

10100100110001101110100 0 1 0 1 1 01 1 1 01 001 001 0 0 1 1 0001101 110100010 110110010

Archivácia a manažment výskumných dát

kurz

Elektronické informačné zdroje pre vedu – Publikačný poradca

Zuzana Stožická

Centrum vedecko-technických informácií SR

NISPEZ^{IV}

Investícia do Vašej budúcnosti



Táto prezentácia je šírená pod licenciou
[Creative Commons 4.0 Attribution](https://creativecommons.org/licenses/by/4.0/)



EURÓPSKA ÚNIA
Európske štrukturálne a investičné fondy
OP Integrovaná infraštruktúra 2014 – 2020



MINISTERSTVO
DOPRAVY A VÝSTAVBY
SLOVENSKEJ REPUBLIKY

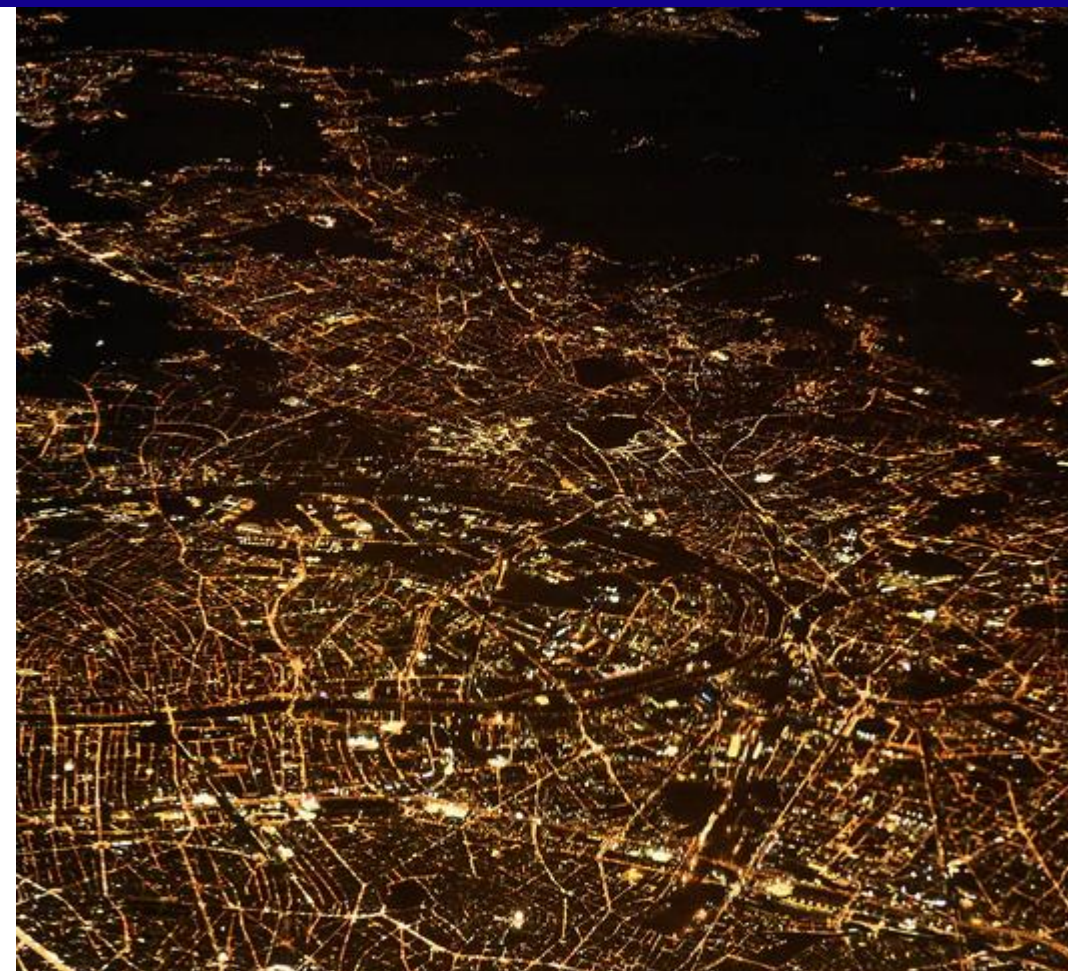


MINISTERSTVO
ŠKOLSTVA, VEDY,
VÝSKUMU A ŠPORTU
SLOVENSKEJ REPUBLIKY



10100100110001101110100 0 1 0 1 1 01 1 1

- Archivácia – pamäť vedeckého záznamu
- Čo sa stane s obsahom časopisu, keď vydavateľ skrachuje?
- Repozitáre
- Manažment výskumných dát – iným slovom zodpovednosť k dátam (a kolegom, ktorí by ich mohli potrebovať)
- Plán manažmentu dát (Data Management Plan, DMP)
- Nástroje, ktoré nám môžu pomôcť



- Krátky, ale informatívny dánsky videokurz o význame dátového manažmentu: <https://vidensportal.deic.dk/RDMelearn>, Danish National Forum for Data Management: doi: 10.11581/dtu:00000048, doi: 10.11581/dtu:00000049, doi: 10.11581/dtu:00000050
- MANTRA Research Data Management training (kurz Edinburskej univerzity): <https://mantra.edina.ac.uk/>
- Po Slovensky: Webinár CVTI SR z mája 2020 [Manažment vedeckých dát a príklad dobrej praxe na Slovensku: digitálny repozitár Katedry mediamatiky a kultúrneho dedičstva na Žilinskej univerzite](https://archiv.nti.sk/archive.php?vid=5841999IUEMF0241##videoplayer) <https://archiv.nti.sk/archive.php?vid=5841999IUEMF0241##videoplayer>

Archivácia – pamäť vedeckého záznamu

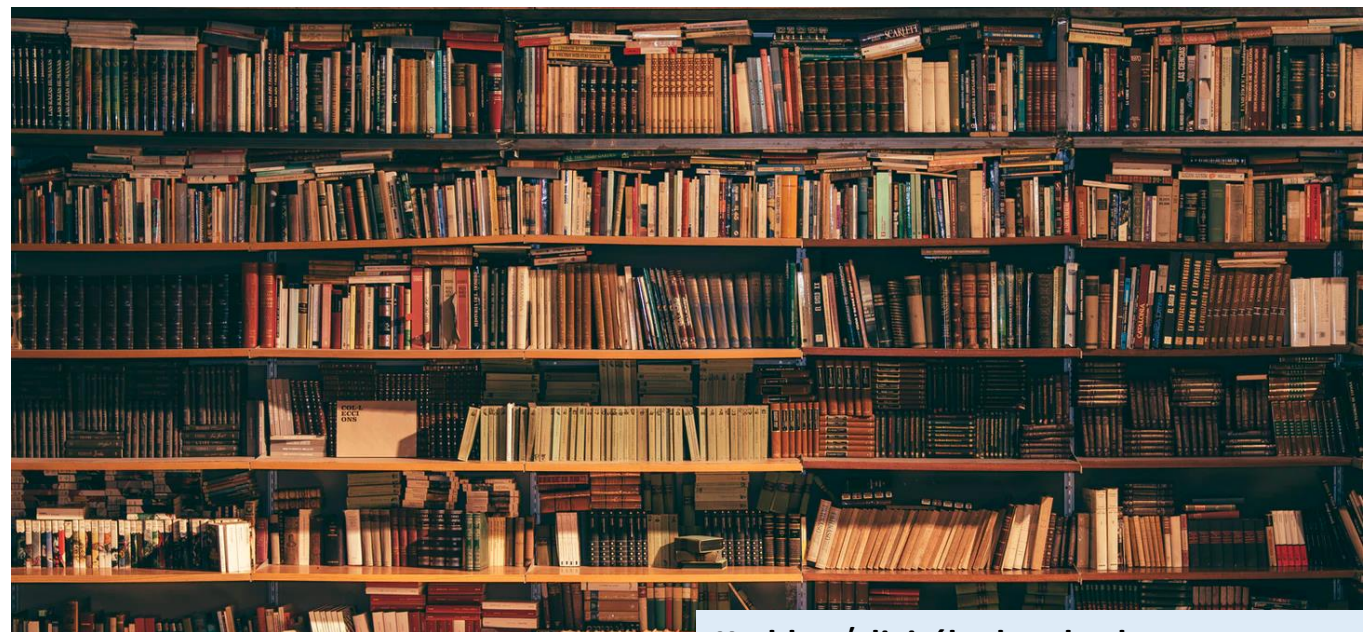
10100100110001101110100 0 1 0 1 1 01 1 1

- Kontinuita vedeckého záznamu
- Vedecké články sú vyhľadávané a citované, (= potrebné), aj storočia po svojom vydaní

Samozrejme nie všetky, ale nikto nevie vopred určiť, ktoré budú „tie šťastné“.

