

# OTVORENÁ VEDA V PRAXI

Otvorené vedecké dáta

## PRINCÍPY FAIR A VEDECKÉ DÁTA

SILVIA SOFIANOS A GABRIELA FIŠOVÁ

CENTRUM VEDECKO-TECHNICKÝCH INFORMÁCIÍ SR

17.10. – 19. 10. 2023

# Princípy FAIR a vedecké dáta

## OBSAH

- Ako na FAIR dáta
- GDPR, citlivé dáta
- **Manažment vedeckých dát, plán manažmentu dát (DMP)**
- **Právne aspekty použitia vedeckých dát + licencie**
- Zhrnutie
- Cvičenie
- Zdroje a samoštúdium



## **MANAŽMENT VEDECKÝCH DÁT**

**PREČO MANAŽOVAŤ VEDECKÉ DÁTA**

**PLÁN MANAŽMENTU DÁT (DMP)**

# Princípy FAIR a vedecké dáta

## ČO ZNAMENÁ MANAŽMENT VEDECKÝCH DÁT

(RDM, Research Data Management)


**Komplex činností realizovaných pred, počas a aj po ukončení výskumu:**

- organizácia, ukladanie, uchovávanie a zdieľanie dát zhromaždených a použitých vo výskumnom projekte
- každodenný manažment výskumných dát počas trvania výskumného projektu (napr. používanie konzistentných konvencií na pomenovanie súborov)
- rozhodnutia o tom, ako sa budú dáta uchovávať a zdieľať po skončení projektu (napr. dlhodobá archivácia a prístupňovanie)



# Princípy FAIR a vedecké dáta

## PREČO MANAŽOVAŤ VEDECKÉ DÁTA

- dáta sú vedeckým produktom (podobne ako články v časopisoch a knihy)
- dátová explózia
- dáta sú krehké a ľahko sa strácajú (najmä digitálne)
- manažment výskumných dát dlhodobo šetrí čas a zdroje
- predchádza chybám a zvyšuje kvalitu analýz
- umožňuje overovať a replikovať zistenia
- uľahčenie zdieľanie výskumných dát  cenné objavy iných subjektov mimo pôvodného výskumu
- požiadavka financovateľov a vydavateľov



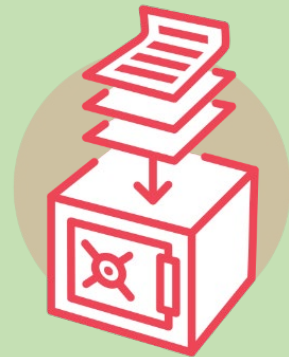


# Princípy FAIR a vedecké dáta

## ČO TREBA ZVÁŽIŤ PRI MANAŽMENTE VEDECKÝCH DÁT



**PLAN & DESIGN**



**MANAGE, STORE, PRESERVE**



**SHARE & PUBLISH**



**DISCOVER, REUSE & CITE**

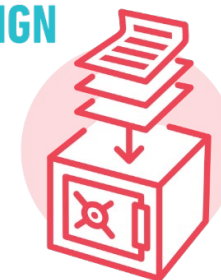
# Princípy FAIR a vedecké dáta



## Čo treba zvážiť pri manažmente vedeckých dát

**Plánovanie** – role a zodpovednosti, potreba nových zručností, potreba zdrojov navyše (ľudské, finančné, materiálne); náklady (manažment dát počas projektu, po ukončení)...

PLAN & DESIGN



MANAGE, STORE, PRESERVE

**Dokumentácia** – zabezpečí zrozumiteľnosť dát iným, ako sú dáta štruktúrované (napr. názvy, kódy, skratky); opisná a kontextová dokumentácia (napr. metódy, tvorba/zber dát); metadáta; pomenovávanie a organizácia dát, záznamov, súborov...



SHARE & PUBLISH

**Formátovanie dát, formáty** – štandardizované procesy a protokoly; dátové formáty a softvér (dlhodobé uchovávanie, zdieľanie); konvertovanie dát...



DISCOVER, REUSE & CITE



# Princípy FAIR a vedecké dáta

## ČO TREBA ZVÁŽIŤ PRI MANAŽMENTE VEDECKÝCH DÁT

**Uchovávanie** – bezpečnosť; osobné, dôverné, citlivé dáta, prístupy;

dĺžka uchovávania, výber dát na dlhodobé uchovávanie...

**Dôvernoscť, etické otázky/súhlasy** – dôverné, citlivé dáta; etická

komisia; informovaný súhlas/y...

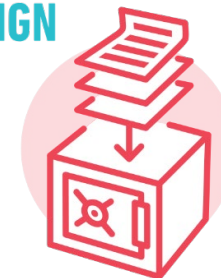
**Autorské práva** – primárne a sekundárne dáta – kto má AP, ktoré dáta

sa môžu zdieľať pod akými licenciami...

**Zdieľanie** – výber dát; ako, kde; spôsob prístupnosti...



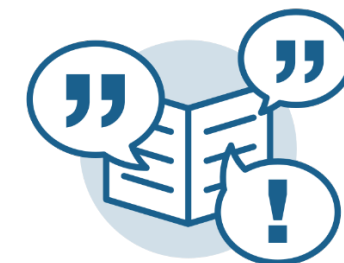
PLAN & DESIGN



MANAGE, STORE, PRESERVE



SHARE & PUBLISH



DISCOVER, REUSE & CITE

## PLÁN MANAŽMENTU DÁT

(DMP, Data Management Plan)

# Princípy FAIR a vedecké dáta

## PLÁN MANAŽMENTU DÁT PREČO?

### Politika povinných DMP

- [program Horizont Európa](#), Poľsko/[Národné centrum vedy](#), Slovensko – tiež sa plánuje zaviesť DMP v grantových agentúrach ([Národná stratégia pre otvorenú vedu](#))
- niektoré grantové agentúry a donori (napr. [ERC](#)) majú svoje **vzory DMP**, do ktorých výskumník vyplní údaje špecifické pre projekt
- výskumné inštitúcie – vo vnútorných predpisoch povinnosť vypracovávať DMP

# Princípy FAIR a vedecké dáta

## PLÁN MANAŽMENTU DÁT PREČO?

### Nástroj, ktorý vedcom umožňuje

- **systematicky premýšľať** o dátovom manažmente
- **robiť strategické rozhodnutia** v správnom čase
- **ujasniť si so spolupracovníkmi očakávania** ohľadom projektu
- **zjednotiť si so spolupracovníkmi postupy a spôsoby dokumentácie**

**DMP nehodnotí výskum – nejde o hĺbkovú dokumentáciu**

# Princípy FAIR a vedecké dáta

## PLÁN MANAŽMENTU DÁT

### KRÁTKY PREHĽAD SO ŠPECIFICKÝM ÚČELOM - FORMÁLNY DOKUMENT

- špecifikuje, **aké dáta a akým spôsobom** budú počas výskumu vytvárané a obsahuje **informácie o ich dostupnosti a využití**
- vypracováva sa na **úplnom začiatku výskumu**, ale **v jeho priebehu sa môže modifikovať** podľa aktuálneho vývoja situácie
- pri tvorbe sa môžu využiť **šablóny/predlohy** (inštitúcia, financovateľ)
- výskumníci môžu vytvoriť **svoj vlastný dokument** – zahrnúť témy, ktoré výskumníci (alebo financovateľ) považujú za dôležité (napr. zdieľanie dát)

# Princípy FAIR a vedecké dáta

## ŠTRUKTÚRA PLÁNU MANAŽMENTU DÁT





## PLÁNOVANIE A DIZAJN



PLAN & DESIGN

- Sú k dispozícii **odporúčania, usmernenia, politiky univerzity, inštitúcie, financovateľa** (napr. bezpečnosť, zdieľanie, uchovávanie, integrita, etika...)?
- Ponúka **univerzita/inštitúcia** nejaké **softvérové nástroje a podporu** v danej oblasti?

# Princípy FAIR a vedecké dáta

## ŠTRUKTÚRA PLÁNU MANAŽMENTU DÁT



# Princípy FAIR a vedecké dáta

existujúce dáta

dáta, ktoré sa objavia /vzniknú počas projektu

## DÁTA

**Existujú už dáta, ktoré môžete využiť?**

**Aký druh dát sa bude zbierať počas výskumu?**

- dotazníky, merania, texty, fotografie, rozhovory, skeny...

**Sú/budú medzi nimi špeciálne, citlivé kategórie dát?**

- osobné údaje, biologické vzorky, komerčné údaje...

**Predpokladaný objem získaných dát?**

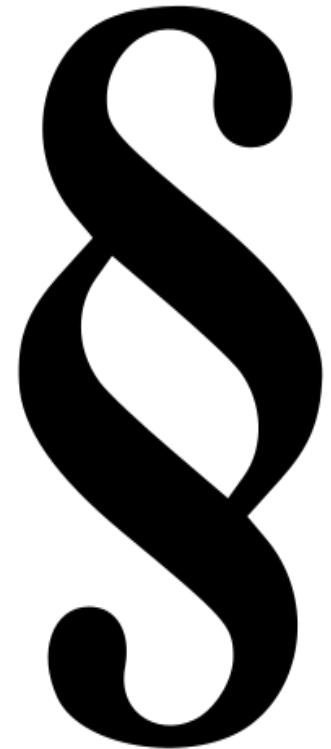
- dopad na výber repozitára (formáty)



## DÁTA

### Aké sú práva a povinnosti spojené s používaním dát iných strán?

- napr. z komerčných databáz – môžu mať v zmluvách zákaz zdieľania podkladových dát po ukončení projektu alebo obmedzenie objemu dát, ktoré môže výskumník extrahovať.
- autorské práva, osobitné právo k databáze, licencia, embargo, obmedzenia, transfer + zber dát mimo EU



# Princípy FAIR a vedecké dáta

## ETICKÉ A PRÁVNE ASPEKTY VÝSKUMU

### Treba pred začatím projektu žiadať o nejaký súhlas?

- osoby: **informovaný súhlas** – napr. pacientov alebo
- organizácie (napr. Štátnu veterinárnu a potravinovú správu, Štátnu ochranu prírody, ŠÚKL atď.), etickú komisiu...

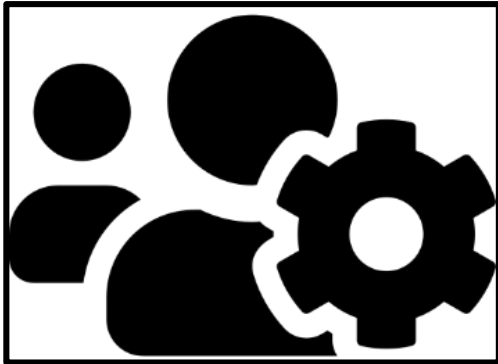
**Zásada dôvernosti, ochrana osobných, citlivých údajov, ochrana detí a zraniteľných skupín** (GDPR – [ZÁKON č.18/2018 Z.z.](#), z 29. novembra 2017 o ochrane osobných údajov a o zmene a doplnení niektorých zákonov – zásada obmedzenia účelu + minimalizácia a ochrana osobných údajov.

### **Dodržiavanie etických štandardov, vedeckej integrity**

(napr. [Deklarácia o posilnení kultúry vedeckej integrity na Slovensku](#))

# Princípy FAIR a vedecké dáta

zodpovednosť  
a povinnosti



## URČENIE ROLÍ A ZODPOVEDNOSTI

**Pre všetky aktivity s dátami v rámci výskumu:**

- stanoviť role/povinnosti a prideliť ich (napr. tabuľka úloh a zodpovedností s menami, opisom zodpovedností)
- hlavný riešiteľ, pokusy, programovanie
- kto bude zodpovedný za vedenie DMP, tvorbu metadát a dokumentácie
- ďalšie vybavenie/expertíza/školenia (**vedúci projektu, výskumní pracovníci, laboratórny/technický personál, externí dodávatelia, podporný personál** – napr. financovanie výskumu, **IT pracovníci, dátový steward/kurátor** a i.)
- kto bude zodpovedný za vedenie DMP, tvorbu metadát a dokumentácie
- potreba ďalšieho vybavenie/expertíz/školení ?



