

Archivácia a manažment výskumných dát

Archivácia

Gabriela Fišová

Centrum vedecko-technických informácií SR



Investícia do Vašej budúcnosti



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



Obsah

- Archivácia – pamäť vedeckého záznamu
- Čo sa stane s obsahom časopisu, keď vydavateľ zanikne?
- Archivačné iniciatívy
- Repozitáre
- Výber repozitára – registre repozitárov
- Dátové repozitáre



Archivácia



Archivácia – pamäť vedeckého záznamu

- kontinuita vedeckého záznamu
- vedecké články a dátové súbory sú vyhľadávané a citované (= potrebné) aj storočia po svojom uverejnení

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?

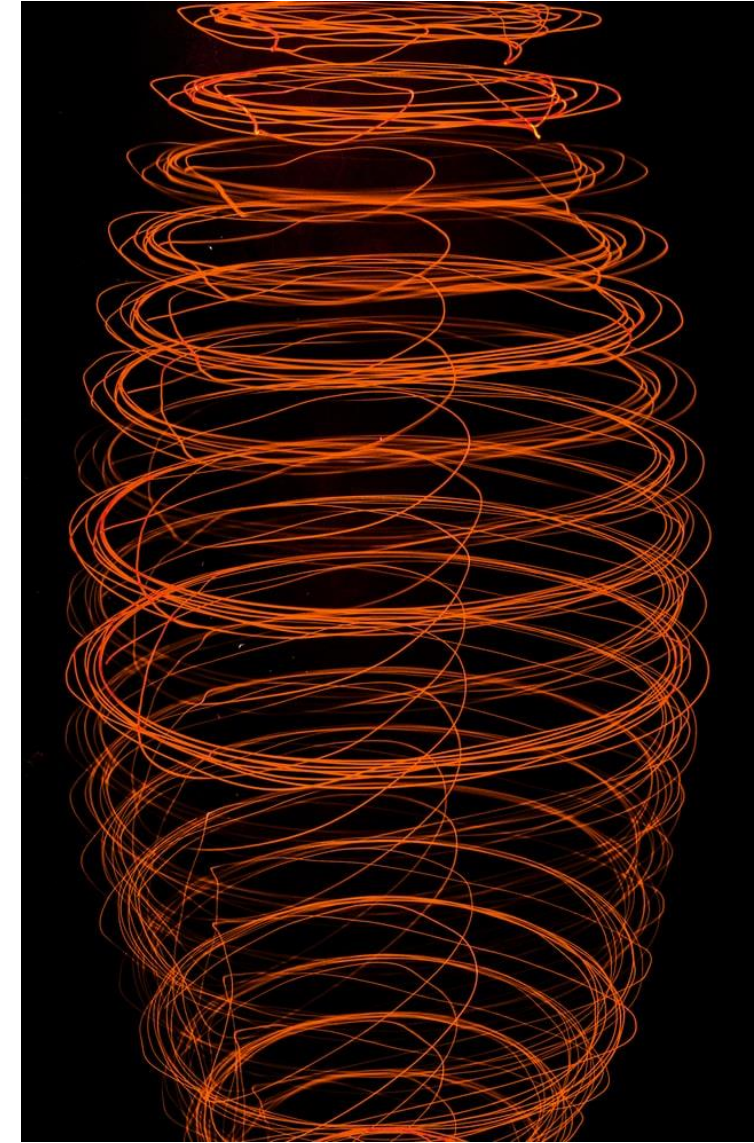


Photo by [Kiarash Mansouri](#) on [Unsplash](#)

Archivácia

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 na budúce použitie, pričom súčasťou je archivácia a/alebo odosielanie dát do repozitára (úložiska).