- Vydavateľstvá vedeckých časopisov, hlavne tie menšie, majú často krátky život
- Čo sa deje s obsahom po zániku časopisu/vydavateľa?
- Kedysi: papierové exempláre časopisu ostávali v knižnici, ktorá ich archivovala
- Dnes: Error 404...

... alebo...



Krehkosť digitálneho obsahu

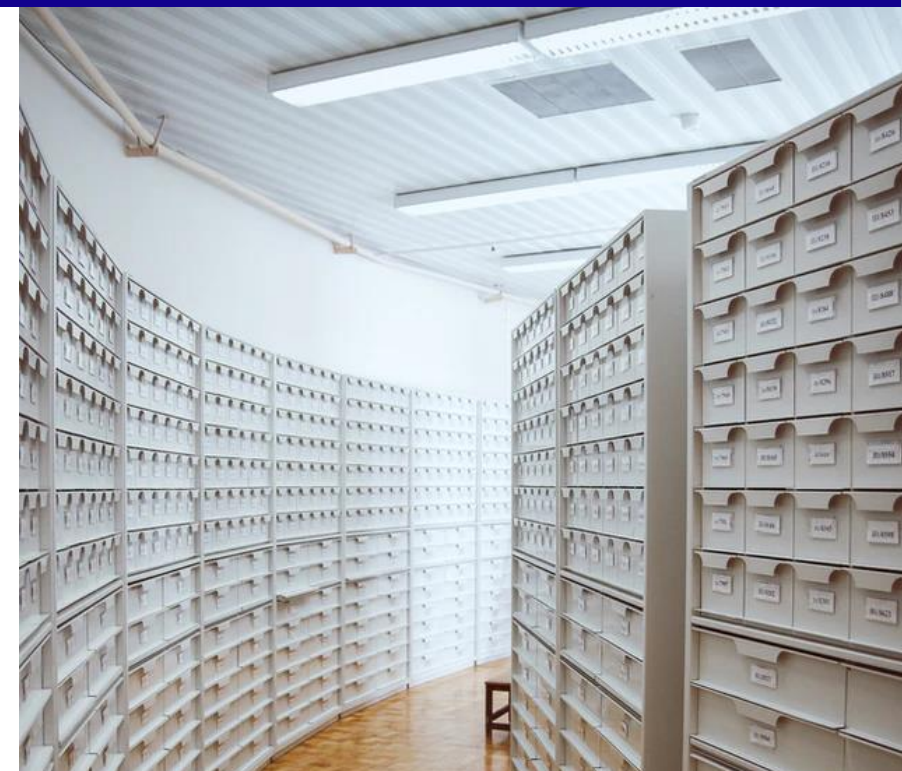
Papierová kniha vydrží storočia. Čo je viac ako napr. CD-nosič.

Na druhej strane, časopis či noviny z menej kvalitného (kyslého) papiera sa po niekoľkých desaťročiach rozpadnú bez náhrady – ak ich obsah chceme zachovať, musíme digitalizovať.

Kam s tým všetkým digitálnym obsahom?

Seriózni vydavatelia sa zapájajú do archivačných iniciatív

- **Portico** (<https://www.portico.org/>) – tmavý archív e-časopisov, kníh a digitálnych kolekcií, podporovaný komunitou vydavateľov a knižníc (sprístupňuje obsah až v okamihu, keď prestane byť klasicky dostupný)
- **JSTOR** (Journal storage, <https://www.jstor.org/>) – od r. 1995, začal ako knižnica archivujúca digitalizované staršie čísla papierových časopisov. 6 % obsahu (spadajúceho do verejnej domény) je voľne dostupných, zvyšok na báze predplatného.
- Ďalšie: **e-Depot** (Holandsko), **OCLC** (British Library), **HathiTrust** (akademické knižnice Big Ten Academic Alliance)...
- ... na Slovensku: projekt DIKDA (SNK)

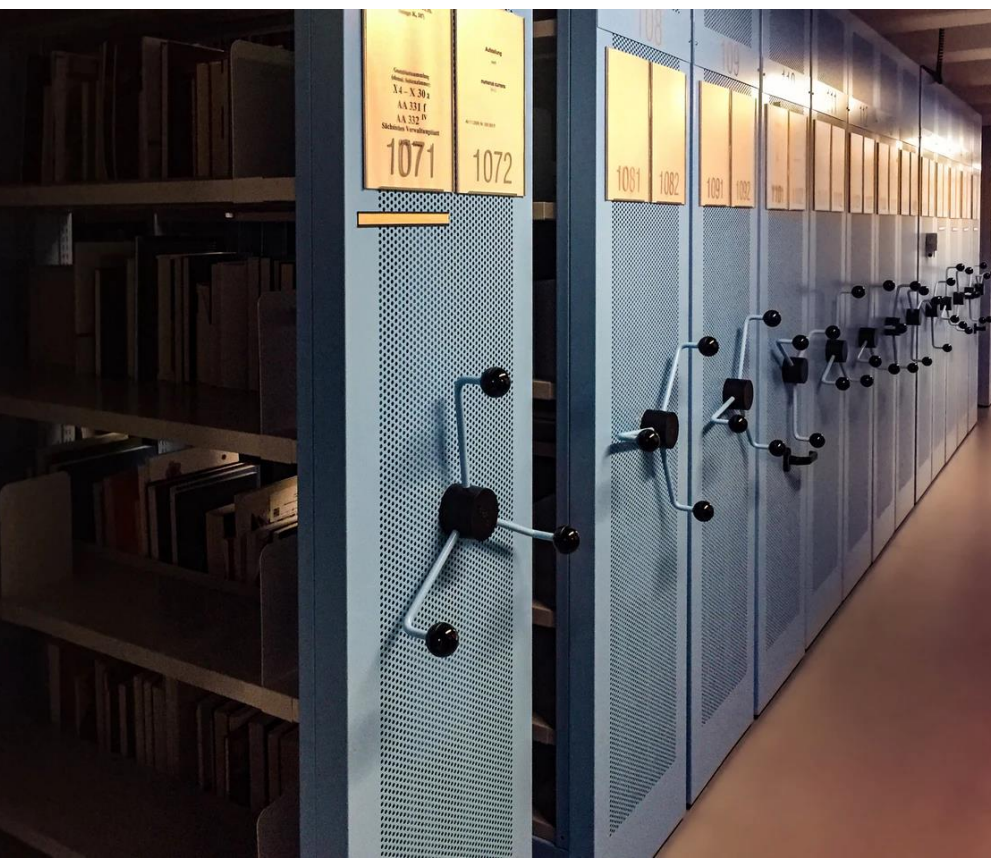


Archivácia – proces podporujúci dlhodobé uchovávanie a ochranu dát/dokumentov a metód/postupov ich čítania a interpretácie

Ochrana (preservation) – súbor postupov a činností zameraných na uchovanie dát po určitú dobu a/alebo zhromažďovanie dát pre budúce použitie, pričom súčasťou je archivácia a/alebo odosielanie dát do repozitára (úložiska).