# Princípy FAIR a vedecké dáta

náklady a  
zdroje



## KALKULÁCIA NÁKLADOV

**Finančné náklady na zdroje** potrebné na zabezpečenie správnej praxe pri manažmente výskumných dát, napr.:

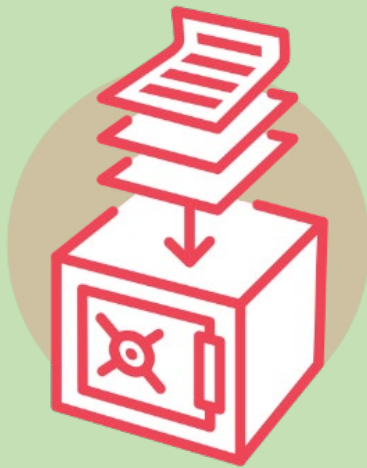
- čas pracovníkov,
- vybavenie,
- **infraštruktúra** (napr. repozitáre) a **nástroje** (napr. softvér na dokumentáciu, organizáciu, spracovanie, ukladanie dát a zabezpečenie prístupu k dátam (počas výskumu, po ukončení)...

**Data management costing tool and checklist**

<https://ukdataservice.ac.uk//app/uploads/costingtool.pdf>

# Princípy FAIR a vedecké dáta

## SPRACOVANIE, ORGANIZÁCIA, DOKUMENTÁCIA, UCHOVÁVANIE



**MANAGE, STORE, PRESERVE**

- Sú k dispozícii **odporúčania, usmernenia, politiky univerzity, inštitúcie, financovateľa** (napr. bezpečnosť, zdieľanie, uchovávanie, integrita, etika...)?
- Ponúka univerzita/inštitúcia nejaké **softvérové nástroje a podporu** v danej oblasti?

# Princípy FAIR a vedecké dáta

organizácia  
dát



Digitalbevaring.dk

**Spôsob organizácie údajov uľahčuje orientáciu v dátových súboroch, prispieva k pochopeniu obsiahnutých informácií a pomáha predchádzať chybám a nesprávnym interpretáciám.**

- organizácia a dokumentácia dát a metadát
- vhodné formáty, štruktúra dátových súborov
- pomenovanie dát, súborov..., verziovanie, štruktúra priečinkov...

# Princípy FAIR a vedecké dáta

## AKO SA BUDÚ DÁTA ZBIERAŤ A SPRACOVÁVAŤ?

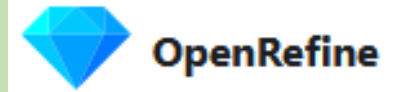
Kontrola kvality dát prebieha vo všetkých fázach výskumu:

1. zber dát
2. zadávanie dát a digitalizácia
3. čistenie a spracovanie „raw“ dát (overovanie dát)
4. vytvorenie a archivácia hlavnej verzie/*master file* (autenticita dát)

## NÁSTROJE

### OpenRefine

nástroj na čistenie tabuľkových dát (tabular data),  
opensource softvér, voľne stiahnuteľný



### Gephi

nástroj na vizualizáciu a skúmanie rôznych druhov diagramov a sietí, open-source softvér, voľne stiahnuteľný



### Nvivo

nástroj na kvalitatívnu analýzu dát, používa v rôznych oblastiach, free trial



### Datawrapper

online nástroj, pomocou neho môžete vytvárať tri druhy vizualizácií: mapy; grafy; tabuľky, bezplatný, bez registrácie

<https://www.datawrapper.de/>



# Princípy FAIR a vedecké dáta

## V AKÝCH FORMÁTOCH SA BUDÚ DÁTA UKLADAŤ?

- výber formátu súboru závisí od fázy výskumu – krátkodobé uchovávanie a dlhodobé uchovávanie
- opakované použitie; široko akceptované v rámci vednej disciplíny

## IDEÁLNY FORMÁT

- **neproprietárny** (nie je krytý patentom či copyrightom firmy)
- **nešifrovaný**
- **nekomprimovaný**
- **bežne používaný** výskumnou komunitou
- **v súlade s otvorenými štandardmi**

## FORMÁTY

- **Metasúbory (containers):** TAR, GZIP, ZIP
- **Databázy:** XML, CSV
- **Zemepisné:** SHP, DBF, GeoTIFF, NetCDF
- **Video:** MOV, MPEG, AVI, MXF
- **Zvuk:** WAVE, AIFF, MP3, MXF
- **Štatistika:** ASCII, DTA, POR, SAS, SAV
- **Obrázky:** TIFF, JPEG 2000, PDF, PNG, GIF, BMP
- **Tabuľkové údaje:** CSV
- **Text:** XML, PDF/A, HTML, ASCII, UTF-8
- **Webové archívy:** WARC

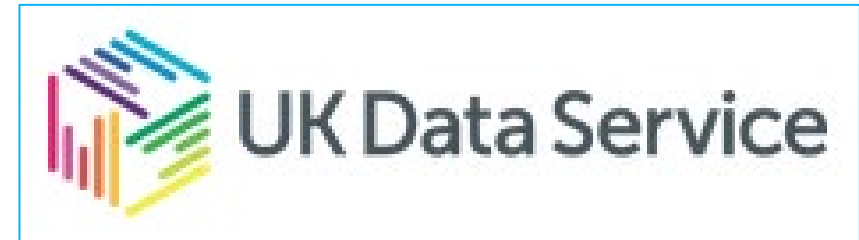


# Princípy FAIR a vedecké dáta

## FORMÁTY



<https://dans.knaw.nl/en/file-formats/>

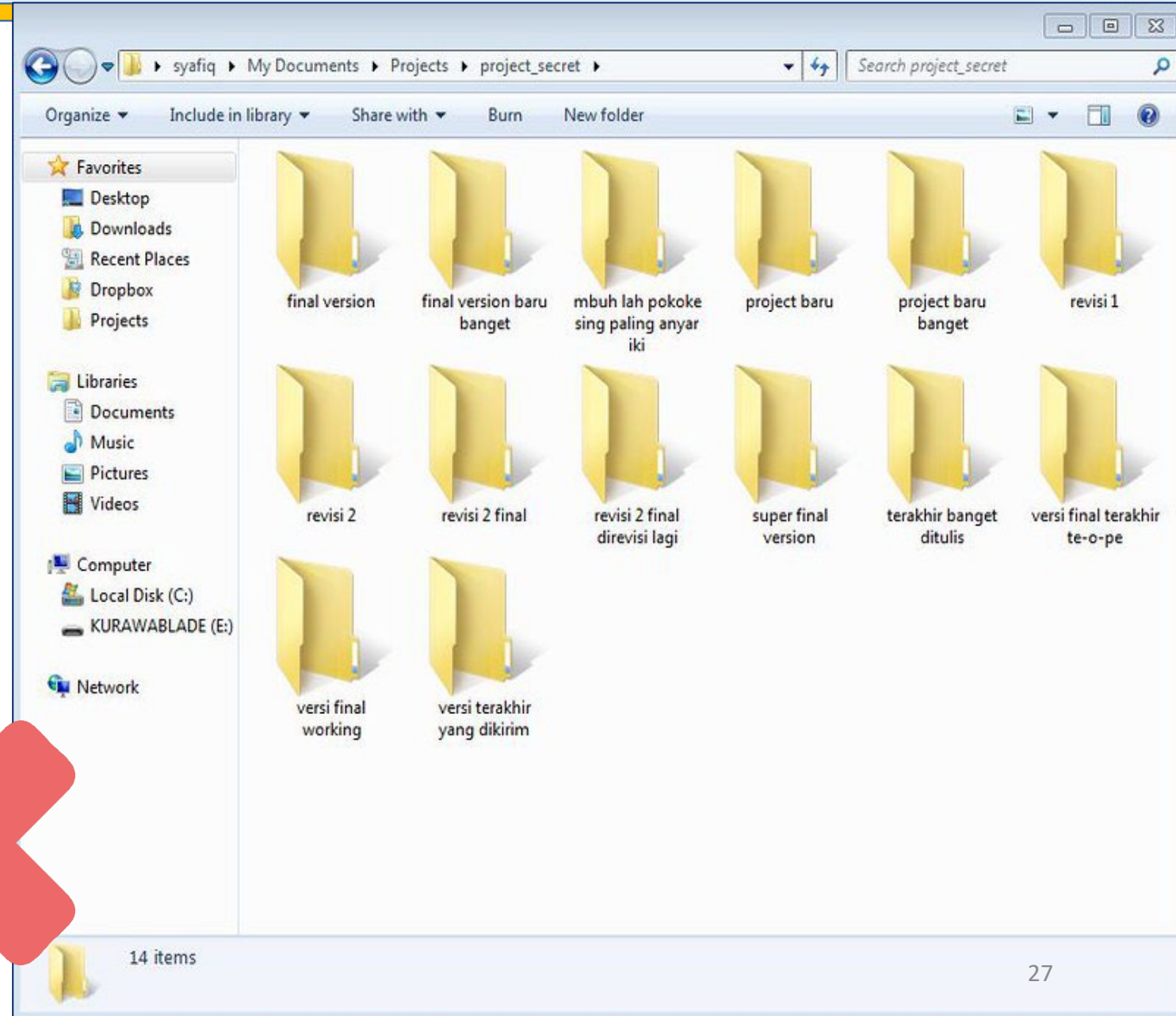


<https://ukdataservice.ac.uk/learning-hub/research-data-management/format-your-data/recommended-formats/>

