanislav Špaček, Petr Velan, Pavel Čeleda and Daniel Tovarník. Encrypted Web Traffic: Event Logs and Packet Traces.</description>
    <description xml:lang="en" descriptionType="TechnicalInfo">Repository Structure - Anonymization – the anonymization folder contains the scripts and settings that were used to anonymize the data capture. - Dataset – the dataset folder contains the host-based and network parts of the dataset in two separate files. - Pcap2flow – the pcap2flow folder contains the tools, settings, and guide to convert the packet dump in the dataset into aggregated IP flows. - README.md This work is licensed under a [Creative Commons Attribution 4.0 International License: https://creativecommons.org/licenses/by/4.0/</description>
  </descriptions>
  - <fundingReferences>
    - <fundingReference>
      <funderName>Horizon 2020 Framework Programme</funderName>
      <funderIdentifier funderIdentifierType="Crossref Funder ID">https://doi.org/10.13039/100010661</funderIdentifier>
      <awardNumber awardURI="https://www.concordia-h2020.eu/">830927</awardNumber>
      <awardTitle>CONCORDIA: Cyber security cOmpeteNce fOr Research anD InnovAtion</awardTitle>
    </fundingReference>
  </fundingReferences>
</resource>

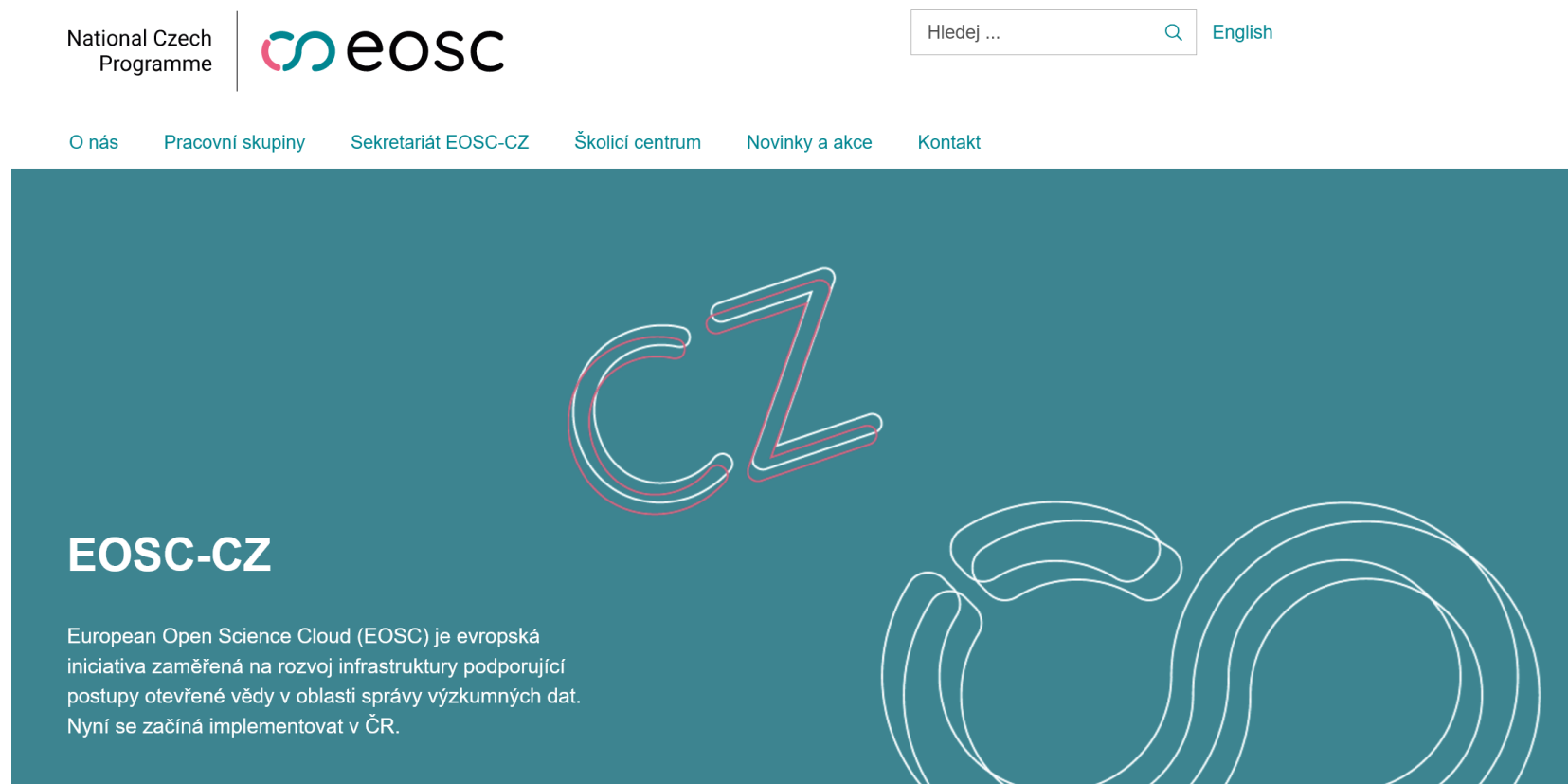
```

# EOSC



# Implementace European Open Science Cloud (EOSC) v ČR

<https://www.eosc.cz/>

A screenshot of the EOSC-CZ website. The header includes the 'National Czech Programme' and 'EOSC' logos on the left, a search bar with the text 'Hledej ...' and a magnifying glass icon, and a language selector set to 'English'. Below the header is a navigation menu with links for 'O nás', 'Pracovní skupiny', 'Sekretariát EOSC-CZ', 'Školící centrum', 'Novinky a akce', and 'Kontakt'. The main content area has a teal background. On the right