PORTICO

Seriózni vydavatelia sa zapájajú do archivačných iniciatív

[Portico](https://www.portico.org/) (<https://www.portico.org/>)

- **tmavý archív** (*dark archive*) e-časopisov, kníh a digitálnych kolekcií



JSTOR

[JSTOR](https://www.jstor.org/) (Journal storage, <https://www.jstor.org/>) – 12 mil. článkov v časopisoch, 100000 kníh, milióny obrázkov (images) a primárne zdroje v 75 disciplínach

- prístup aj k otvorenému obsahu

Archivácia



- ďalšie: [e-Depot](#) (Holandsko), [OCLC](#) (British Library), [HathiTrust](#) (konzorcium univerzít)
- na Slovensku: projekt [DIKDA](#) (SNK)



[LOCKSS](#) (*Lots Of Copies Keep Stuff Safe*, www.lockss.org) – Stanford University, od r. 1998

- systém, ktorý kontroluje viaceré kópie dokumentu voči sebe navzájom, aby odhalil náhodne vzniknuté chyby, prípadné pokusy o zmenu či potlačenie materiálu, ktorý bol vydaný

[CLOCKSS](#) (*Controlled LOCKSS*, <https://clockss.org/>) – dlhodobé uchovávanie, chráni obsah pred stratou, poškodením aj pre budúce generácie
- **tmavý archív** (dark archive)



Repozitáře



Repozitáre

Pojem **digitálny repozitár** zaviedli v roku 1995 Robert Kahn a Robert Wilensky

- definovali teoretické princípy otvoreného univerzálneho informačného systému ([Kahn-Wilensky Framework](#))
- definovali komponenty otvoreného systému na uchovávanie, prístup a manažment (organizáciu, riadenie) informácií, architektúru systému oddeľujú od uchovávaného obsahu

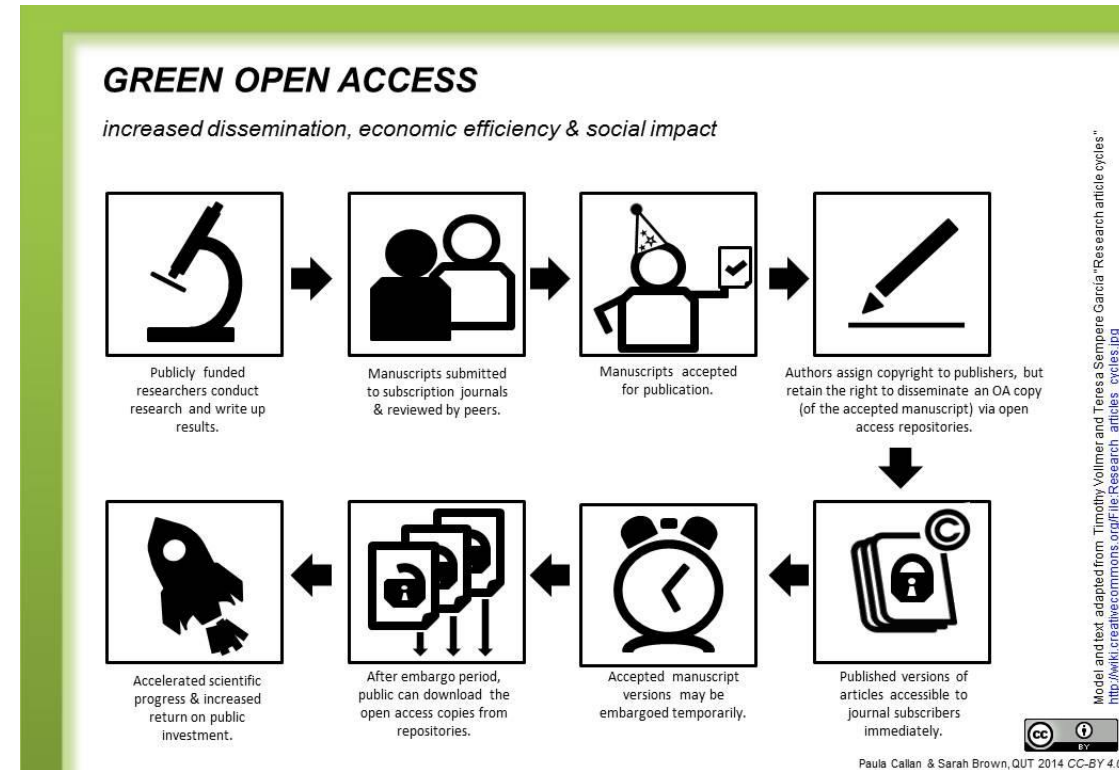
Digitálny repozitár/úložisko (Digital Repository/Storage)

– **systém na prevádzku súboru služieb**, ktoré sa zaoberajú získavaním, správou a sprístupňovaním digitálneho obsahu.