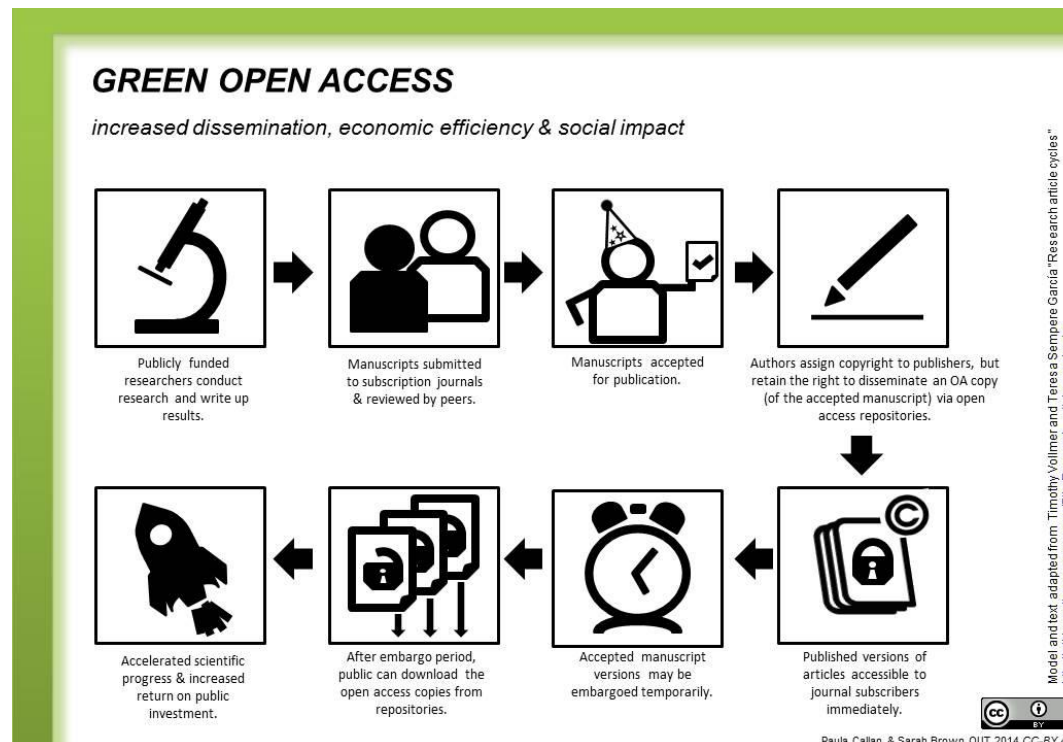
- rsity, od r. 1998. Systém na báze softvéru
i voči sebe navzájom, aby odhalil náhodne
ol vydaný. Keby bola na internete len jedna kópia



Autoarchivácia (autor si archivuje svoje dielo sám) v repozitári – cesta nezávislá od vydavateľov

Repozitár – štruktúrované digitálne úložisko, systém pre prevádzku súboru služieb, ktoré sa zaoberajú získavaním, správou a prístupňovaním digitálneho obsahu

- „**digital repository**“ – termín zaviedli 1995 R. Kahn, R. Wilensky (teoretické princípy: Kahn-Wilensky Framework, KWF. Definujú komponenty, architektúru systému oddeľujú od uchovávaného obsahu)



Repozitár – štruktúrované digitálne úložisko, systém pre prevádzku súboru služieb, ktoré sa zaoberajú získavaním, správou a sprístupňovaním digitálneho obsahu

- **Inštitucionálne** repozitáre – služby orientované primárne dovnútra inštitúcie s cieľom mať pod kontrolou publikačné (dátové a iné) výstupy a poskytovať zamestnancom potrebné informačné zdroje, ale aj navonok: reprezentovať inštitúciu kvalitným obsahom
 - Otvorené **predmetové** repozitáre – zameriavajú sa na konkrétnu vednú oblasť (napr. **ArXiv**, od r. 1991)
 - **Národné** repozitáre – obsahujú publikačné výstupy krajiny
 - **Dátové, softvérové, multiodborové** (napr. Zenodo)...
- 



INŠTITUCIONÁLNY REPOZITÁR SAV

ČINNOSTI A OHLASOV

IÝCH VÝSTUPOV ▾

Y

Repozitár SAV je inštitucionálne digitálne úložisko pre dlhodobé uloženie a sprístupnenie publikácií ([zamestnaneckých diel](#)) vytvorených zamestnancami niektorej z organizácií Slovenskej akadémie vied, budované v súlade s ustanoveniami [Smernice o inštitucionálnom repozitári SAV](#). Repozitár je registrovaný v [Directory of Open Access Repositories \(OpenDOAR\)](#). (Auto)archivácia publikácií v repozitári je uplatnením [zelenej cesty Open Access](#) k výsledkom výskumu SAV realizovaných z verejných zdrojov. Repozitár SAV je prístupný cez [Katalóg publikáčnej činnosti](#) na stránke Ústrednej knižnice SAV, ktorý obsahuje metadáta (bibliografické záznamy) k publikáciám. Prostredníctvom katalógu môže:

- **používateľ** (verejnosť) vyhľadávať a **používať** publikácie v súlade so **Všeobecnými podmienkami používania repozitára SAV** a v súlade s ustanoveniami **Autorského zákona**,
- **autor** vkladá svoje publikácie podľa schválenej **metodiky pre autora** a určovať rozsah použitia autorského diela používateľom (prístupupovanie). Videonávod **Ako nahráť prácu do** **Inštitucionálneho repozitára SAV** nájdete **tu**.

Repozitár

- Vnútri každého repozitára je špecializovaný systém súborov
- Fond tvoria **digitálne objekty**

Digitálny objekt obsahuje:

- Jedinečný identifikátor (**handle**)
- Vlastnosti, metadáta (properties)
- Obsah v bitoch (content)
- Digitálny podpis (signature), zabezpečenie autenticity obsahu
- Záznamy o použití, vyhľadávaní objektu (transaction log)

Otvorené softvéry pre repozitáre:

Dspace, Eprints, Fedora, Invenio...

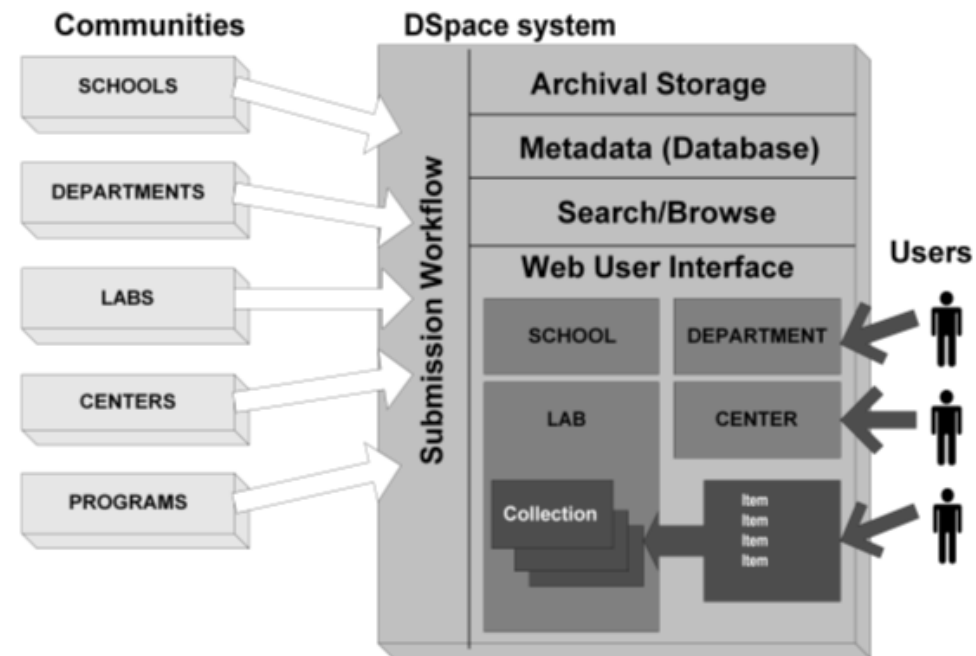
Dôveryhodnosť repozitárov:

PLATTER (Planning Tool for Trusted Electronic Repositories)

DRAMBORA (Digital Repository Audit Method Base on Risk Assessment)

Funkcie repozitára:

- Spracovanie dokumentov
- Tvorba metadát
- Uloženie a archivácia digitálneho obsahu
- Zabezpečenie integrity a autenticity
- Sprístupnenie dokumentov



Repozitár by mal spĺňať 10 podmienok:

OASIS: nezisková organizácia, vznikla v r. 1993 ako Standard Generalized Markup Language (SGML) Open

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 sprístupňovania a kontextu použitia.
8. Musí byť schopný zabezpečiť splnenie požiadaviek na **sprístupňovanie** digitálnych objektov konkrétnej komunity.
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.

Ako nájsť vhodný repozitár?