side, there is a large graphic of the letters 'CZ' in a stylized, outlined font. Below this, the text 'EOSC-CZ' is displayed in a bold, white font. Underneath, a paragraph in white text reads: 'European Open Science Cloud (EOSC) je evropská iniciativa zaměřená na rozvoj infrastruktury podporující postupy otevřené vědy v oblasti správy výzkumných dat. Nyní se začíná implementovat v ČR.' To the right of this text is a decorative graphic of several overlapping, white-outlined circles of varying sizes.

# EOSC v ČR – očekávané služby Národní datové infrastruktury (NDI)

- **Služby** očekávané od **NDI** vychází z výstupů pracovní skupiny [Základní služby](#).
  - **Pracovní skupiny** pro implementaci EOSC v ČR **jsou stále otevřené** novým členům, viz [Pracovní skupiny implementace EOSC v ČR](#).
- Navrhované služby se dají rozdělit do několika základních kategorií:
  - **Důvěryhodné repozitáře**
  - Služby pro podporu práce s **persistentními identifikátory** (PID)
  - Služby pro řízení **přístupu** k datům, **autentizaci a autorizaci** uživatelů (AAI) a podporu jejich **spolupráce**
  - Služby pro **podporu** a plánování **správy výzkumných dat**
  - **Machine actionability**
  - Služby pro **FAIRifikaci** výzkumných **dat**
  - Služby **transferu dat**
  - **Kvalita** služeb
  - **Kompatibilita** s **existujícími** platformami
  - **Statistiky a přehledy**
  - Národní příspěvek k **budování EOSC**

# Důvěryhodné repozitáře

- **Národní repozitářová platforma (NRP)** nad NDI.
  - Součástí i **Národní metadatový adresář (NMA)**.
- Místo, kam můžete (i velká) data **nejen uložit**, ale také
  - spolehnout se, že tam budou **v následujících letech dostupná** a
  - budete moci **s daty pracovat**.
- Určeno i pro „živá“ data, tj. podpora **aktivní práce s daty**, ne jen uložení archivních dat.