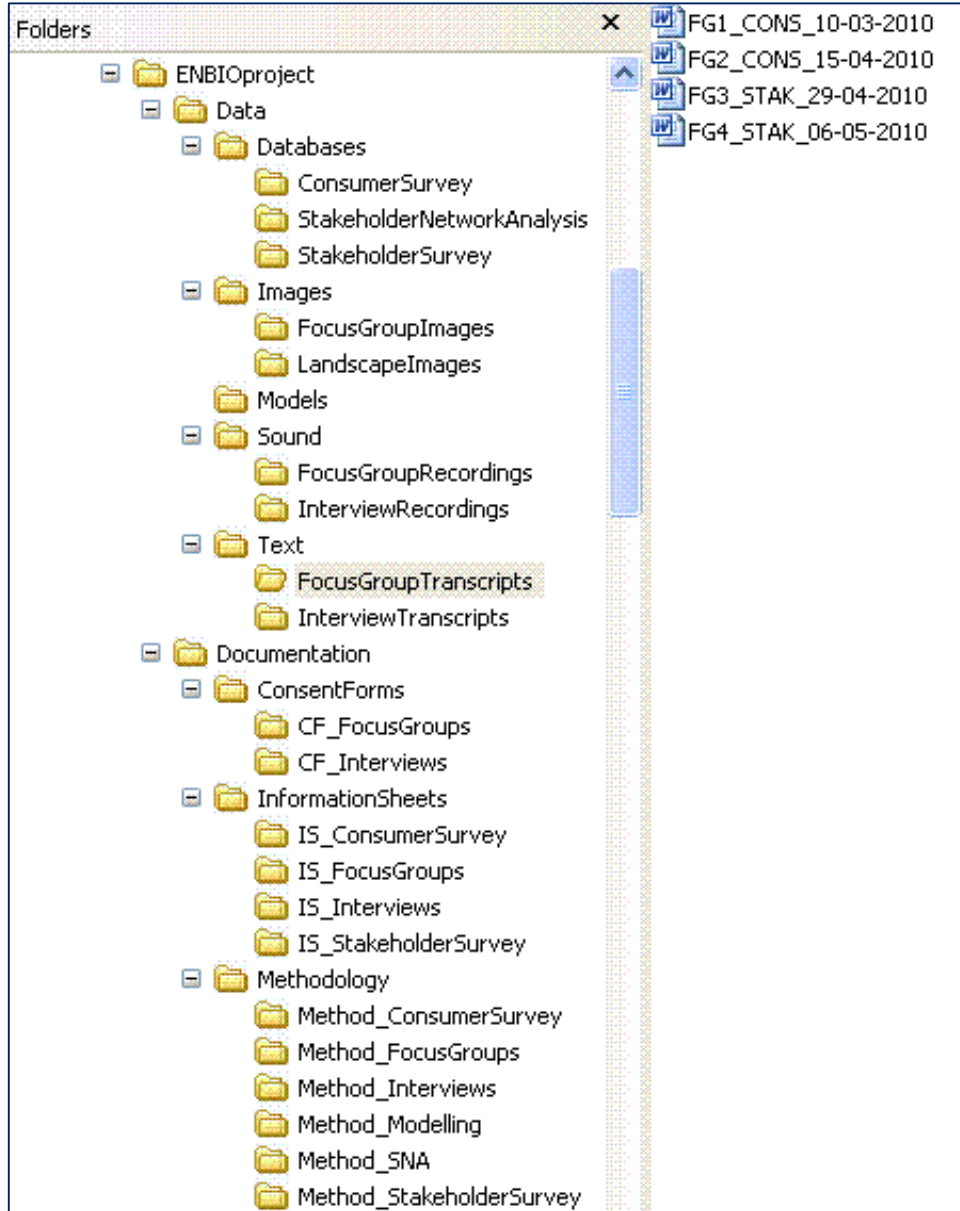


# Princípy FAIR a vedecké dáta

**AKÁ JE STRATÉGIA  
ORGANIZÁCIE/ŠTRUKTÚRY  
A POMENOVÁVANIA  
PRIEČINKOV, ZLOŽIEK,  
SÚBOROV?**



# Princípy FAIR a vedecké dáta



## ŠTRUKTÚRA PRIEČINKOV, ZLOŽIEK, SÚBOROV

- ľahké vyhľadávanie (Dôkladne si premyslite, ako najlepšie usporiadať súbory v priečinkoch, aby ste uľahčili vyhľadávanie a organizáciu súborov a verzií (aj sebe))
- spolupráca s inými (Pri práci v spolupráci s ostatnými je potreba usporiadanej štruktúry ešte vyššia.)
- hlboká alebo plytká hierarchia (Zvážte najlepšiu hierarchiu pre súbory a rozhodnite sa, či je vhodnejšia hlboká alebo plytká hierarchia. Je dobré obmedziť úroveň priečinkov na tri alebo štyri hlboké a nemať v každom zozname viac ako desať položiek.)
- nezávislé zbierky údajov ( vytvoriť pre každú zbierku samostatný priečinok s údajmi.)



# Princípy FAIR a vedecké dáta

## POMENOVANIE PRIEČINKOV, ZLOŽIEK, SÚBOROV (*FILE NAMING CONVENTION*)

Umožňuje efektívnu identifikáciu, lokalizáciu a používanie výskumných dátových súborov

### Stratégia pomenovania súborov

- stratégia pomenovania súborov (*file naming convention*) má byť **konzistentná v čase a medzi rôznymi ľuďmi**
- **názov súboru** je hlavný identifikátor súboru (konzistentné, popisné)

### Prvky, ktoré treba zohľadniť pri tvorbe stratégie pomenovania

- Číslo verzie
- Dátum vytvorenia (formát dátumu: RRRR-MM-DD/RR-MM-DD)
- Meno tvorcu
- Opis obsahu
- Názov výskumného tímu/oddelenia spojeného s dátami
- Dátum zverejnenia
- Číslo projektu

### Hromadné premenovanie súborov

**Bulk Rename Utility** –

*aplikácia Microsoft, bezplatná*

**Ant Renamer**

**RenameIT**

**Rename4Mac**

# Princípy FAIR a vedecké dáta

## POMENOVANIE PRIEČINKOV, ZLOŽIEK, SÚBOROV (*FILE NAMING CONVENTION*)

### Odporúčaná postup

- **krátke** ale výstižné **názvy**
- používať **názvy súborov** na **klasifikáciu typov súborov**
- používať **dátumy** vo formáte RRRR-MM-DD
- **oddelenie prvkov v názve súboru**: **používať spojovník (-)** alebo **podčiarkovníky (\_)**, **nepoužívajte** medzery, bodky a špeciálne znaky `~ ! @ # $ % ^ & * ( ) ` ; < > ? , [ ] { } ""`
- 3-písmenovú príponu súboru vyhradiť pre **kódy formátu súboru** (napr. .doc, .xls, .mov, .tif).
- zahrnúť do názvov súborov **označenie verzií**, ak treba
- **číselná zložka**, napr. číslo predmetu alebo číslo verzie – **použiť úvodné nuly** (001, 002 atď.), aby sa súbory zoradili v postupnom poradí

# Princípy FAIR a vedecké dáta

## POMENOVANIE PRIEČINKOV, ZLOŽIEK, SÚBOROV (*FILE NAMING CONVENTION*)

Dokumentácia: zaznamenajte zloženie svojej konvencie pomenovania súborov, napr.:

`<date><type><ID1><gender><age><municipality><datatype><ID2>`

### Príklad pomenovania súborov kvalitatívnych dát

20130311\_interview2\_audio.wav

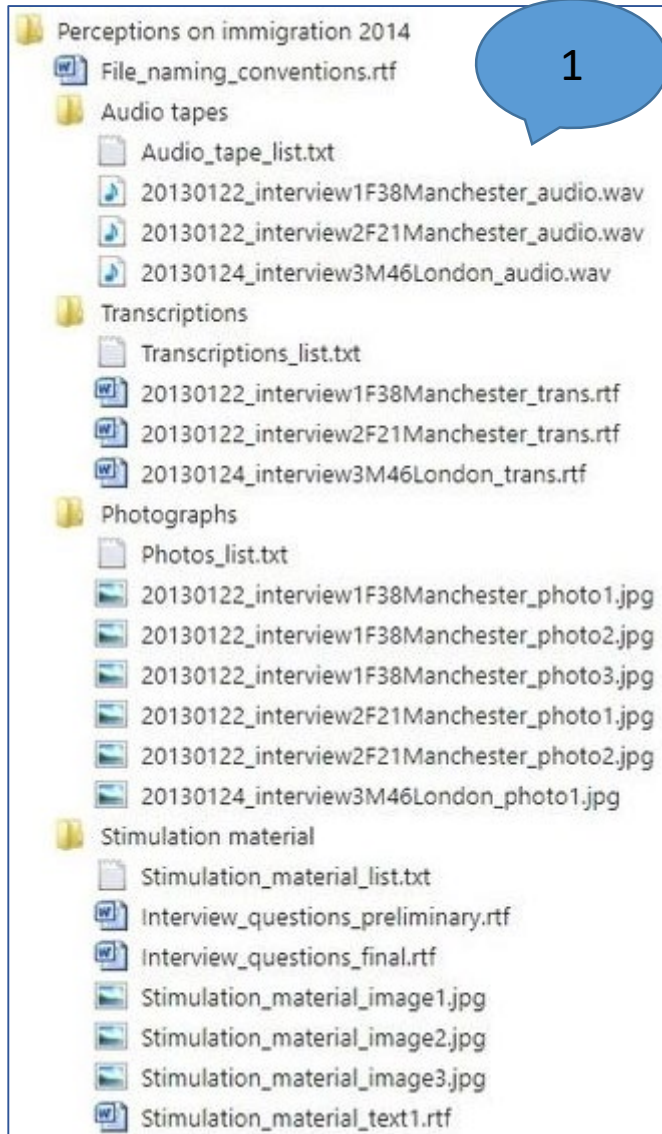
**dátum** - **typ udalosti+ID** - **typ dát** - **formát**

- `<date>` - dátum zhromaždenia dát (formát dátumu by mal byť RRRR-MM-DD);
- `<type>` - typ udalosti/dát;
- `<ID1>` - ID udalosti zberu;
- `<gender>` - pohlavie respondenta;
- `<age>` - vek respondenta;
- `<obec>` - obec bydliska respondenta;
- `<datatype>` - typ dát, ktoré súbor obsahuje, napríklad "trans" znamená prepis, "audio" znamená zvukový záznam a "image" znamená fotografiu;
- `<ID2>` - identifikačné číslo, ktoré sa používa na oddelenie snímok spojených s udalosťou zberu.



# Princípy FAIR a vedecké dáta

## POMENOVANIE PRIEČINKOV, ZLOŽIEK, SÚBOROV (*FILE NAMING CONVENTION*)



### PRÍKLADY

#### Súbory s FNC

- 20170503\_BIOProject\_DesignDocument\_Jones\_v1-01.docx
- 20170709\_BIOProject\_MasterData\_Smith\_v1-00.xlsx
- 20170825\_BIOProject\_Test1\_Data\_Gonzalez\_v1-01.xlsx
- 20170825\_BIOProject\_Test2\_Data\_Gonzalez\_v2-01.xlsx
- 20171023\_BIOProject\_MeetingNotes\_Kramer\_v1-00.docx

#### Súbory bez FNC

- Test\_data\_2017
- Research\_Data\_2016
- Project\_Design\_2017
- Observations\_2016
- Notes\_2016

Ďalšie príklady na:  
<https://dmeg.cessda.eu/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-and-folder-structure>

# Princípy FAIR a vedecké dáta



## AKO SA BUDE RIEŠIŤ VERZIOVANIE?

Spôsob, akým spravujeme viaceré varianty/verzie toho istého dokumentu, súboru


### Stratégia kontroly verzií

- vytvoriť **hlavný súbor/master file** + prijať opatrenia na zachovanie jeho autenticity
- **koľko verzií súboru** uchovávať, **ktoré verzie** uchovávať, **ako dlho** a **ako verzie organizovať**
- **ktoré verzie míľnikov** sa majú uchovávať (napr. **hlavné verzie namiesto menších verzií**: verziu 02-00 ale nie 02-01) + **určiť jedno miesto na ukladanie** míľnikov a hlavných verzií
- **jedinečná identifikácia rôznych verzií** súborov pomocou systematického pomenovania (napr. napríklad pomocou čísel verzií alebo dátumov)
- zaznamenávať **zmeny** vykonané v súbore **pri vytváraní novej verzie** (vzťahy medzi položkami)
- **sledovať umiestnenie súborov**, ak sú uložené na rôznych miestach
- **pravidelne synchronizovať súbory na rôznych miestach** (napr. softvér MS SyncToy)

# Princípy FAIR a vedecké dáta

## AKO SA BUDE RIEŠIŤ VERZIOVANIE?

### SPÔSOBY KONTROLY VERZIÍ

- **dátum zaznamenaný v názve súboru** alebo v rámci súboru (napr. **HealthTest-2008-04-06**)
- **číslovanie verzie v názve súboru** (napr. **HealthTest-00-02** alebo **HealthTest\_v2**)
- **história súborov, tabuľka kontroly verzií alebo poznámky zahrnuté v súbore**, kde sú zaznamenané verzie, dátumy, autori a podrobnosti o zmenách v súbore
- funkcie na kontrolu verzií v rámci použitého softvéru
- softvér na správu verzií (napr. [Subversion](#))
- služby na zdieľanie súborov (napr. Dropbox  Docs !)
- kontrola práv na úpravu súborov (kto môže upravovať)
- manuálne spájanie záznamov alebo úprav viacerými používateľmi

[Projekt Dolt](#)

d01thUb



### BIG DATA

„V prípade veľkých súborov dát nemusí byť ukladanie všetkých revízií každého záznamu správnym prístupom. ...