Digitálny archív (Digital Archive) – organizovaná zbierka digitálnych dokumentov (born digital alebo digitalizovaných) zhromažďovaná s cieľom ich dlhodobého uchovávania a ochrany.

Autoarchivácia - autor si archivuje svoje dielo sám, napr. v repozitári – cesta nezávislá od vydavateľov (**zelená cesta OA**)



Webinár o **zelenom otvorenom prístupe** – Preprinty a ich miesto vo vedeckej komunikácii: <https://video.nti.sk/rec/?735676N02NKILA99>

Repozitáre

Obsah

- vnútri každého repozitára je špecializovaný systém súborov
- fond tvoria **digitálne objekty**

Digitálny objekt obsahuje:

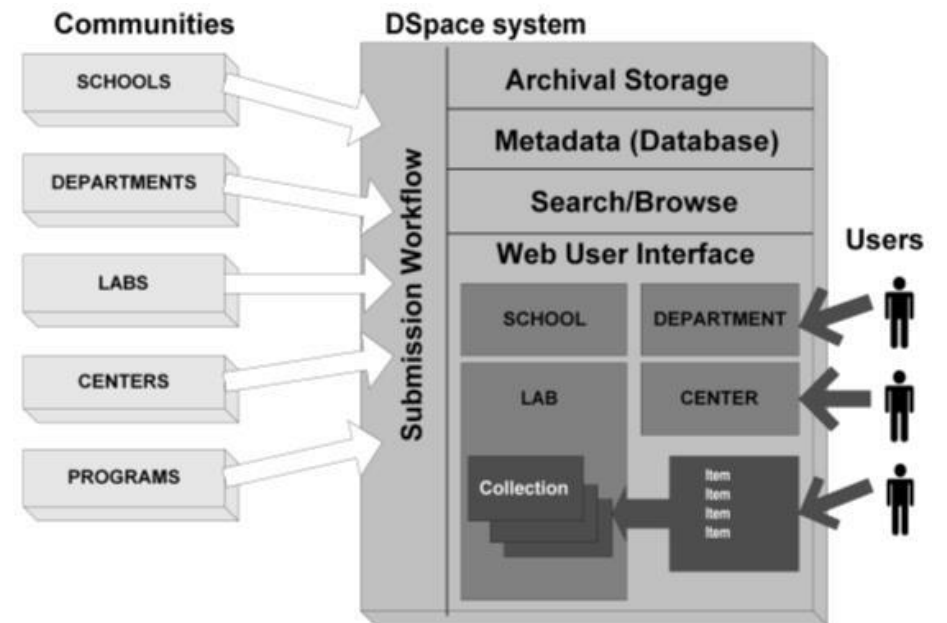
- jedinečný identifikátor (**handle**)
- vlastnosti, metadáta (properties)
- obsah v bitoch (content)
- digitálny podpis (signature), zabezpečenie autenticity obsahu
- záznamy o použití, vyhľadávaní objektu (transaction log)

Otvorené softvéry pre repozitáre (*softvér s otvoreným zdrojovým kódom*): napr. DSpace, Fedora, EPrints, Invenio (ZENODO)

Hostingové riešenia: napr. SimpleDL, DSpaceDirect.

Funkcie repozitára:

- spracovanie dokumentov
- tvorba metadát
- uloženie a archivácia digitálneho obsahu
- zabezpečenie integrity a autenticity
- prístupnenie dokumentov



Repozitáre

Typy

Predmetové/Odborové (tematické)

arXiv.org



[Predmetové repozitáre](#) (Subject-based/Disciplinary)

Inštitucionálne (*Autoarchivácia + Zelená cesta*) –

[DR Žilinskej univerzity](#), [repozitár UMB](#), [Inštitucionálny repozitár SAV](#),



Národné



Multiodborové (univerzálne/všeobecné)



Dátové



(GIGA)ⁿ DB



[Data Repository Guidance - Nature](#)
Predmetové, všeobecné...