Medzinárodné registre repozitárov

- Registry of Open Access Repositories, ROAR (<http://roar.eprints.org/>)
- Open Directory of Open Access Repositories, OpenDOAR (<https://v2.sherpa.ac.uk/opensoar/>)
- Registry of Research Data Repositories (<https://www.re3data.org/>)

ROARMAP: Registry of Open Access Repository Mandates (politiky,
<http://roarmap.eprints.org/>)



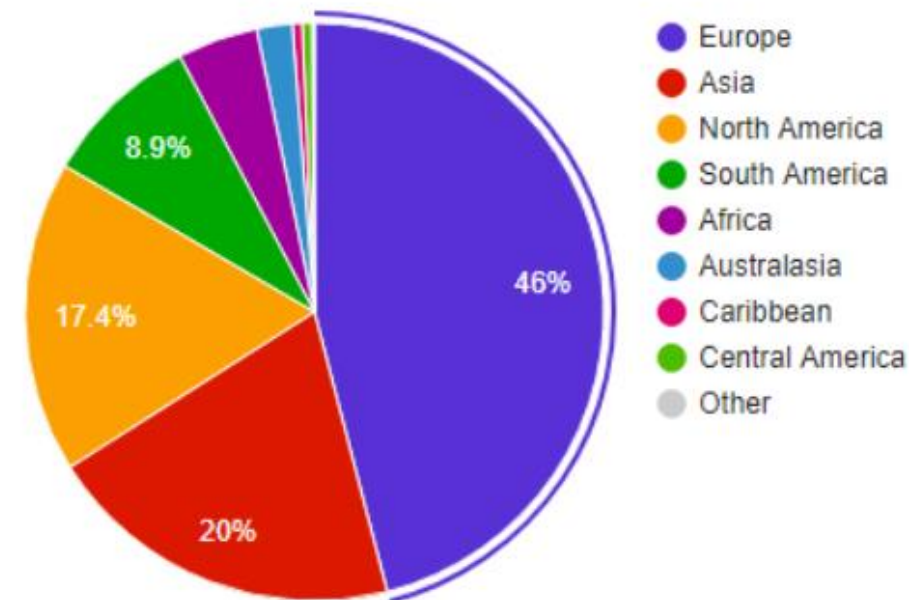
Kde si overiť, ktorú verziu svojho článku smiem dať do repozitára?

Sherpa Romeo: <https://v2.sherpa.ac.uk/romeo/>



OpenDOAR

Proportion of Repositories by Continent - Worldwide



Total = 3519 repositories

OpenDOAR - 19-Jun-2018

Vy a dáta?

- Venujete na Vašom pracovisku dátam takú pozornosť, akú si zaslúžia?
- Máte jednotnú politiku týkajúcu sa dát?
- Aký typ dát vytvárate?
- Ako ich označujete?
- Kde ich uchovávate?
- Kto s nimi smie pracovať?
- Zdieľate dáta so zvyškom vedeckej komunity, alebo ich hromadíte len v rámci pracoviska („data hoarding“)?

Dátová explózia

90 % svetových dát sa vyprodukovalo za posledné 2 roky – vďaka novým nástrojom a metódam vo vede (ako urýchľovače, teleskopy skúmajúce vesmír, družice skúmajúce Zem, sekvenovanie génov, magnetická rezonancia a pod. v biomedicíne, ale aj v iných vedách, napr. archeológii, 3D skenovanie a modelovanie, nástup počítačového spracovania vo všetkom, od environmentalistiky po lingvistiku...)

80 % z dát, ktoré pribúdajú, nie sú štruktúrované – sú užitočné len pre tých, ktorí ich získali, ale pri dobrom spracovaní a manažovaní by mohli poslúžiť aj mnohým ďalším.

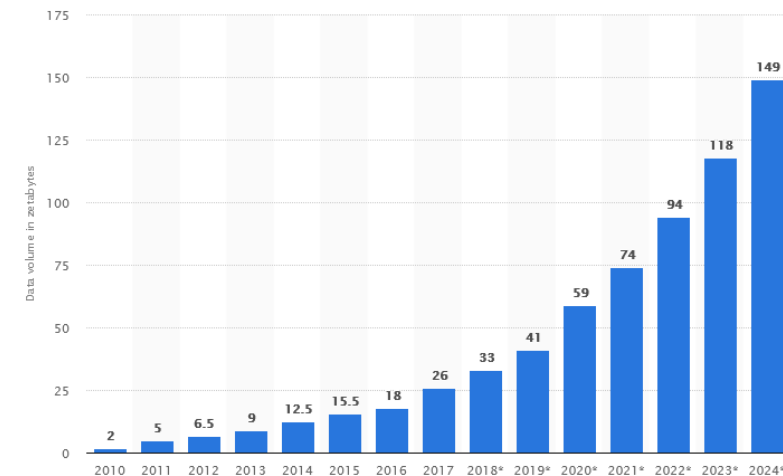
Dáta však musia byť:

- kvalitné,
- dobre opísané,
- dobre štruktúrované,
- dobre uskladnené a prístupné

(v rámci možností – napr. po anonymizácii osobných údajov, vyriešení licencií a pod.)

Komu patria výskumné dáta?

- Spravidla výskumnej inštitúcie
- Význam inštitucionálnych politík
- Dáta nepodliehajú copyrightu
- Na ich využitie možno udeliť licenciu



Prečo sa starať o dáta?

- Článok vo vedeckom časopise nepovie všetko
- Podkladové dáta: „Pohľad pod kapotu výskumu.“
- Objem a rôznorodosť produkovaných dát rastie
- Je v záujme vedcov aj výskumných inštitúcií, aby mali v dátach poriadok
- Dáta ako služba a reprezentácia inštitúcie
- Záujem o dáta rastie (otvorená veda, spolupráce, interdisciplinárnosť...)
- Aj staršie dáta sú zaujímavé, vznikajú nové cesty ich využitia
- Tvorcovia dát nemusia ani tušiť, na aký účel môžu byť ich dáta v budúcnosti pre niekoho užitočné

Pre inšpiráciu k inštitucionálnej politike ohľadom dát:

<https://www.jisc.ac.uk/guides/research-data-management>

Koho to zaujíma?

- Financovateľov výskumu (H2020, Horizont Europe)
- Výskumné inštitúcie
- Vydavateľov
- Ostatných vedcov

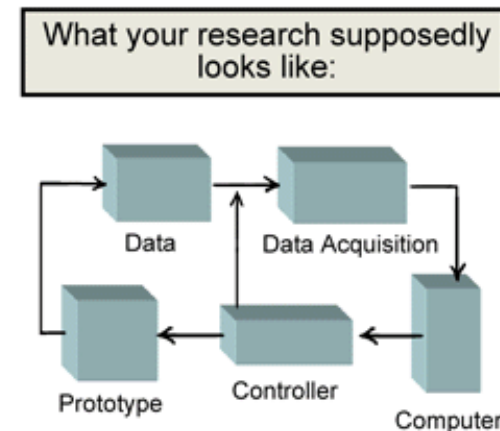


Figure 1. Experimental Diagram

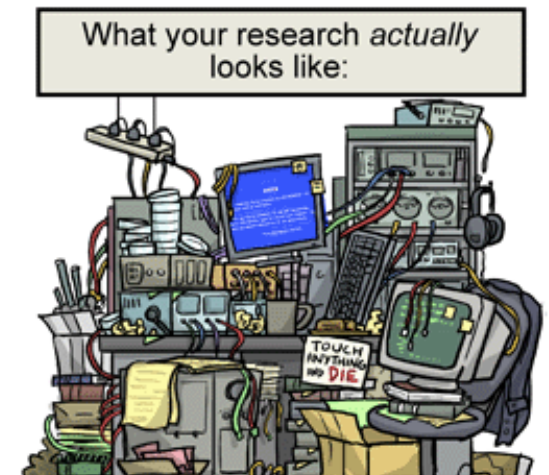
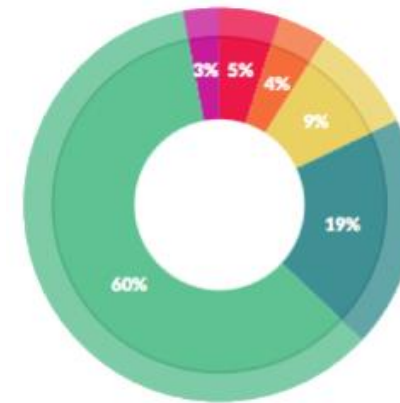


Figure 2. Experimental Mess

Manažment výskumných dát patrí k dobrej praxi

- Otázka kontinuity v laboratóriách (Vyznáte sa v dátach svojich kolegov/štoventov? Vyznali by sa kolegovia v tých vašich? Aj po desiatich rokoch?)
- „skryté znalosti“ – informácie, ktoré „predsa všetci v labáku vedia“, ale nikto ich nenapíše/môžu sa ľahko stratiť/dáta bez nich môžu zmiast čitateľa.
- Dobrá organizácia dát redukuje námahu potrebnú na ich vyhľadanie a pochopenie systému, čím zvyšuje podiel času, ktorý môžete stráviť štúdiom vašej témy.
- Otázka bezpečnosti dát („Stratil sa USB kľúč. Dva roky práce v ťahu!“)
- Vopred sa rozhodnete, čo, s kým a kedy budete zdieľať (v rámci laboratória a so spolupracovníkmi zdieľame dáta hneď po ich získaní, so svetom zvyčajne až po tom, ako dáta poslúžili primárnemu účelu).
- Keď už zdieľať, tak v podobe kvalitných dát (vyhľadateľnosť, usporiadanosť, otvorené formáty...). Dáta by mali byť využiteľné a **zrozumiteľné bez interakcie s autorom**
- Zdieľanie dát zvyšuje **reprodukovateľnosť**, **dôveryhodnosť**, aj **finančnú efektívnosť výskumu** (v podobe možnosti opätovného využitia dát).