Časté aktualizácie, ktoré sa nikdy nečítali, by sa teda nemuseli – z hľadiska citovania údajov – verziovať.“

# Princípy FAIR a vedecké dáta

## AKO ZABEZPEČÍTE SVOJE DÁTA? OCHRANA DÁT (DATA SECURITY)

- ochrana pre neoprávneným prístupom k dátam a neoprávnenou manipuláciou či vymazaním súborov
- ochrana osobných údajov, dôverných informácií

### PRIJATÍM TECHNICKÝCH ALE AJ ORGANIZAČNÝCH OPATRENÍ:

- heslá + kontrolovaný prístup k jednotlivým dátovým súborom, **napr. pridelenie oprávnení "bez prístupu", "len na čítanie", "na čítanie a zápis" alebo "len pre správcu"**
- šifrovanie/kódovanie (data encryption) : nástroje (softvér)  
[BitLocker](#) – program Microsoftu, standard on selected editions of Windows; for the encryption of disk volumes and USB devices, dostupný v niektorých verziách operačného systému Windows  
[FileVault2](#) – program na šifrovanie disku v systéme Mac OS X 10.3 Panther, standard on Apple Macs; for full disc encryption, proprietary  
[VeraCrypt](#) – multi-platform encryption software (Windows, Mac and Linux); for full-disk and container encryption, open-source, pod licenciou převážně Apache License 2.0, části pod licencií TrueCrypt 3.0  
[Axcrypt](#) – open-source file-level encryption for Windows, enkrypcia/šifrovanie súborov
- rozdelenie dát (napr. **uchovávanie mien a adries v rôznych dátových súboroch**)
- fyzické zabezpečenie ochrany dát (**prístup do miestností, k počítačom, transport citlivých dát...**)
- bezpečnosť sietí a počítačových systémov (**externé siete, firewall, pravidelné aktualizácia softvéru...**)
- bezpečná likvidácia dát (**nestačí iba použiť funkciu „vymazať/delete“**)
- bezpečné uchovávanie (napr. **služby zdieľania súborov ako Google Docs, OneDrive a Dropbox, nemusia byť vhodné pre dôverné/citlivé dáta; cloudové úložiská ako Disk Google, Dropbox, OneDrive, iCloud alebo YouSendIt sa ľahko používajú, ale nemusia byť trvalé alebo bezpečné**)
- **ľudský faktor**, obzvlášť dôležité pri spolupráci a vo veľkých tímoch (**jasné pravidlá, politiky, usmernenia, jasná komunikácia, obmedzenie prístupu k dátam na základe potreby...**)

## AKO BUDETE UKLADAŤ A ZÁLOHOVAŤ DÁTA V PRIEBEHU PROJEKTU?

Zálohovanie – umožňuje obnoviť dáta v prípade poškodenie/straty pôvodných dát

**Aké otázky si položiť v rámci analýzy rizík pri plánovaní stratégie zálohovania výskumných dát?**

### OTÁZKY PRI ANALÝZE RIZÍK SÚVISIACICH SO STRATÉGIOU ZÁLOHOVANIA VÝSKUMNÝCH DÁT

- Prijala inštitúcia nejakú politiku súvisiacu so zálohovaním? Má už nejaké riešenia? –Väčšina univerzít má politiku pre súbory uchovávané v rámci univerzitnej siete.
- Ktoré informácie/dáta/systémy zálohovať? –Môžu to byť **všetky (full back-up- plná záloha)**, niektoré alebo len tie časti, ktoré sa zmenili. Inteligentné zálohovanie – **rozdielová záloha (differential back-up -nedochádza k opätovnému zálohovaniu všetkých súborov, ale pouze súborov, ktoré byly od počáteční plné zálohy změněny. Rozdílem oproti přírůstkové metodě je pak to, že nedochází pouze k aktualizaci souborů o provedené změny, ale dochází k novému uložení celého souboru, který byl od počáteční plné zálohy změněn. Tímto dochází k vyřešení problémů s možnou ztrátou dat vlivem ztráty jedné předchozí zálohy, jelikož je možný přístup k celým změněným souborům\*) a přírůstkové (incremental back-up, při tomto druhu zálohování dat se plná záloha vytváří pouze při počátečním zálohování a dále se již zálohují pouze změny souborů, které byly od poslední plné či přírůstkové zálohy změněny. Tyto změny se zapisují přímo do konkrétních souborů, které byly vytvořeny počáteční plnou zálohou\* [https://is.muni.cz/th/vxsr1/PREDELAVANA\\_FINAL.pdf](https://is.muni.cz/th/vxsr1/PREDELAVANA_FINAL.pdf) )** zálohy sa nazývajú aj "inteligentné" zálohy. Ak sa denne mení len malé percento vašich údajov, je plytvanie časom a miestom na disku, aby sa každý deň spúšťala plná záloha.
- Aké formáty používať?–Zálohy hlavných kópií súborov (master copies) by mali byť vo formátoch ([file formats](#)) vhodných na dlhodobé digitálne uchovávanie, t. j., otvorené alebo štandardné formáty a nie proprietárne formáty.
- Ako často zálohovať?–Zvážiť, ako často sa robia zmeny dát, a aké množstvo dát ste pripravený stratiť medzi zálohovaním. Zvážiť zálohovanie po každej zmene dátového súboru alebo v pravidelných intervaloch, napr. denne alebo týždenne.
- Koľko kópií vyhotoviť? –aspoň 3 kópie dát, pričom aspoň 1 má byť uložená mimo miesta realizácie výskumu (offsite).



## Aké otázky si položiť v rámci analýzy rizík pri plánovaní stratégie zálohovania výskumných dát?

### OTÁZKY PRI ANALÝZE RIZÍK SÚVISIACICH SO STRATÉGIU ZÁLOHOVANIA VÝSKUMNÝCH DÁT

- **Kde uchovávať záložné súbory? Koľko miesta bude treba?** –V závislosti od formy zálohovania a rizika spojeného so stratou dát je najvhodnejšie záložné súbory na sieťovom pevnom disku (networked hard drive). Pri kritických dátach, nedostupných inde, sa odporúča aj uchovávanie offline na optických médiách, vymeniteľných/prenosných pevných diskoch (removable hard drives) alebo magnetických páskach (magnetic tapes). ... Nespoliehať sa na usb kľúče (pen drives). Fyzické médiá majú byť uložené bezpečne na inom mieste. Väčšina výrobcov poskytuje odporúčania, za akých podmienok je najlepšie uchovávať fyzické médiá. Záložné súbory obsahujúce osobné údaje vyžadujú šifrovanie (encryption) a majú sa ukladať bezpečne.
- **Ako organizovať záložné súbory?** –Zaistiť, aby záložné súbory na vymeniteľných médiách boli dobre označené, s uvedením obsahu a dátumu/času, a dobre organizované. Bez manažmentu dát, môže byť ťažké dosiahnuť cieľ – obnova stratených dát.
- **Aké nástroje sú k dispozícii?** –Na zálohovanie často používaných a kritických dátových súborov je vhodné používať automatizovaný proces zálohovania. Systémy Windows ([File History](#)) aj MacOS ([Time Machine](#)), ktoré zálohovanie veľmi uľahčujú.
- **Ako uchovávať osobné údaje?** -Ak dáta obsahujú osobné údaje, treba dbať na to, aby sa vytvoril len minimálny počet potrebných kópií, napríklad hlavný súbor a jedna záložná kópia, ktorá je zašifrovaná a bezpečne uložená. Veľa kópií osobných údajov - problém likvidovať po ukončení projektu.
- **Ako budete overovať integritu záložných súborov?** -Pri zápise alebo kopírovaní záloh môže dôjsť k chybám. **Odporúča sa častá kontrola integrity záložných súborov.** Môžete to urobiť pomocou takzvaných nástrojov na kontrolu integrity kontrolných súčtov (checksum tools), ako sú [MD5summer](#) or [Checksum Checker](#). Checksum - reťazec čísel - z bitových hodnôt (jednotiek a núl) súboru. Sledovanie toho, či sa „odtlačok prsta“ daného súboru mení, umožňuje zistiť, či bol súbor nejakým spôsobom úmyselne alebo neúmyselne zmenený. UK Data Service prirovnáva checksums k digital fingerprints.
- **Ako dlho sa budú zálohy uchovávať a ako sa budú likvidovať?** –Ak treba často zálohovať obrovské množstvá dát (large amounts of data), nemusí byť možné uchovávať všetky zálohy počas celého trvania projektu. **V prípade citlivých dát sa uistite, že akékoľvek vymazané dáta sú naozaj vymazané (gone) a nedajú sa žiadnym spôsobom obnoviť.**

# Princípy FAIR a vedecké dáta

## ZÁLOHOVANIE V SKRATKE: „pravidlo 3-2-1“



Nástroje na kontrolu integrity súborov prostredníctvom kontrolných súčtov (checksum tools): [MD5summer](#) alebo [Checksum Checker](#)

Nástroje na zálohovanie: systém Windows ([File History](#)); systém MacOS ([Time Machine](#))

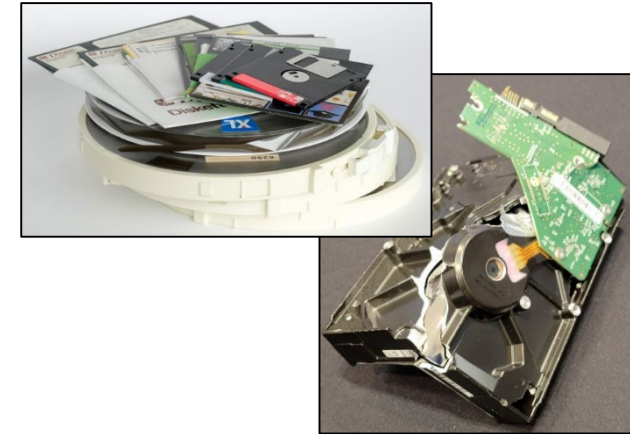
# Princípy FAIR a vedecké dáta

## DLHODOBÉ UCHOVÁVANIE VÝSKUMNÝCH DÁT PO UKONČENÍ PROJEKTU

**Ktoré z množstva získaných dát majú byť uchovávané dlhodobo?**

**Posúdiť hodnotu svojich dát, dlhodobo uchovávať len dáta s vysokou hodnotou:**

- všetky podkladové dáta súvisiace s publikáciou
- dáta, ktorých zber je náročný alebo nákladný
- dáta, ktoré sa nedajú nahradiť/znovu vygenerovať



## KTORÉ DÁTA MAJÚ BYŤ ZNIČENÉ/VYMAZANÉ? AKO?

- biologický materiál, ktorý sa časom znehodnotí
- dôverné údaje alebo osobné údaje, ktoré nemožno anonymizovať

Nástroj DCC ['Five steps to decide what data to keep: a checklist for appraising research data'](#)

## AKO DLHO PO UKONČENÍ PROJEKTU DÁTA UCHOVÁVAŤ?

- minimálne obdobie pre dáta, ktoré sú podkladom pre publikácie
- dodržiavať právne predpisy, príp. nariadenia platné pre váš typ dát, pravidlá či usmernenia inštitúcie...

Softvér na vymazanie súborov z pevných diskov: [BCWipe](#), [WipeFile](#), [DeleteOnClick](#) a [Eraser](#) pre platformy Windows; a [Permanent Eraser](#) pre platformy MacOS.