Softvérové



Repozitáre

Dátové repozitáre

Dátový repozitár (úložisko údajov/dát) - obrovská databázová infraštruktúra, ktorá zhromažďuje, spravuje a ukladá rôzne súbory údajov na analýzu, distribúciu a tvorbu požadovaných výstupov.

Typy dátových repozitárov

Data Warehouse

Data Lake

Data Marts

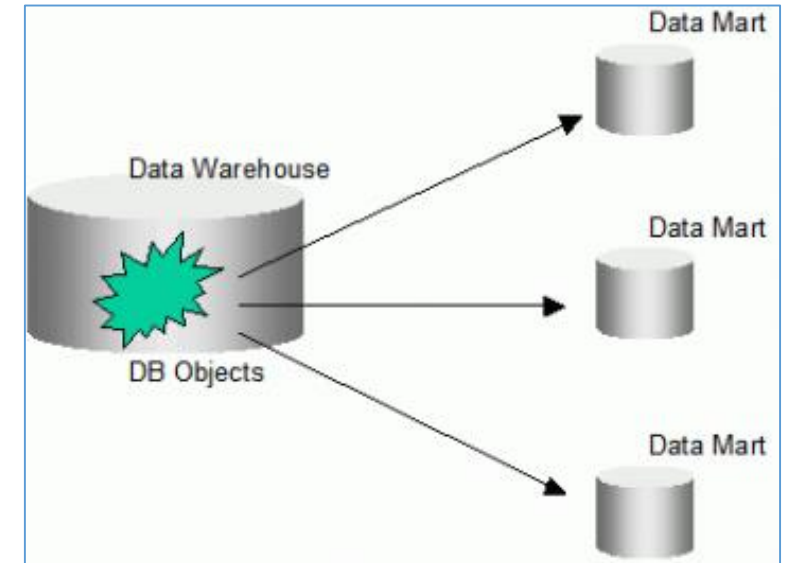
Data Cubes

Metadata Repositories



Data warehouse

Data Mart



Repozitáře

Data Warehouse		Data Lake
Strukturovaná, předem zpracovaná	Data	Surová, strukturovaná, nestrukturovaná
Schema-on-write	Zpracování	Schema-on-read
Drahé pro velké objemy dat	Uložení dat	Navrženo pro běžná (levná) úložiště
Méně agilní. Nelze snadno měnit účel dat ze dne na den.	Flexibilita	Vysoce agilní. Mohu měnit konfigurace a účel dat.
Pokročilá	Bezpečnost	Pokročilejší
Byznysoví analytici	Primární uživatele	Datoví vědci



Repozitáre

Dôveryhodné digitálne repozitáre (TDR, Trustworthy Digital Repositories)

Referenčný model OASIS

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 sprístupňovania a kontextu použitia.
8. Musí byť schopný zabezpečiť splnenie požiadaviek na **sprístupňovanie** 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.

Repozitáre

Dôveryhodnosť – audit

DRAMBORA interactive

Digital Repository Audit Method
Based on Risk Assessment



DRAMBORA (Digital Repository Audit Method Base on Risk Assessment)

- metóda interného auditu, ponúka kvantifikovateľný pohľad na závažnosť rizík, ktorým repozitáre čelia, schopnosť dosahovať stanovené ciele