What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets: 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%

Typy výskumných dát

- **Dáta plynúce z pozorovaní:** dáta získané v reálnom čase, zvyčajne nenahraditeľné
- **Experimentálne dáta:** laboratórne výskumy (génové sekvencie, chromatogramy...)
- **Simulačné dáta:** testovacie modely (ekonomické, klimatické...)
- **Skompilované dáta:** text mining, data mining...
- **Tzv. referenčné dáta:** zbierka menších datasetov (napr. databanky génových sekvencií)

Formy výskumných dát:

- Text, tabuľky
- Laboratórne protokoly,
- Archeologické správy,
- Dotazníky, prepisy,
- Audio/video pásky,
- Obsah databáz,
- Modely, algoritmy, programy,
- Metodológie a pracovné postupy,
- Artefakty...

Korešpondencia...

Grantové žiadosti...

Technické správy...

Výskumný cyklus a dáta

10100100110001101110100 0 1 0 1 1 01

Uvážiť požiadavky investora/grantovej agentúry, vypracovať DMP

- Jasné citácie
- Zaisťiť akumuláciu kreditu (uznania)

hypotéza

- Jasné práva na používanie
- Zaisťiť poctivé citovanie zdrojov

zber dát

?
otvorené IT rozhrania
([GitHub](#))

Analýza a
spracovanie

- Využívať infraštruktúru dostupných služieb
- Pripojiť k výsledkom trvalé identifikátory (**DOI...**)
- Pripojiť k výsledkom štandardizované **metadáta**
- Uverejniť metadáta s použitím **otvorenej licencie?**

uloženie dát a
výsledkov

dlhodobé
uchovávanie

- Používať služby, ktoré zabezpečujú uchovanie a integritu materiálov

uverejnenie
a šírenie

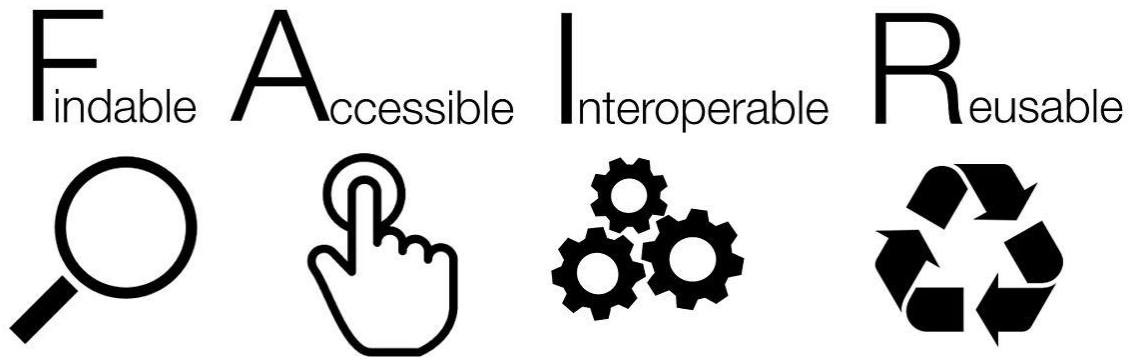
- **Otvorené licencie**
- Otvorená evaluácia (open peer-review)
- Zaisťiť prepojenia medzi publikáciami, dátami a metódami
- Využívať inštitucionálne a tematické repozitáre (nájdite ten najlepší pre vás na **re3data.org**)

opätovné
použitie

Životný cyklus
výskumu z pohľadu
výskumných dát.

**Otvorenosť a
štandardy v boji
proti chaosu.**

Princípy FAIR: Findable, Accessible, Interoperable, Reusable



FAIR – princípy manažmentu dát

(pôvod: konferencia v Leidene, 2014. V roku 2016 boli princípy publikované v článku Wilkinson et al. 2016: <https://doi.org/10.1038/>)

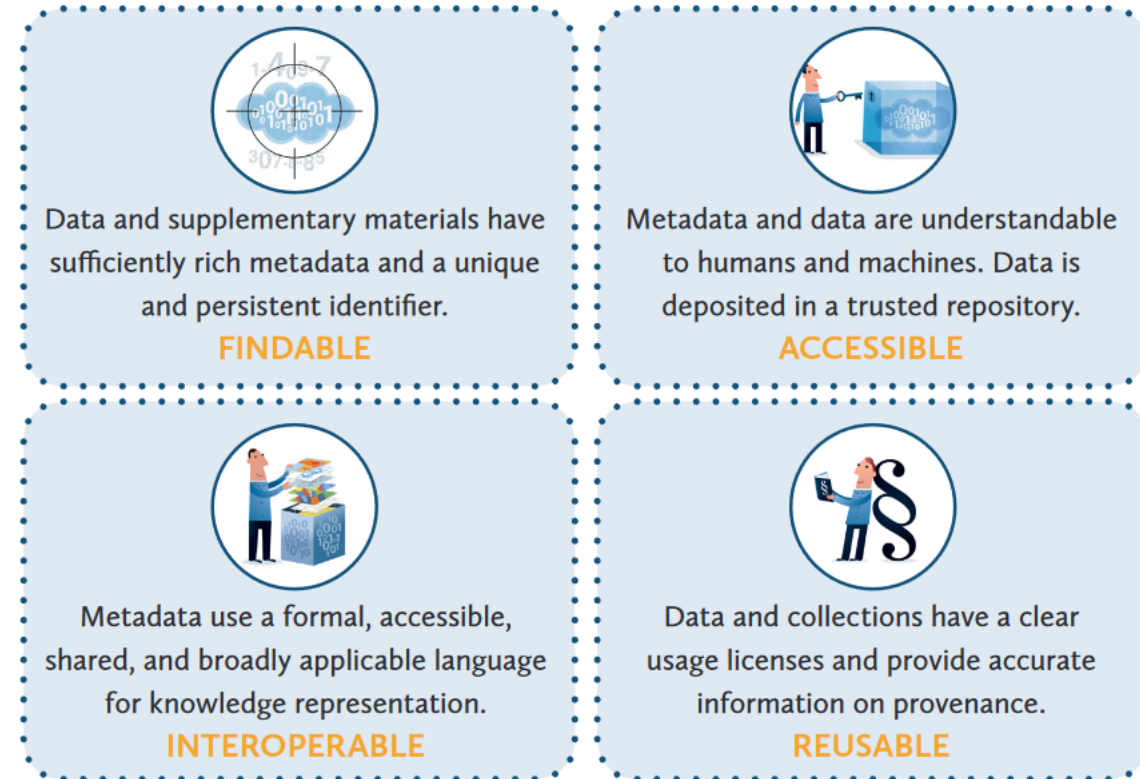
Findable – nájdateľné (metadáta, identifikátory)

Accessible – dostupné (ak aj nie sú „open“, je dôležité, aby bol ich status jasne definovaný)

Interoperable – interoperabilné (dôležitosť prepojenia a kompatibility s inými – napr. štandardy metadát ako Dublin Core, <https://dublincore.org/>)

Reusable – znovu využiteľné (význam licencií, napr. Creative Commons)

What is FAIR DATA?



<https://libereurope.eu/article/implementing-fair-data-principles-role-libraries/>

Princípy FAIR treba dodržiavať počas celého výskumného cyklu.