# Princípy FAIR a vedecké dáta

## ZDIEĽANIE, PUBLIKOVANIE: OPAKOVANÉ POUŽITIE, CITOVANIE



SHARE & PUBLISH



DISCOVER, REUSE & CITE

- Sú k dispozícii odporúčania, usmernenia, politiky univerzity, inštitúcie, financovateľa (napr. bezpečnosť, zdieľanie, uchovávanie, integrita, etika...)?
- Ponúka univerzita/inštitúcia ponúka nejaké softvérové nástroje a podporu v danej oblasti?

# Princípy FAIR a vedecké dáta

## AKO ZABEZPEČÍTE ZROZUMITEĽNOSŤ DÁT PRE INÝCH VEDCOV...?

- opis použitých štandardov a metód pomôže pochopiť iným, ako ste realizovali projekt (dokumentácia – kontextové informácie, na úrovni dát, na úrovni štúdie, nástroje; readme súbor)

## AKO BUDETE DÁTA DOKUMENTOVAŤ, OPISOVAŤ?

**Metadát je veľa – ktoré budú kľúčové pre projekt?**

### Metadáta môžu byť:

**Opisné metadáta** – informácie o obsahu a kontexte dát

*Například:* názov, tvorca, kľúčové slová predmetu a opis (abstrakt)

**Štrukturálne metadáta** – opisujú fyzickú štruktúru zložených dát

*Například:* použitý fotoaparát, clona, expozícia, formát súboru a vzťah k iným údajom alebo súborom

**Administratívne metadáta** – informácie používané na manažment dát

*Například:* kedy a ako boli vytvorené, kto k nim môže pristupovať, softvér potrebný na ich použitie a autorské práva

Zdroj: <https://guides.lib.unc.edu/metadata/definition>



# Princípy FAIR a vedecké dáta

**AKO BUDETE DÁTA DOKUMENTOVAŤ, OPISOVAŤ?**

**Metadát je veľa – ktoré budú kľúčové pre váš projekt?**

## METADÁTA A metadátové štandardy

- všeobecný napr. [Dublin Core](#)
- odborové metadátové štandardy, napr. [DDI](#), [Darwin Core](#),  
[EML](#), [VRA Core](#)...

Odkazy na metadátové štandardy – DCC a RDA

<https://www.dcc.ac.uk/guidance/standards/metadata>



Metadata Standards Catalog

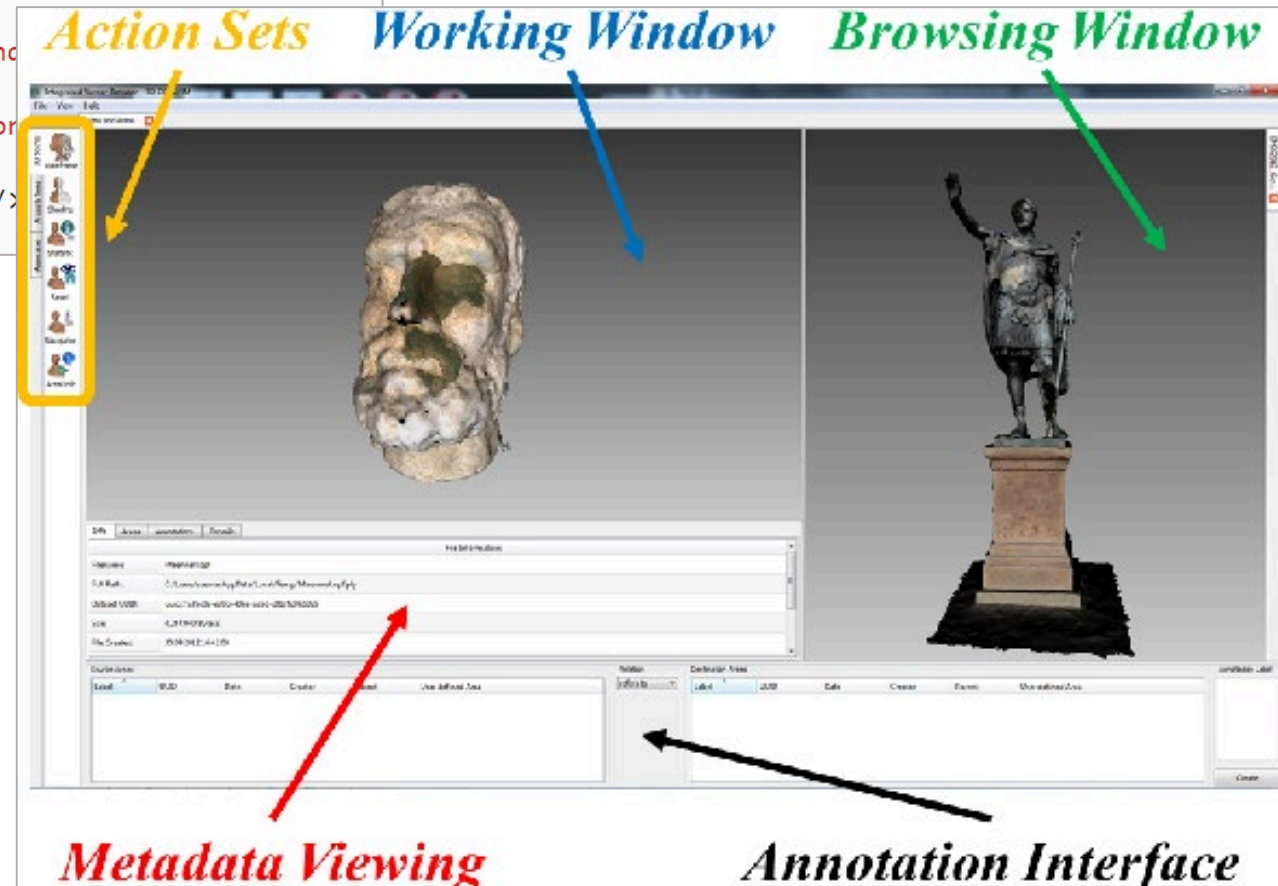


# Princípy FAIR a vedecké dáta

## METADÁTA - PRÍKLADY



```
<head profile="http://dublincore.org/documents/dcq-html/">
<title>Dublin Core</title>
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<link rel="schema.DCTERMS" href="http://purl.org/dc/terms/" />
<meta name="DC.format" scheme="DCTERMS.IMT" content="text/html" />
<meta name="DC.type" scheme="DCTERMS.DCMIType" content="Text" />
<meta name="DC.title" content="Dublin Core" />
<meta name="DC.publisher" content="Jimmy Wales" />
<meta name="DC.subject" content="Dublin Core Metadaten-Elemente, Anwend
<meta name="DC.creator" content="Björn G. Kulms" />
<meta name="DCTERMS.license" scheme="DCTERMS.URI" content="http://www.gnu.or
<meta name="DCTERMS.rightsHolder" content="Wikimedia Foundation Inc." />
<meta name="DCTERMS.modified" scheme="DCTERMS.W3CDTF" content="2006-03-08" />
</head>
```



[Zdroj obrázka:](#) Serna, S.P., Schmedt, H., Ritz, M., & Stork, A. (2012). Interactive Semantic Enrichment of 3D Cultural Heritage Collections. *IEEE Conference on Visual Analytics Science and Technology*.

**AKO BUDETE DÁTA DOKUMENTOVAŤ, OPISOVAŤ?**

**RIADENÉ SLOVNÍKY**

**Príklady a odkazy na riadené slovníky**

<https://guides.lib.utexas.edu/metadata-basics/controlled-vocabs>



<https://www.openaire.eu/rdm-glossary>

<https://bartoc.org/>

<https://bartoc.org/vocabularies>

[Library of Congress  
Authorities for subject  
headings](https://authorities.loc.gov/)

<https://authorities.loc.gov/>

[ICD-11: International Classification of Diseases 11th  
Revision](https://www.who.int/standards/classifications/classification-of-diseases)

[The global standard for diagnostic health information](https://www.who.int/standards/classifications/classification-of-diseases)

<https://www.who.int/standards/classifications/classification-of-diseases>

**The Astronomy  
Thesaurus**

<https://www.mso.anu.edu.au/library/thesaurus/>

[Art & Architecture Thesaurus® Online](http://www.getty.edu/research/tools/vocabularies/aat/)

**Getty Research Institute**

<http://www.getty.edu/research/tools/vocabularies/aat/>

# Princípy FAIR a vedecké dáta

## RIADENÉ SLOVNÍKY - PRÍKLAD

AGROVOC Multilingual Thesaurus Content language English ▾

Alphabetical Hierarchy

... > Pinopsida > Pinales > Pinaceae > Abies > Abies magnifica

**PREFERRED TERM** ⓘ **Abies magnifica** 📄

---

**BROADER CONCEPT** Abies (en)

---

**ENTRY TERMS** ⓘ *California red fir* (en)  
ⓘ *red fir* (en)

---

**HAS TAXONOMIC RANK** species (taxa) (en)

---

**IN OTHER LANGUAGES**

ⓘ تنوب بهي	Arabic
ⓘ 加州红冷杉	Chinese
ⓘ 红果冷杉	
ⓘ 加利福尼亚冷杉	
ⓘ <b>Abies magnifica</b>	Czech
ⓘ <i>Abies nobilis magnifica</i>	
ⓘ <i>jedle nádherná</i>	
ⓘ <b>Abies magnifica</b>	French
ⓘ <i>Sapin rouge californien</i>	
ⓘ <b>Abies magnifica</b>	Georgian
ⓘ <b>Abies magnifica</b>	German



## AKÝM SPÔSOBOM SA BUDÚ ZDIEĽAŤ DÁTA, KTORÉ MAJÚ BYŤ ZVEREJNENÉ?

- ktoré dáta sa budú zdieľať, kde (**dátový repozitár, dátový časopis**), ako ich nájdú iní (metadáta), opakované použitie (licencie)

### Dátový repozitár

obrovská databázová infraštruktúra, ktorá zhromažďuje, spravuje a ukladá rôzne súbory dát na analýzu, distribúciu a tvorbu požadovaných výstupov.

### Typy dátových repozitárov

Data Warehouse

Data Lake

Data Marts

Data Cubes

Metadata Repositories

### Ako nájsť vhodný OA repozitár?

Medzinárodné registre repozitárov s otvoreným prístupom

- **Registry of Open Access Repositories**, [ROAR](http://roar.eprints.org/), (<http://roar.eprints.org/>)
- **Open Directory of Open Access Repositories**, [OpenDOAR](https://v2.sherpa.ac.uk/opensoar/) (<https://v2.sherpa.ac.uk/opensoar/>)
- [re3data](https://www.re3data.org/) - **Registry of Research Data Repositories**, (<https://www.re3data.org/>)
- **Registry of Open Access Repository Mandates**, [ROARMAP](http://roarmap.eprints.org/), (politiky, <http://roarmap.eprints.org/>)

# Princípy FAIR a vedecké dáta

## OVERENIE DÔVERYHODNOSTI REPOZITÁRA

### Referenčný model OASIS pre architektúru zameranú na služby

*OASIS: nezisková organizácia, vznikla v r. 1993*

#### Repozitár by mal spĺňať 10 podmienok:

1. Musí prijať **záväzok dlhodobej a udržateľnej správy** digitálnych objektov pre cieľovú komunitu.
2. Musí **preukázať spôsobilosť a zdroje** pre zabezpečenie udržateľnosti.
3. Musí mať **jasné formálne a legislatívne rámce** ako východisko svojich záväzkov.
4. Musí mať efektívnu a udržateľnú rámcovú **stratégiu**.
5. Musí definovať a prijať **inštitucionálne interné politiky** pre získavanie a ukladanie digitálnych objektov.
6. Musí byť schopný poskytnúť **zabezpečenie integrity**, autenticity a dostupnej použiteľnosti digitálnych objektov.
7. **Vytvára a uchováva relevantné metadáta** o digitálnych objektoch, obsahu, ako aj udalostiach súvisiacich s objektmi, vrátane podmienok prístupňovania a kontextu použitia.
8. Musí byť schopný zabezpečiť splnenie požiadaviek na **prístupňovanie** digitálnych objektov konkrétnej komunite.
9. Musí mať spracovanú **stratégiu dlhodobej ochrany** vrátane riešenia bezpečnostných konfliktov a havárií systému.
10. Musí mať **technickú infraštruktúru adekvátnu udržateľnej údržbe** a zaistenia digitálnych objektov.