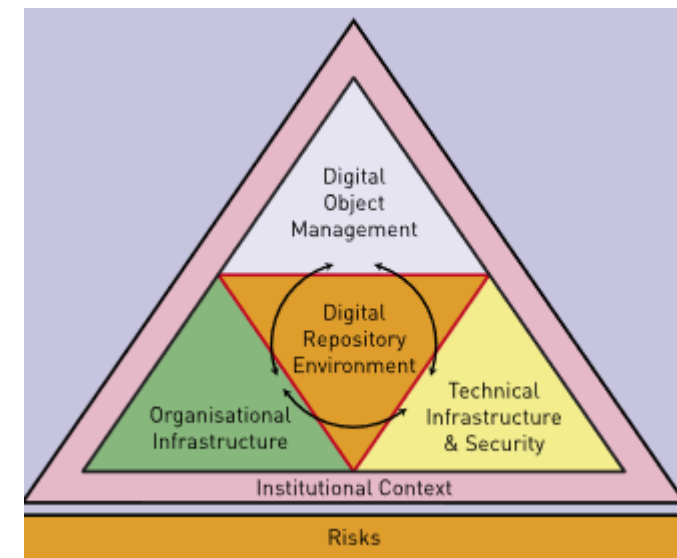
PLATTER (A Planning Tool for Trusted Electronic Repositories)

- súbor nástrojov, dopĺňa nástroj DRAMBORA

nestor-Network of Expertise in Long-Term Storage and Long-Term availability of Digital Resources,

Nemecko

Digital Preservation Capability Maturity Model (DPCMM)



<https://www.dcc.ac.uk/news/dpe-releases-platter-planning-tool-trusted-electronic-repositories>
https://digitalpreservationeurope.eu/publications/reports/Repository_Planning_Checklist_and_Guidance.pdf

V českém jazyku: https://digitalpreservationeurope.eu/platter/platter_akm2008.pdf

https://wikisofia.cz/wiki/D%C5%AFv%C4%9Bryhodn%C3%BD_repozit%C3%A1%C5%99_audity_a_certifikace

<https://digitalpreservationeurope.eu/competence-centres/list/?id=92>

<https://www.dpconline.org/handbook/institutional-strategies/audit-and-certification>

Repozitáre

Dôveryhodnosť – certifikácia



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



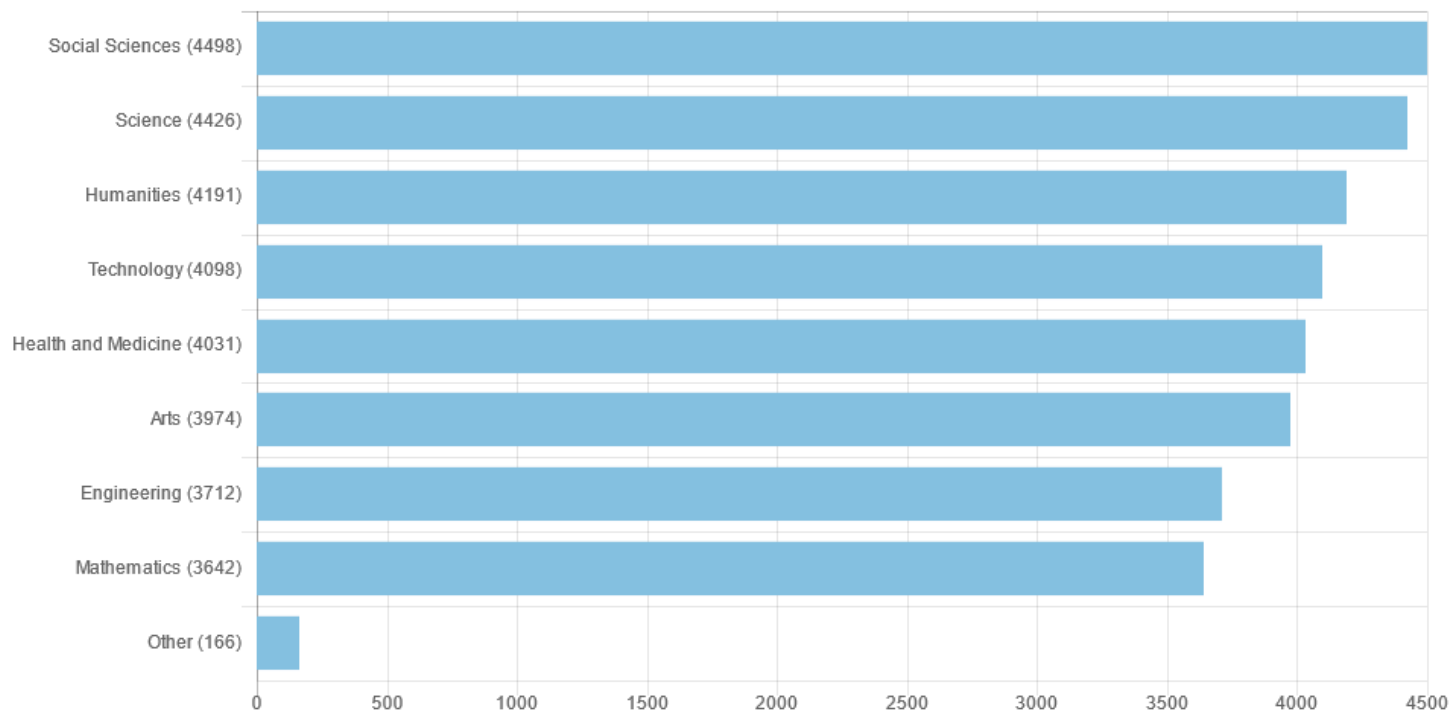
Nestor Seal - podľa normy nemeckej normy DIN 31644