Keď dáta získavate, myslíte si, že je všetko jasné. Môže sa vám zdať zbytočné zaznamenávať banálne veci o organizácii súborov, význame skratiek a pod. Ale skúste sa na výsledky projektu pozrieť o päť rokov („Čo znamená toto? Ako to, že neviem nájsť x...?“)

Pozor, **FAIR nerovná sa OPEN data!** Dáta môžu (a mali by) byť FAIR, aj keď nebudú prístupné v plnej podobe celej verejnosti.

Aj keď dáta z nejakého dôvodu (osobné údaje, práva priemyselného vlastníctva...) nemôžu byť zdieľané otvorene, môžete vytvoriť opis dát a zdieľať ho verejne, aby tí, pre ktorých by dáta mohli byť užitočné, mohli podľa pravidiel požiadať o prístup k dátam.

Aplikácia princípov FAIR berie do úvahy:

- **Dokumentácia** – poskytuje kontext a robí dáta zrozumiteľné aj pre ostatných vedcov.
- **Metadáta** – štruktúrované dáta o dátach, zvyšujú objaviteľnosť dát a uľahčujú ich strojové spracovanie
- **Formáty dát** – dôležité najmä keď treba kombinovať viac súborov dát z rôznych zdrojov
- **Prístupové práva k dátam** – kto a za akých podmienok môže k dátam pristupovať (musí sa určiť vopred, napr. v informovaných súhlasoch pacientov v rámci klinických štúdií musí byť vopred špecifikované, za akým účelom a za akých podmienok sa dáta použijú, čo z nich sa môže zverejniť a čo nie)
- **Trvalé identifikátory** – DOI sa môže použiť aj na vedecké dáta, ktoré budú vďaka nemu ľahko vyhľadateľné. Okrem DOI ale existujú aj identifikátory fyzických vzoriek, organizácií a mnohých iných vecí vo výskume (tie môžeme použiť pri tvorbe dokumentácie k dátam).

Manažment výskumných dát

Dokumentácia

- Na začiatku projektu sa dohodnúť s kolegami na budúcej štruktúre dát (ako sa budú zapisovať, do akých adresárov sa bude čo dávať, ako sa budú nazývať súbory...)
- Dokumentácia by mala obsahovať všetko potrebné o zbere dát...
 - **metódy**
 - **nástroje**
 - **softvér**
- ... a všetky relevantné záznamy počas celého výskumného procesu:
 - **kto s dátami pracoval**
 - **ako sa spracovávali**
 - **ako súvisia** s inými súbormi dát alebo publikáciami



Manažment výskumných dát

Metadáta

- Uznávaným štandardom pre metadáta je Dublin Core
- Rôzne vedné odbory a oblasti ľudského záujmu majú vlastné metadátové štandardy, napr.:
 - Darwin Core – pre geografický výskyt druhov
 - EML – Ecological Metadata Language pre ekológiu
 - VRA Core – Visual Resources Association pre vizuálne umenie
 - V behaviorálnych, spoločenských a ekonomických vedách sa často používa Data Documentation Initiative (štandard pre prieskumy, dotazníky, štatistické súbory)...
- Ak štandardy vo Vašej oblasti neexistujú, spíšte do zvláštneho súboru všetky informácie, ktoré považujete za potrebné pre porozumenie daným dátam.



Dublin Core Metadata Initiative



FINDABLE, INTEROPERABLE, REUSABLE

Manažment výskumných dát

Formáty dát

- Je dôležité používať formáty súborov zaužívané vo Vašom vednom odbore
- Stáva sa, že výrobcovia prístrojov majú vlastné softvéry spracovávajúce výstupy z prístrojov vo vlastných formátoch (ktoré výskumníci s prístrojmi od inej firmy nemusia vedieť otvoriť)

- **Ideálny formát je:**

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

INTEROPARABLE, REUSABLE

Dobrá prax pre formáty súborov podľa Stanfordskej knižnice:
<https://library.stanford.edu/research/data-management-services/data-best-practices/best-practices-file-formats>

- 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

Manažment výskumných dát

Prístupové práva k dátam

- Niektoré dáta sú citlivé a nie je žiaduce, aby ich mohol vidieť hocikto (či už kvôli právam pacientov, alebo právam priemyselného vlastníctva)
- V takom prípade je dobré **zverejniť metadáta**, aby mohol o prístup k dátam požiadať výskumník, ktorý na to má relevantný dôvod
- Je kľúčové zverejniť aj **podmienky**, za akých môže byť udelený prístup a **kontakt** na oprávnenú osobu, ktorá môže prístup udeliť.
- ...dost' často sa však stáva, že sa vedci bránia myšlienke zdieľania dát len z nezvyku, alebo nešpecifických obáv.



Licencie pre otvorené dáta

- Najpoužívanéjšie licencie sú Creative Commons, napr. CC-BY, CC-BY-SA
- Open Source licencia pre otvorené dáta v IT (softwar, skripty, programy): GPL (General Public Licence)
- Aj metadáta majú mať svoju licenciu (dosť často to býva CC0 – public domain)
- CiteAs (<https://citeas.org/>) nástroj na citovanie rôznych výskumných produktov

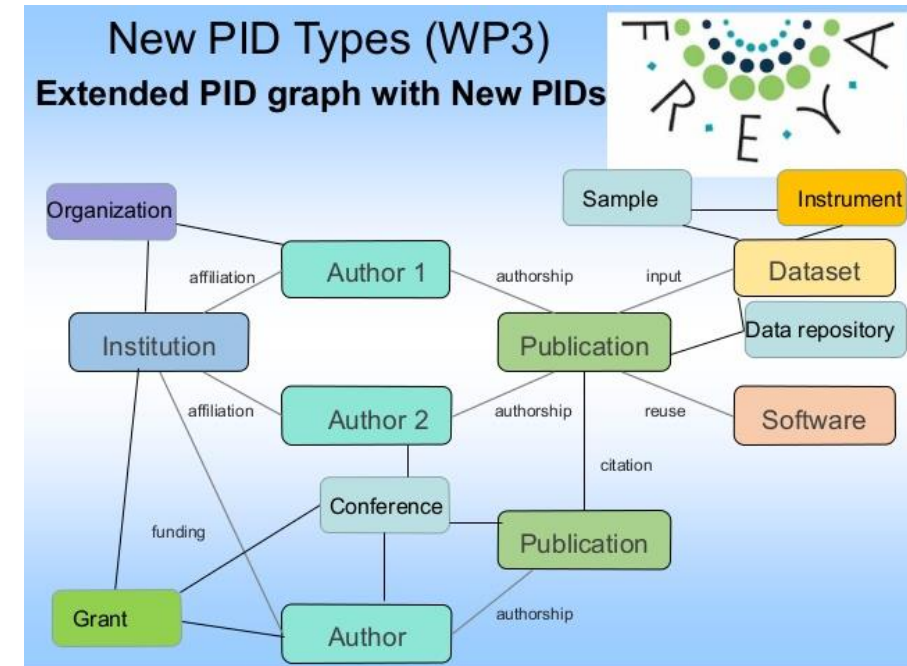
Jednoduchým spôsobom ako zdieľať dáta je zverejniť ich v **repozitári** (dátovom, odborovom, inštitucionálnom, alebo všeobecnom ako napr. **Zenodo**, ktoré už pri vkladaní obsahu dáva užívateľom návod ako postupovať v súlade s princípmi FAIR. Poskytuje aj DOI a licencovanie podľa voľby vkladajúceho a zároveň sa stará o dlhodobé uchovávanie).

Trvalé identifikátory

- používa sa na jednoznačné identifikovanie objektu (online aj offline)
- udržiavané aktívne v priebehu času
- Inštitucionálne spravované systémy trvalých identifikátorov:

Digital Object Identifiers (DOI) a Handle systém, ORCID...