# Princípy FAIR a vedecké dáta

## OVERENIE DÔVERYHODNOSTI REPOZITÁRA – CERTIFIKÁTY



### ISO 16363 Audit and certification of trustworthy digital repositories

- komplexný externý audit + certifikácia
- repozitáre s rôznym obsahom, on-site

### CoreTrustSeal

- dátové repozitáre
- menej metrík, kratšia doba trvania, každé tri roky, off-site



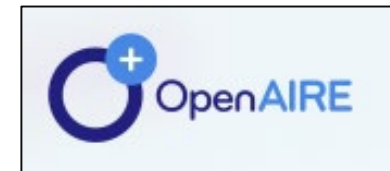
### Nestor Seal

podľa normy nemeckej normy DIN 31644

### Európska komisia a projekty financované EK:

FAIR dáta v dôveryhodných repozitároch

<https://www.openaire.eu/find-trustworthy-data-repository>



# Princípy FAIR a vedecké dáta



## DÁTOVÉ REPOZITÁRE - PRÍKLADY

Repository Name	Information on fees/costs	Size limits	Integrated with <i>Scientific Data's</i> manuscript submission system	Re3data / FAIRsharing entry
<a href="#">Dryad Digital Repository</a>	\$120 USD for first 20 GB, and \$50 USD for each additional 10 GB	<a href="#">None stated</a>	<b>Yes ✓</b>	<a href="#">view FAIRsharing entry</a>
<a href="#">figshare</a>	100 GB free per <i>Scientific Data</i> manuscript.	1 TB per dataset	<b>Yes ✓</b> - To qualify for the 100 GB of free storage, data must be uploaded to figshare via our submission system. <a href="#">Download instructions.</a>	<a href="#">view FAIRsharing entry</a>
<a href="#">Harvard Dataverse</a>	<a href="#">Contact repository</a> for datasets over 1 TB	2.5 GB per file, 10 GB per dataset	No	<a href="#">view re3data entry</a>
<a href="#">Open Science Framework</a>	<a href="#">Free of charge</a>	5 GB per file, multiple files can be uploaded	No	<a href="#">view FAIRsharing entry</a>
<a href="#">Zenodo</a>	<a href="#">Donations towards sustainability encouraged</a>	50 GB per dataset	No	<a href="#">view re3data entry</a>
<a href="#">Science Data Bank</a>	<a href="#">Free of charge</a>	8 GB per file, no limit to dataset size	No	<a href="#">view FAIRsharing entry</a>

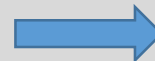
### [Data Repository Guidance – Nature](#)

Usmernenia + odkazy na odborové repozitáre

<https://www.nature.com/sdata/policies/repositories>

a všeobecné repozitáre

<https://www.nature.com/sdata/policies/repositories#general>



# Princípy FAIR a vedecké dáta

## DÁTOVÉ ČASOPISY

International Journal of  
**Epidemiology**

Issues

More Content ▼

Submit ▼

**Geoscience Data Journal**

Open Access

Co-Editors-in-Chief: Katherine Royse & Jian Peng

JOURNAL METRICS >

Online ISSN: 2049-6060

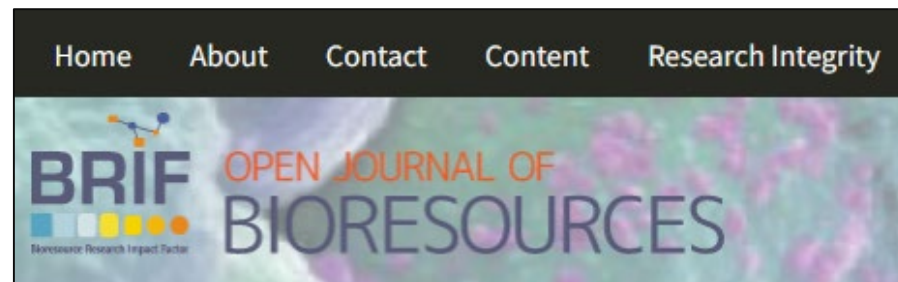
<https://academic.oup.com/ije>

<https://brill.com/view/journals/rdj/rdj-overview.xml>

<https://rmets.onlinelibrary.wiley.com/journal/20496060>

<https://openhumanitiesdata.metajnl.com/>

<https://openbioresources.metajnl.com/>



Viac na:

<https://www.wiki.ed.ac.uk/display/datashare/Sources+of+dataset+peer+review>

# Princípy FAIR a vedecké dáta

## TVORBA PLÁNU MANAŽMENTU DÁT NÁSTROJE



☐ Ponúka univerzita/inštitúcia ponúka nejaké softvérové nástroje a podporu na tvorbu DMP?



# Princípy FAIR a vedecké dáta

## Tvorba plánu manažmentu dát - NÁSTROJE

### ➤ OpenAIRE–Argos

<https://argos.openaire.eu/splash/>

<https://argos.openaire.eu/splash/about/how-it-works.html>

### ➤ DMPOnline (DCC, DigitalCurationCentre, UK)

<https://www.dcc.ac.uk/dmponline>

### ➤ DMPTool (USA univerzity)

<https://dmptool.org/>

### ➤ Data Stewardship Wizard

<https://ds-wizard.org/> ; videoprezentácia:

<https://www.youtube.com/watch?v=pvN0gPqBMzk>

## Príklady DMP dostupné na

➤ RIO - <https://riojournal.com/>

➤ Zoznam [zverejnených DMP](#) DMPOnline

➤ Zoznam [zverejnených DMP](#) DMPTool

➤ [Katalóg DMP](#) LIBER

➤ [Zoznam DCC](#)

argos

Plan and follow  
your data



openaire.eu/argos

- Get familiar with the **DMP Lifecycle** and excel in **Open and FAIR RDM planning**
- **Co-create DMPs** and manage workload
- **Publish and cite** DMPs as living documents
- **Configure** DMPs to tailored community needs
- **Link** DMPs to research outputs, EOSC services and the **OpenAIRE Research Graph**



DMPTool  
Build your Data Management Plan

Welcome to the DMPTool  
Create data management plans that meet institutional and funder requirements.

Get started

DMPTool by the Numbers

54,108 Users	51,335 Plans More	290 Participating Institutions More
-----------------	-------------------------	---

Top Templates

Digital Curation Centre  
Template USP - Baseado no DCC

DMPONLINE Home Public DMPs Funder requirements Help

Plan to make data work for you

Data Management Plans that meet institutional funder requirements.

Sign in Create account

\* Email

\* Password

Forgot password?  
 Remember email

Sign in

Sign in with your institutional credentials

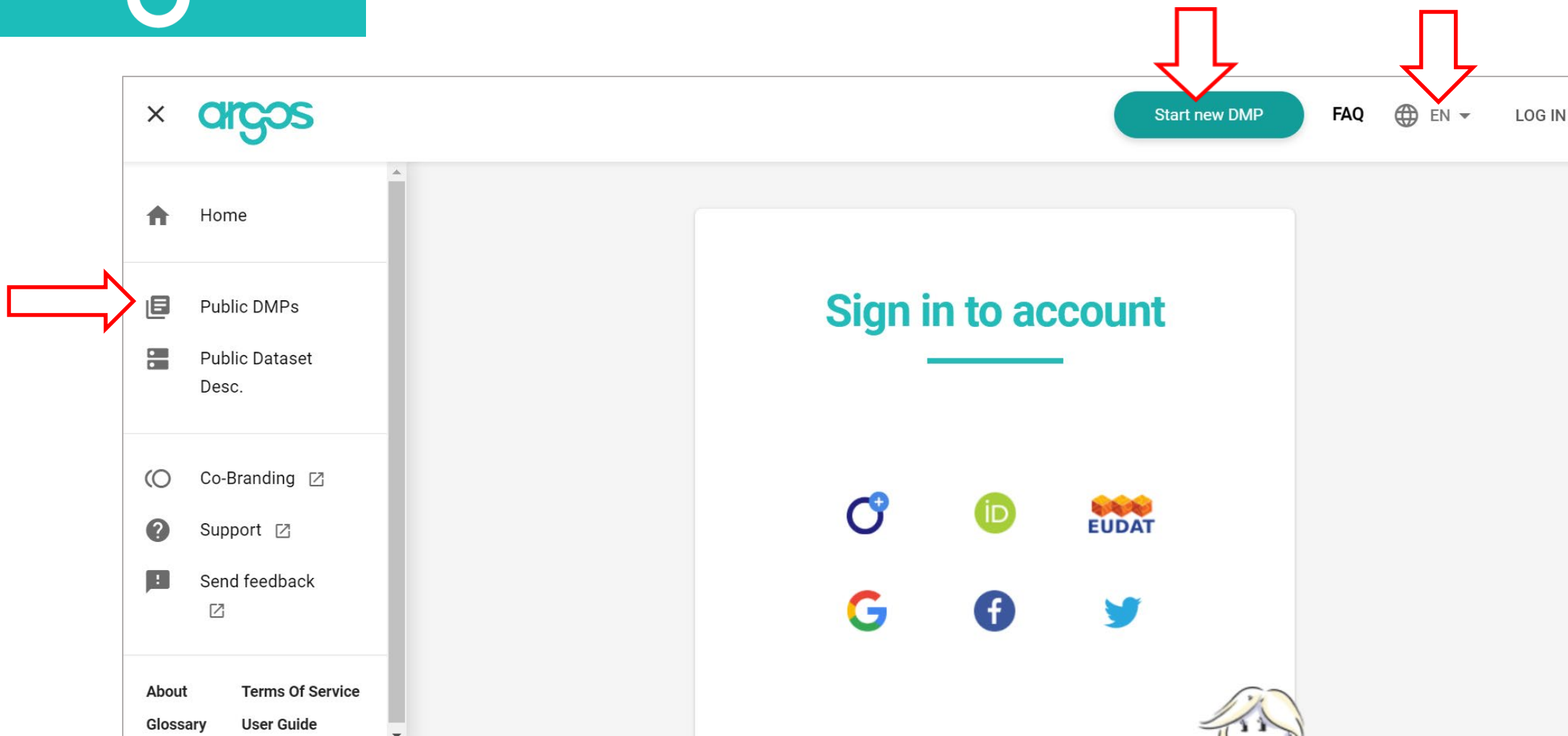
DMPOnline helps you to create, review, and share data management plans that meet institutional and funder requirements. It is provided by the Digital Curation Centre (DCC).