- Poskytování služeb dle vhodných standardů/certifikací pro provoz **důvěryhodných repozitářů**.
  - Důvěryhodné repozitáře dle definic EU, EOSC, grantových agentur apod.
  - Pravděpodobně CoreTrustSeal.
  - Rodina standardů ISO 27 tisíc apod.?
- **Nestaráte se o vlastní infrastrukturu, zabezpečená úložiště apod., ale**
  - **využíváte služby**.
  - NDI umožní **postavení vlastních instancí repozitářů** (oborových, institucionálních, ...).



# Precizní kontrola vlastníka dat nad řízením přístupu k datům i metadatům

- **Data** uložená v NDI jsou **pod** plnou **kontrolou** jejich **vlastníků**.
  - **Zpřístupnění** dalším osobám se děje **na pokyn** jejich **vlastníka**.
- Už samotná **informace o existenci** dat či metadata **může být citlivá** informace.
  - **Řízení přístupu** až na úrovni řízení **viditelnosti existence** samotných **metadat** či jejich **částí** apod.
- **Federované AAI** služby
  - Autentizace uživatelů přes **federované** služby **přihlašování** (eduID.cz apod.),
  - tj. s využitím **existující identity** z instituce, **bez** nutnosti **vytváření** dalšího **účtu**, hesla atd.

# Služby pro sdílení dat s jinými uživateli

- **Sdílení dat** by mělo být možné jednak **s konkrétními osobami a skupinami**
  - výběr např. z institucionálních identit daných federovaným přihlášením apod.,
  - případně **sdílením „odkazem“**.
- Může být řešeno i formou **„data visiting“**.
  - **Řízené** povolením přístupu vybraným uživatelům **do „domovského“ úložiště** s danými daty,
  - **bez** nutnosti **přenosu do jiné lokality**.
- Možnost **anonymních view-only přístupů** k datům.
- Pro podporu předání dat pro **zaslepené recenzní řízení** apod.
- Stále častější **povinnost sdílení podkladových dat k publikacím**.
  - Někdy již v okamžiku recenzního řízení.
  - Ukázka existující implementace např. v platformě OSF: <https://help.osf.io/hc/en-us/articles/360042097853-Create-a-View-only-Link-for-a-Registration>

# Služby pro podporu a plánování správy výzkumných dat

- **Definované** (strojově zpracovatelné) **politiky** práce s daty **v repozitářích** v NDI.
- **Podpora tvorby DMP.**
  - **Vygenerování DMP**
    - nebo vybraných částí
  - **na základě politik** použitého **repozitáře**, znalosti licencí, formátů, velikostí dat, cílených retencí dat (maximální či minimální doba uchovávání např. na základě zákonných požadavků) apod.
  - Možnost **napojení/integrace existujících nástrojů** pro DMP.
    - Např. Data Stewardship Wizard.

# Metriky a měření plnění FAIR atributů pro výzkumná data

- Pro datasey může být zajímavé nějakým způsobem **měřit** plnění jednotlivých **FAIR atributů**.
  - NRP EOSC CZ by měla **definovat metriky** hodnocení a **poskytovat nástroje** pro **měření plnění FAIR atributů**.
- Hodnocení na základě informací o
  - datech,
  - metadatech,
  - nastavení a politikách konkrétního repozitáře apod.
- Příkladem implementace podobného nástroje může být např. FAIR-Checker (<https://fair-checker.france-bioinformatique.fr/>) nebo FAIR metriky pro DMP v Data Stewardship Wizard.



Zdroj: <https://researchers.ds-wizard.org/>

# Otázky?



Zdroj: Communicate\_communication\_conference\_2028004 od OpenClipart-Vectors z Pixabay



Spolufinancováno  
Evropskou unií

# Děkuji za pozornost!

E: ruzicka@ics.muni.cz

E: info@esoc.cz W: <https://www.eosc.cz/>



Spolufinancováno  
Evropskou unií