Persistent Uniform Resource Locator (PURL), Uniform Resource Name (URN)...



iniciatíva [FREYA](#)

FINDABLE, ACCESSIBLE, INTEROPERABLE, REUSABLE

Manažment výskumných dát

F

Metadáta
Trvalý identifikátor

A

Kto a ako má prístup

I

metadátové štandardy
štandardné formáty dát

R

Presná dokumentácia

DATA MANAGEMENT PLAN – Plán manažmentu dát



Plán manažmentu dát – Data Management Plan (DMP)

- Viac než formalita – DMP je nástroj, ktorý umožňuje vedcovi
 - systematicky premýšľať o dátovom manažmente
 - robiť strategické rozhodnutia v správnom čase
 - ujasniť si so spolupracovníkmi očakávania ohľadom projektu
 - zjednotiť si so spolupracovníkmi postupy a spôsob dokumentácie.
- O dátach treba rozmýšľať (a rozhodovať sa) PRED, POČAS aj PO uskutočnení výskumu. DMP sa vypracováva na úplnom začiatku, ale v priebehu výskumu sa môže modifikovať podľa aktuálneho vývoja situácie.
- Čoraz viac sa razí politika povinných DMP (ako súčasť žiadosti o grant)
- Niektoré grantové agentúry a donori (napr. ERC) mávajú vzory DMP, do ktorých výskumník vyplní údaje špecifické pre svoj projekt.
- Výskumné inštitúcie vo vyspelých krajinách už majú vo vnútorných predpisoch povinnosť vypracovávať DMP
- Existujú v nich aj pracovné pozície „podpora manažmentu dát“.

Na aké otázky odpovedáte v pláne manažmentu dát

- Aký druh dát budete zbierať počas výskumu? (dotazníky, merania, texty, fotografie, skeny...)
- Budú medzi nimi špeciálne, citlivé kategórie dát? (osobné údaje, biologické vzorky, komerčné údaje, ktoré by mohli ovplyvniť obchodnú súťaž...)
- Musíte pred začatím projektu žiadať o súhlas (osoby – napr. pacientov, alebo organizácie, napr. Štátnu veterinárnu a potravinovú správu, Štátnu ochranu prírody, Etickú komisiu...)?
- Potrebujete pred začatím projektu splniť nejaké právne povinnosti?
- Ste si vedomí práv a povinností pri používaní dát iných strán (napr. z komerčných databáz – niektoré môžu mať v zmluvách zákaz zdieľania podkladových dát po ukončení projektu, alebo obmedzenie objemu dát, ktoré môže výskumník extrahovať)?
- V akých formátoch budete dáta ukladať?
- Aký je predpokladaný objem získaných dát?
- Máte stratégiu pomenovávania súborov a adresárov?

Na aké otázky odpovedáte v pláne manažmentu dát

- Ako budete dáta dokumentovať, opisovať? Metadát je veľa – ale ktoré budú kľúčové pre váš projekt?
- Ako zabezpečíte, že budú vaše dáta zrozumiteľné pre iných?
- Ako zabezpečíte svoje dáta (aby k nim nemali prístup neoprávnené osoby – napr. dáta sú chránené heslom na univerzitnom serveri, citlivé dáta v papierovej podobe zamknuté v skrinke)?
- Ako zaistíte zálohovanie dát?
- Akým spôsobom budete zdieľať dáta, ktoré majú byť zverejnené (napr. v repozitári)?
- Ako zabezpečíte dlhodobé uchovávanie? Ako dlho (napr. minimum 5 rokov pre dáta, ktoré sú podkladom pre publikácie)?
- Ktoré z množstva získaných dát majú byť uchovávané dlhodobo a ktoré majú byť zničené/vymazané?

Pomoc pri vytváraní DMP:

Pokyny pre data management pri projektoch **Horizont 2020**:

https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Webový nástroj **DMP online**

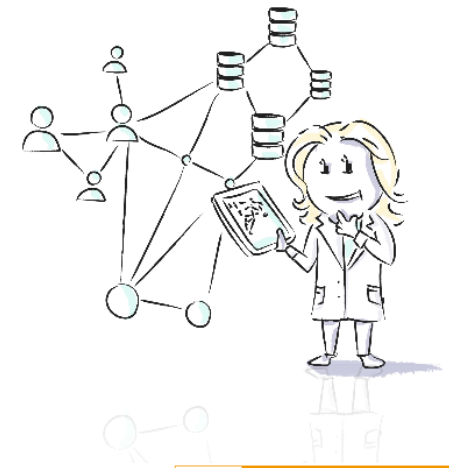
<https://dmponline.dcc.ac.uk/>



[ABOUT](#) [RESOURCES](#) [CONTACT](#) [LOG IN](#)

Plan and follow your data

- Create** machine actionable DMPs.
Configure to best fit your discipline.
Link to EOSC components out of the box.
Share easily in your repository.



Bring your Data Management Plans closer to where data are generated, analysed and stored.

Plan to make data work for you

Data Management Plans that meet institutional funder requirements.



Journal of the American Chemical Society

Sign in Create account

* **Email**

* **Password**

[Forgot password?](#)

☐ Remember email

Sign in

- or -

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).



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	 290 Participating Institutions
More	More	More

Top Templates

- Digital Curation Centre
- Template USP - Baseado no DCC
- Template USP - Mínimo
- NSF-SBC: Social, Behavioral, Economic Sciences
- UNICAMP-GENÉRICO. Aplicável a todas as áreas

Icon Description	Value
Three people icons	59,972 Users
Building icon	314 Organisations
Document icon	65,212 Plans
Globe icon	89 Countries

101001001110001101110100 0 1 0 1 1 01 1 1

Zaujímavé stránky:

Guidelines on FAIR Data Management in H2020:

https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

Managing and Sharing Data (UK Data Archive): <https://ukdataservice.ac.uk/media/622417/managingsharing.pdf>

RDM Rose (vzdelávacie materiály): <http://rdmrose.group.shef.ac.uk/>

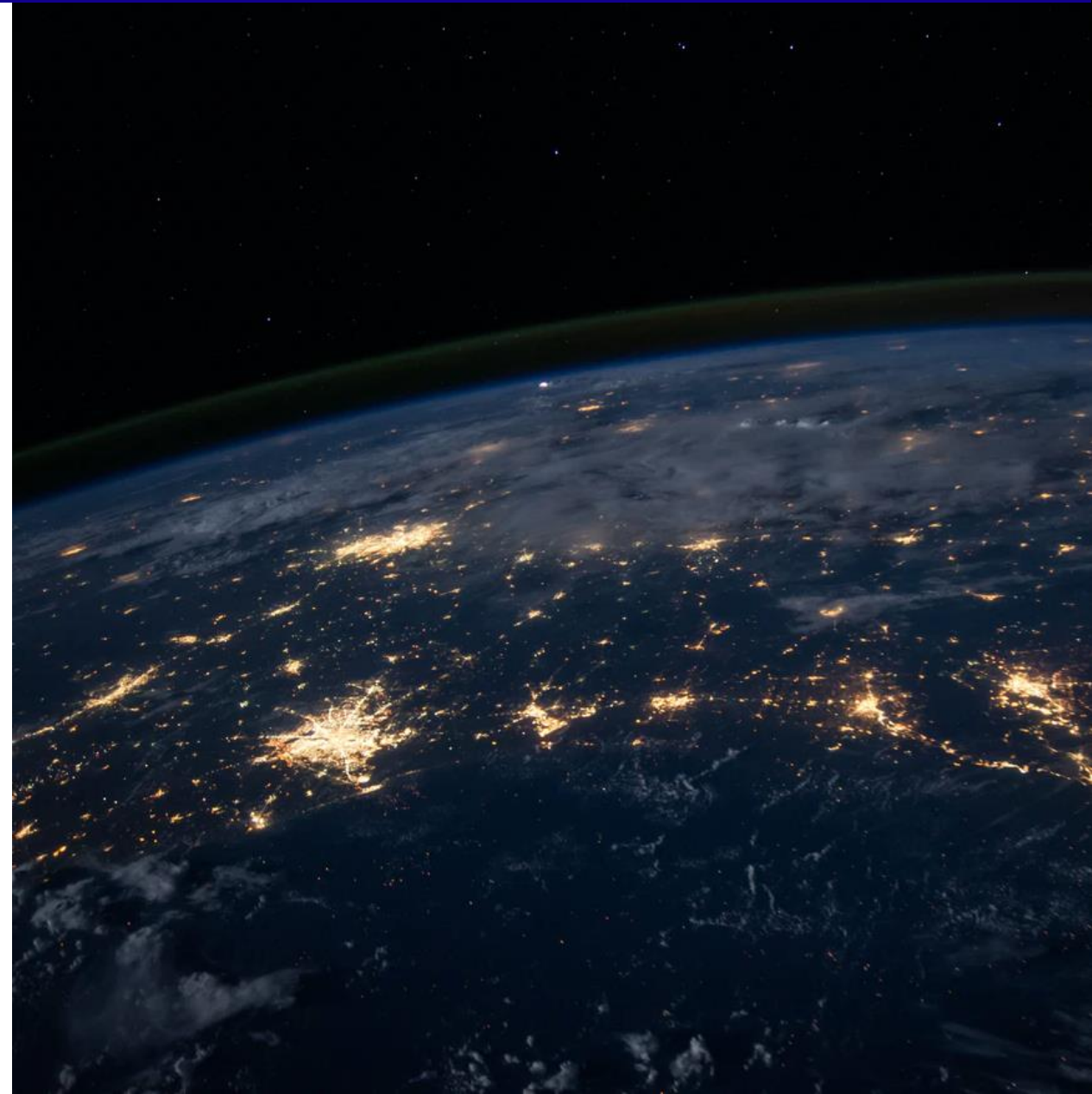
Creating a Data Management Plan Document (Open Science Framework): <https://help.osf.io/hc/en-us/articles/360019931133-Creating-a-data-management-plan-DMP-document>