90 539 Users 303 Organisations 99 545 Plans 89 Countries

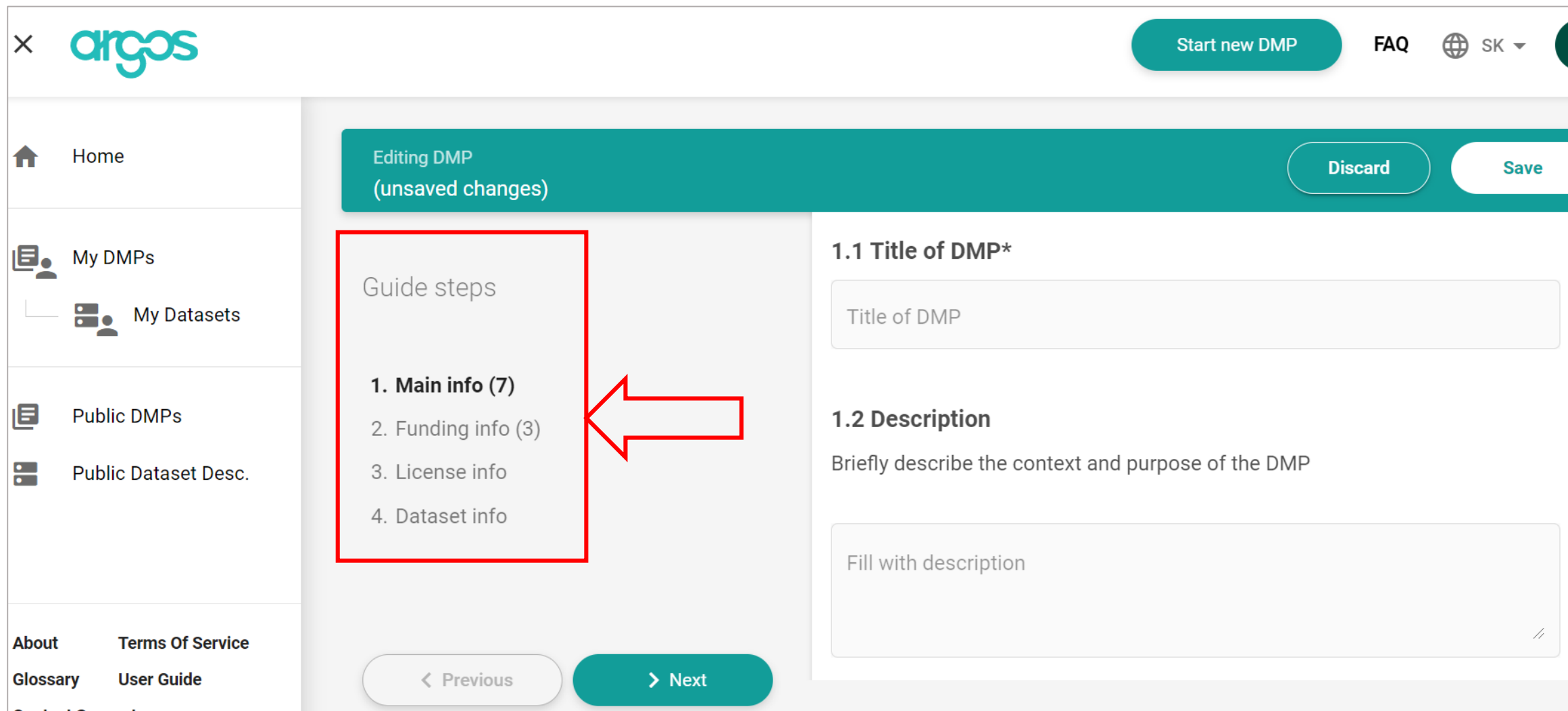
DSW Product Solutions Learn About Get Started

## Data Stewardship Wizard

Create, plan, collaborate, and bring your data management plans to life with a tool trusted by thousands of people worldwide — from data management pioneers, to international research institutes.



The screenshot shows the argos website interface. A red arrow points to the 'Start new DMP' button in the top right corner. Another red arrow points to the 'Public DMPs' link in the left sidebar. A third red arrow points to the 'Public Dataset Desc.' link in the left sidebar. The main content area displays a 'Sign in to account' prompt with social login options for OpenID, iD, EUDAT, Google, Facebook, and Twitter. The bottom right corner of the main content area features a small cartoon elephant logo.



Editing DMP (unsaved changes) Discard Save

Guide steps

1. Main info (7)
2. Funding info (3)
3. License info
4. Dataset info

1.1 Title of DMP\*

Title of DMP

1.2 Description

Briefly describe the context and purpose of the DMP

Fill with description

< Previous > Next

# Princípy FAIR a vedecké dáta

## Open Research Data and Data Management Plans

Information for ERC grantees  
by the ERC Scientific Council




European Research Council  
Executive Agency

Established by the European Commission



European Research Council (ERC)

ERC Data Management Plan  
Template



ERC OPEN RESEARCH  
DATA MANAGEMENT PLAN (DMP)

European Research Council  
Established by the European Commission

Project Acronym	Project Number

Template for the ERC Open Research Data Management Plan (DMP). The following sections should describe how you plan to make the project data Findable, Accessible, Interoperable and Reusable (FAIR). Each of the following five issues should be addressed with a level of detail appropriate to the project.

SUMMARY (dataset<sup>1</sup> reference and name; origin and expected size of the data generated/collected; data types and formats)

**Súhrn:** v 1 DMP môže byť viac súborov, dát, názov súboru dát, názov, odkaz, pôvod, očakávaný objem dát, typy a formáty

<sup>1</sup> Several datasets may be included into a single DMP.

Page 2 of 5

## ERC OPEN RESEARCH DATA MANAGEMENT PLAN (DMP)

1. MAKING DATA FINDABLE (dataset description: metadata, persistent and unique identifiers e.g., DOI)

**Vyhľadateľnosť:** opis súboru dát: metadáta, perzistentné a jedinečné identifikátory, napr. DOI

2. MAKING DATA OPENLY ACCESSIBLE (which data will be made openly available and if some datasets remain closed, the reasons for not giving access; where the data and associated metadata, documentation and code are deposited (repository?); how the data can be accessed (are relevant software tools/methods provided?))

**Prístupnosť:** ktoré dáta budú voľne prístupné, ktoré ostatnú zatvorené, prečo, uloženie dáta a metadát, kódov, ako sa dá dostať k dátam (softvéru, metódam)



# Princípy FAIR a vedecké dáta

## ERC OPEN RESEARCH DATA MANAGEMENT PLAN (DMP)

3. **MAKING DATA INTEROPERABLE** (which standard or field-specific data and metadata vocabularies and methods will be used)

**Interoperabilita:** ktoré štandardy a špecializované metadátové štandardy a slovníky sa budú používať

4. **INCREASE DATA RE-USE** (what data will remain re-usable and for how long, is embargo foreseen; how the data is licensed; data quality assurance procedures)

**Opakované použitie:** ktoré dáta bude možné použiť opakovane a ako dlho, embargo, licencie, zaistenie kvality dát



European Research Council  
Established by the European Commission

## ERC OPEN RESEARCH DATA MANAGEMENT PLAN (DMP)

5. **ALLOCATION OF RESOURCES and DATA SECURITY** (estimated costs for making the project data open access and potential value of long-term data preservation; procedures for data backup and recovery; transfer of sensitive data and secure storage in repositories for long term preservation and curation)

**Alokácia zdrojov a bezpečnosť dát:** odhad nákladov na prístupnosť dát v otvorenom režime a potenciálna hodnota dlhodobého uchovávaní dát; zálohovanie dát a obnova dát; transfer citlivých dát a bezpečné uchovávanie v repozitároch pri dlhodobom uchovávaní a spracovaní

**DISCLAIMER.** Please note that the ERC Data Management Plan is not a part of the Ethics Review. It is the responsibility of the Principal Investigator to inform the ERCEA Ethics Team of any ethics issues/concerns regarding the collection, processing, sharing and storage of data in relation to the project.

**Etické otázky**

# Princípy FAIR a vedecké dáta

## Plán manažmentu dát Humanitné vedy

Humanitné vedy  
(humanities)

Technické vedy  
(engineering and technology)

Lekárske vedy  
(medical and health sciences)

Spoločenské vedy  
(social sciences)

An example of a Data Management Plan for a project aimed at transliterating, translating and additional descriptions of the first volume of 18th-century manuscripts describing the several-year journey of Nathaneal Jacob Gerlach from Gdańsk around Europe.

The Project assumed that the work would be done with the use of a manuscript already digitized and made available in an electronic version. The project will result in ready-made transliterations, translations and other manuscript-related studies.



DATA MANAGEMENT PLAN – FORM FOR GDAŃSK UNIVERSITY OF TECHNOLOGY EMPLOYEES
<b>1. Data description and collection or re-use of existing data</b>
<b>1.1. How will new data be collected or produced and/or how will existing data be re-used?</b> Source (raw) data: Scans in jpeg format from the Pomeranian Digital Library available under the public domain (CC0 license). Unique URI for the source data: <a href="https://pbc.gda.pl/dlibra/publication/10113/edition/6021">https://pbc.gda.pl/dlibra/publication/10113/edition/6021</a> New data generated during the project: electronic documents with transliterations, translations, and descriptions of additional page elements such as seals, notes, etc. Each page of transliterations, translations will be corresponding to each scanned image in source data. The descriptions will include references - links to specific pages in the source data.
<b>1.2. What data (for example the kinds, formats, and volumes) will be collected or produced?</b> Source (raw) data: jpeg images collected in one folder. Each image has a unique name consisting of a prefix and a sequence number. Number of files: 363. Total size of files: 300MB New data generated during the project: documents in .docx and .txt and .pdf formats. Each *.docx document will be converted to open formats (.rtf, .odt) . Pdf files will contain images join with translations and transliterations. Estimated number of files: 726 txt files, 3 .docx files, 3 pdf files. Estimated size of files: 900MB
<b>2. Documentation and data quality</b>
<b>2.1. What metadata and documentation (for example methodology or data collection and way of organising data) will accompany data?</b> The data will be stored in an ordered structure of folders: separated folders for source data, translations, transliterations, and descriptions. Names of txt files will be corresponding to source files names. Each final document docx, and pdf will be described with metadata in the Dublin Core format. and will be share in Pomeranian Digital Library with a unique URI specified for the digital library system (dLibra). Based on all the files collected during the project, will be generated the dataset in zip format. The dataset will be made available through the open research data repository - MOST Wiedzy Open Research Data Catalogue witch support different metadata standards and schemas (eg. DataCite, DublinCore). The authors will be identified and authorized by ORCID number.

### Opis dát

Ako sa budú dáta zbierať/generovať, opakované používať?

Aké dáta – druh, formát, objem?

### Dokumentovanie a kvalita dát

Metadáta a dokumentácia – napr. spôsob zberu a organizácie dát

**Právne  
požiadavky**

**Zdieľanie dát,  
dlhodobé  
uchovávanie**

<p><b>2.2. What data quality control measures will be used?</b></p> <p>The specificity and scope of the project does not require the development and application of defined quality control procedures. In the individual stages of the project will participate People with knowledge and experience in a given field (eg in translation and transliteration of texts). Finally, the consistency of the data compiled with the source data will be verified</p>
<p><b>3. Storage and backup during the research process</b></p>
<p><b>3.1. How will data and metadata be stored and backed up during the research process?</b></p> <p>Data generated during the project will be stored on the computers of the team members and each day the data will be backed up in a dedicated external disk space based on disk arrays with the automatic backup functionality.</p>
<p><b>3.2. How will data security and protection of sensitive data be taken care of during the research?</b></p> <p>During the project, We will not collect process any personal and sensitive data.</p> <p>The access to the computers of the team members and to the external disk service will be restricted only to authorized users and will be protected by a strong password.</p>
<p><b>4. Legal requirements, codes of conduct</b></p>
<p><b>4.1. If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?</b></p> <p>During the project, We will not collect process any personal and sensitive data.</p>
<p><b>4.2. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?</b></p> <p>In the case of obtaining data or results that may constitute the matter of intellectual property, the rights to them are acquired by the person who obtained them.</p> <p>The data generated during the project will be made available under a CC-BY license.</p>
<p><b>5. Data sharing and long-term preservation</b></p>
<p><b>5.1. How and when will data be shared ? Are there possible restrictions to data sharing or embargo reasons?</b></p> <p>All data obtained as a result of the project will be made available immediately after the end of the project.</p> <p>The dataset with all orderly data will be made available through the open research data repository - MOST Wiedzy Open Research Data Catalogue that fulfills FAIR requirements.. Additionally, single studies and documents will be share in Pomeranian Digital Library.</p>
<p><b>5.2. How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?</b></p> <p>All data obtained as a result of the project will be qualified for long-term preservation.</p>

**Aké opatrenia na kontrolu  
kvality dát sa použijú?**

**Uchovávanie a zálohovanie počas výskumu**

<p>Both The MOST Wiedzy Open Research Data Catalogue, and The Pomeranian Digital Library guarantee that the data will be available for a minimum of 10 years.</p>
<p><b>5.3. What methods or software tools will be needed to access and use the data?</b></p> <p>The open file format allows using of commercial software as well the free, standard, and open software (e.g. txt reader, PDF viewer, open office). There are no requirements for the operating system.</p>
<p><b>5.4. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?</b></p> <p>The datasets made available through the MOST Wiedzy Open Research Data Catalogue will have the DOI assigned. The MOST Wiedzy Open Research Data Catalogue also supporting assigning DOI for each version of the existing dataset.</p> <p>The Pomeranian Digital Library guarantee a unique URL for each resource.</p>
<p><b>6. Data management responsibilities and resources <b>Manažment dát</b></b></p>
<p><b>6.1. Who (for example role, position, and institution) will be responsible for data mangement (i.e the data steward)?</b></p> <p>The Project PI (Imię Nazwisko) will be responsible for the overall data quality.</p> <p>The Open Science Competence Center established by PG (pg.edu.pl/openscience) will be responsible for support in the preparation and maintenance of the Data Management Plan and for support in the process of preparing and sharing datasets in the MOST Wiedzy Open Research Data Catalog.</p>
<p><b>6.2. What resources (for example financial and time) will be dedicated to data management and ensuring the data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?</b></p> <p>There will be no additional financial costs of data management involved in the project.</p>



# Princípy FAIR a vedecké dáta



Funding & tender opportunities  
Single Electronic Data Interchange Area (SEDIA)

HOME SEARCH FUNDING & TENDERS HOW TO PARTICIPATE PROJECTS & RESULTS WORK AS AN EXPERT SUPPORT

## Reference Documents

### Programming period

2021-2027

### Filter by programme:

Horizon Europe (HORIZON)

Clear filter

Declaration of Honour (DoH) annual updates

Declaration of honour (DoH) prizes

### Project reporting templates

Periodic report (HE)

Certificate on the financial statements (CFS)

Time declaration

Data management plan (HE)

Progress report (HE MSCA)

Contractor details and project abstracts (HE PCP PPI)

End of phase/project results and conclusions (HE PCP PPI)

Report on cumulative expenditure

<https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/how-to-participate/reference-documents?selectedProgrammePeriod=2021-2027&selectedProgramme=HORIZON>

# Princípy FAIR a vedecké dáta

[Projekt FAIR-IMPACT](#) - [aktuálna výzva, otvorená do 1.](#)

[novembra 2023](#) - záujemcovia sa môžu uchádzať

o podporu zo strany programových mentorov, ktorí úspešných uchádzačov budú školiť v zavádzaní postupov podporujúcich vznik FAIR dát.

- zameraná na špecializované poradenstvo a individuálnu podporu
- vhodná pre začínajúcich alebo menej pokročilých žiadateľov v oblasti zavádzania princípov FAIR dát do praxe
- projekt FAIR-IMPACT poskytne úspešným uchádzačom špecializované poradenstvo a individuálnu podporu

**Do výzvy sa môžu zapojiť:**

- Iniciatívy na národnej úrovni,
- Organizácie vykonávajúce výskum,
- Prevádzkovatelia úložísk a poskytovatelia dát.

eosc | FAIR-IMPACT  
Expanding FAIR solutions across EOSC



**2nd Open Call for Support**

**APPLY till  
1st November 17:00 CET**

**National Level Initiatives  
Research Performing Organisations  
Repositories and Data Service Providers**

FAIR Implementation Team | FAIR-IMPACT.eu | Funded by the European Union

**Termín na podanie žiadosti je do 1. novembra 2023.**

**Ak sa chcete o výzve dozvedieť viac, či máte otázky, registrujte sa na seminár, ktorý sa uskutoční 18. októbra v čase od 15:00-16:00.**

**[ERA Portál Slovensko \(eraportal.sk\)](https://eraportal.sk/aktuality/projekt-fair-impact-otvoril-druhe-kolo-podpornej-vyzvy/?era=1): <https://eraportal.sk/aktuality/projekt-fair-impact-otvoril-druhe-kolo-podpornej-vyzvy/?era=1>**

## ZHRNUTIE

- Čo je manažment vedeckých dát
- Aké činnosti alebo faktory treba zväžiť pri manažmente vedeckých dát
- Čo je DMP, prečo ho treba robiť, čo je jeho obsahom
- Nástroje na tvorbu DMP
- Ukážky DMP

# Princípy FAIR a vedecké dáta

## Cvičenie

**1) Pozrieť si a porovnať DMP z rôznych vedných odborov a rôznych DMP nástrojov** – čo všetko obsahujú, z toho, čo sme si povedali, napr.: aké dáta zbierajú, v akých formátoch sa budú dáta ukladať...

Na stránke: <https://otvorenaveda.cvtisr.sk/plan-manazmentu-dat/>

Technické vedy: [Design of a low-cost 3D head scanner](#) – DMPONLINE

Spoločenské vedy: [Worlding Public Cultures: The Arts and Social Innovation](#) – DMPONLINE

Humanitné vedy: [The Italian saxophone – a new instrument for an old musical tradition](#) - DMPOnline

Lekárske vedy: [Headache and sleep disturbances after traumatic brain injury – a sham controlled repetitive transcranial](#)  
– DMPTuuli

**2) Pozrieť si ARGOS – verejné DMP a verejné dátové súbory**

**OpenAIRE–Argos:**

- <https://argos.openaire.eu/splash/>
- <https://argos.openaire.eu/splash/about/how-it-works.html>

# Princípy FAIR a vedecké dáta

## Manažment vedeckých dát, právne aspekty použitia vedeckých dát, licencie: zdroje, odkazy, ďalší študijný materiál

### Zdroje

- <https://pitt.libguides.com/managedata>
- <https://ukdataservice.ac.uk/learning-hub/research-data-management/>
- <https://dmeg.CESSDA.eu/Data-Management-Expert-Guide>
- <https://dmeg.CESSDA.eu/Data-Management-Expert-Guide/1.-Plan/Adapt-your-DMP-Part-1>
- <https://library.soton.ac.uk/researchdata/storage#s-lib-ctab-14781697-4>
- [https://guides.nyu.edu/data\\_management/osf](https://guides.nyu.edu/data_management/osf)
- <https://www.cos.io/services/training>
- <https://www.openaire.eu/research-data-management-handbook>
- <https://opencscience.upol.cz/otvarena-data/repozitare-a-datove-casopisy/>
- Daniela Ukropová. Metadáta v slovenských knižniciach. Itlib. <https://itlib.cvtisr.sk/%c4%8cl%c3%a1nky/clanek2107/>
- Archivácia výskumných dát.+ Manažment výskumných dát. In: Zuzana Stožická, Gabriela Fišová, Silvia Sofianos, & Simona Hudecová. (2023, March 28). Webináre pre STU (február-marec 2023). Zenodo. <https://doi.org/10.5281/zenodo.7785206>
- Ball, A. (2014). 'How to License Research Data'. DCC How-to Guides. Edinburgh: Digital Curation Centre. Available online: Available online: </resources/how-guides>
- <https://www.dcc.ac.uk/guidance/how-guides/license-research-data>
- Adamová, Zuzana: Otvorené dáta. Prezentácia pre CVTI, 2023.
- [Dánsky online kurz](#) o dôležitosti dobrého manažmentu dát v anglickom jazyku. Danish National Forum for Data Management: doi: 10.11581/dtu:00000048, doi: 10.11581/dtu:00000049, doi: 10.11581/dtu:00000050
- <https://library.soton.ac.uk/researchdata/storage#s-lib-ctab-14781697-4>

Ďalšie použité zdroje aj s hypertextovými odkazmi sú uvedené priamo na slajdoch.



# Princípy FAIR a vedecké dáta

## Manažment vedeckých dát, právne aspekty použitia vedeckých dát, licencie: zdroje, odkazy, ďalší študijný materiál

### Odkazy na ďalší študijný materiál:

- <https://openscience.cuni.cz/OSCI-63.html>
- <https://ukdataservice.ac.uk/learning-hub/research-data-management/plan-to-share/resources-for-research-data-management/>
- <https://dmeg.CESSDA.eu/>  
CESSDA ERIC (CESSDA, Consortium of European Social Science Data Archives; ERIC, European Research Infrastructure Consortium): provides large-scale, integrated and sustainable data services to the social sciences. It brings together social science data archives across Europe, with the aim of promoting the results of social science research and supporting national and international research and cooperation.
- 4 Principles to Follow about Document Version Control and how to maintain it without going mad. Dostupné na <https://medium.com/agileinsider/4-principles-to-follow-about-document-version-control-d9116b21f05e>
- 3 dôvody, prečo začať s verziovaním. Dostupné na <https://www.websupport.sk/blog/2013/02/3-dovody-preco-zacat-s-verziovanim/>
- <https://itlib.cvtisr.sk/wp-content/uploads/2022/12/Pavcik.pdf>
- LSE Library (2008) Versions toolkit for authors, researchers and repository staff. . London School of Economics and Political Science, London. Dostupné na <http://eprints.lse.ac.uk/64948/>
- [https://e-justice.europa.eu/119/SK/glossaries\\_and\\_translations](https://e-justice.europa.eu/119/SK/glossaries_and_translations)
- Ludwig, Jess: Digital Humanities, Data Literacy Skills and AI: Understanding the Way Things Work, 10. august 2023. Dostupné na : [https://scholarlykitchen.sspnet.org/2023/08/10/guest-post-digital-humanities-data-literacy-skills-and-ai-understanding-the-way-things-work/?informz=1&nbd=8cb2f73c-376d-47f9-9da5-037a94bc9ba1&nbd\\_source=informz](https://scholarlykitchen.sspnet.org/2023/08/10/guest-post-digital-humanities-data-literacy-skills-and-ai-understanding-the-way-things-work/?informz=1&nbd=8cb2f73c-376d-47f9-9da5-037a94bc9ba1&nbd_source=informz)
- <https://www.cms.hu-berlin.de/en/dl-en/dataman-en/support/material>
- <https://mantra.ed.ac.uk/>
- <https://otvorenaveda.cvtisr.sk/plan-manazmentu-dat/>
- Príklady publikovaných DMP: publikačná platforma RIO - <https://riojournal.com/>, DMPOnline [zverejnené plány](#), verejné DMP na [DMPTool](#), [katalog LIBER](#), [zoznam](#) na stránkach Digital Curation Centre radený.
- <https://openscience.cuni.cz/OSCI-63.html>
- Copyright a verejné licencie: webinár - <https://otvorenaveda.cvtisr.sk/copyright-a-verejne-licencie/>
- Preklad vzdelávacieho materiálu organizácie CC. Dostupný na <https://zenodo.org/record/8344376> ; <https://otvorenaveda.cvtisr.sk/publikacie/>
- <https://resources.data.gov/>

**Zdroje obrázkov:** sú uvedené priamo na slajdoch s príslušnými obrázkami, logá sú printscreeny z webových stránok.

# Ďakujem za pozornosť.

**Kontakt:**

[gabriela.fisova@cvtisr.sk](mailto:gabriela.fisova@cvtisr.sk)

[otvorenaveda@cvtisr.sk](mailto:otvorenaveda@cvtisr.sk)

<https://otvorenaveda.cvtisr.sk/>



**Centrum vedecko-technických informácií**

Odbor podpory otvorenej vedy a výskumu

Lamačská cesta 8/A

840 05 Bratislava