Research Data Alliance <https://rd-alliance.org/>

CODATA – Committe on DATA of the International Science Council (<https://codata.org/about-codata/>)

Ako zodpovedný manažment dát prospieva vedcom a vede

- Zdieľanie dát zvyšuje **dosah** výskumu a **prestíž** výskumníka aj výskumnej organizácie, pre ktorú pracuje.
- Zdieľanie dát zvyšuje ekonomickú efektivitu výskumu (pridaná hodnota opätovného zdieľania + možnosť zabrániť zbytočnému duplikovaniu).
- Zdieľanie dát tiež prispieva k **transparentnosti a reprodukovateľnosti** výskumu a pôsobí ako prevencia „sloppy science“ (nedbalej vedy) , pretože si podkladové údaje môže každý skontrolovať (recenzent v rámci recenzného konania, iný vedec z danej oblasti v rámci prieskumu literatúry, aj spätne po rokoch) a vidieť, čo s nimi výskumník robil.



Záznamy z predchádzajúcich webinárov:

<https://eiz.cvtisr.sk/eiz-webinare/>

Informácie o EIZ z ponuky CVTI SR:

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

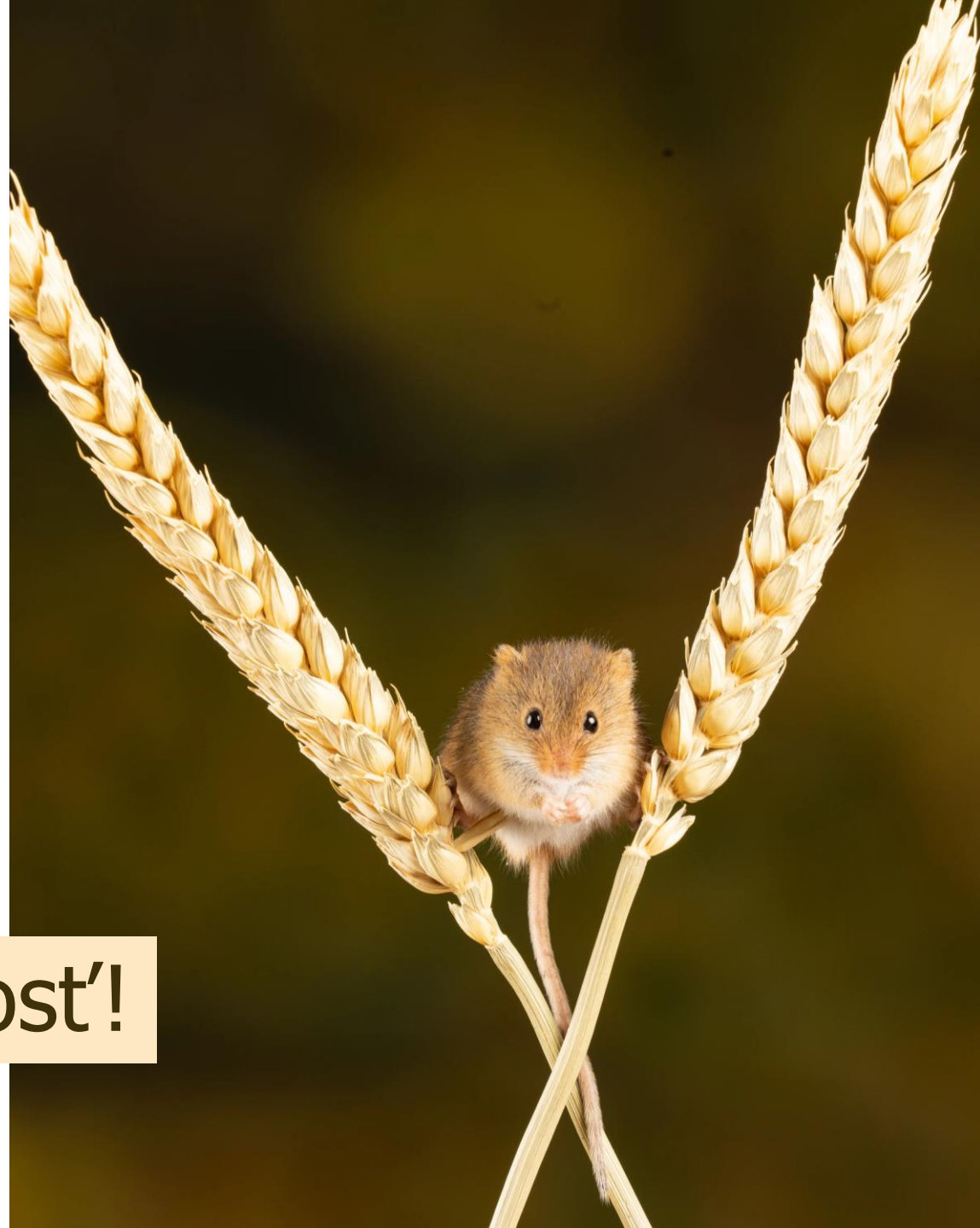
Informácie o Open Access:

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

Dotazník pre vedeckých pracovníkov a doktorandov:

<http://anketa.cvtisr.sk/index.php/755813/lang-sk>

Ďakujem za pozornosť!



10100100110001101110100 0 1 0 1 1 01 1 1

Zdroje obrázkov (ak nie sú uvedené priamo pri obrázkoch), alebo zaujímavé linky pre ďalšie štúdium

- **Str. 2:** svetlá mesta v noci z leteckej perspektívy: Denis Kummer, unsplash, <https://unsplash.com/photos/52gEprMkp7M>
- **Str. 3:** Ilustrácia k chybovému hláseniu Google 404 error page <https://kinsta.com/blog/error-404-not-found/>,
obrázok knižnice: Alfons Morales, unsplash: <https://unsplash.com/photos/YLSwjSy7stw>
- **Str. 4:** obrázok archívu: Ula Kuzma, unsplash, <https://unsplash.com/photos/9i4DHIC80AQ>
- **Str. 5:** obrázok archívu: C M, unsplash https://unsplash.com/photos/X_j3b4rqnlk
- **Str. 6:** Green Open Access, Australasian OA Strategy Group: <https://aoasg.org.au/what-is-open-access/>
- **Str. 7:** inštitucionálny repozitár SAV: <https://uk.sav.sk/pre-vedcov/open-access/institucionalny-repozitar-sav/>
- **Str. 8:** schéma repozitára DSpace, Smith et al. 2003: <http://dlib.org/dlib/january03/smith/01smith.html>
- **Str. 9:** graf počtu repozitárov podľa kontinentu: Bawack a Nkolo 2018, Open Access Movement: Reception And Acceptance By Academic Libraries In Developing Countries" (2018).Library Philosophy and Practice (e-journal):
<https://core.ac.uk/download/pdf/188141049.pdf>
- **Str. 10:** Dátová explózia s predikciou do r. 2024, Statista 2020, <https://www.statista.com/statistics/871513/worldwide-data-created/>
- **Str. 11:** Managing qualitative data at UofG, <https://uofgpgrblog.com/pgrblog/2018/2/9/managing-qualitative-research-data-at-uofg>
- **Str. 12:** Graf o práci s dátami <https://www.integrity-research.com/big-data-big-mess-fixing-isnt-easy/>
- **Str. X:** Ďakujem za pozornosť: obrázok: Nick Fewings, unsplash, <https://unsplash.com/photos/-dtKoaHpi9M>