Repozitáre

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/>)
- **Registry of Research Data Repositories**, [re3data](https://www.re3data.org/), (<https://www.re3data.org/>)
- **Registry of Open Access Repository Mandates**, [ROARMAP](http://roarmap.eprints.org/), (politiky, <http://roarmap.eprints.org/>)
- **SherpaRomeo** (<https://v2.sherpa.ac.uk/romeo/>)
umožňuje overiť, ktorá verzia článku sa môže uložiť do repozitára



Zdroj: [OpenDOAR Statistics](https://v2.sherpa.ac.uk/view/repository_visualisations/1.html), Content Subjects Overview, 02-02-2023
https://v2.sherpa.ac.uk/view/repository_visualisations/1.html

Zhrnutie na záver

Prečo archivovať dáta / články / knihy...?

Čo je archivácia?

Čo je tmavý archív?

Aké systémy dlhodobého uchovávanía majú vydavatelia k dispozícii?

Digitálny repozitár – charakteristika, typy (inštitucionálny, predmetový/tematický, národný, multiodborový/univerzálny, dátový, príklady)

Repozitár - obsah

- funkcie
- softvér
- dôveryhodnosť (, referenčný model OASIS – 10 princípov, DRAMBORA, certifikácia...)

Výber repozitára – registre repozitárov

Archivácia – ďalšie zdroje, odkazy, použité zdroje

Zdroje a odkazy na štúdium:

- Diego Menchaca. (2019). *6 repositories to share your research data*. <https://www.teamscopeapp.com/blog/6-repositories-to-share-your-research-data>
- ASTERA. *Data Repository: Importance, Challenges, and Best Practices*. <https://www.astera.com/type/blog/data-repository/>
- Nature.com. *Recommended Data Repositories*. <https://www.nature.com/sdata/policies/repositories>
- Metadátové repozitáre: https://en.wikipedia.org/wiki/Metadata_repository
- <https://www.dataversity.net/what-is-a-metadata-repository/>
- <https://www.dataversity.net/metadata-repository-basics-from-database-to-data-architecture/>
- <https://www.questionpro.com/blog/data-repository/>
Data Repository: What it is, Types and Guide
- https://link.springer.com/referenceworkentry/10.1007/978-0-387-39940-9_909
- Manuál EK k programovému obdobiu HE: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf
Repozitáre - strana 48, 49
- V slovenskom jazyku: 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>

Použité zdroje:

Zdroje uvedené na jednotlivých snímkach a prezentácie a tieto zdroje:

- Archivácia a manažment výskumných dát, Zuzana Stožická. CVTI SR. <https://eiz.cvtisr.sk/kurzy/eiz-pre-vedu-publikacny-poradca/>
- Webová stránka: <https://www.ica.org/en/why-archiving>
- Manažment výskumných dát. Archivácia a repozitáre. Ľubica Jamborová. CVTI SR. <https://eiz.cvtisr.sk/kurzy/eiz-pre-vedu-publikacny-poradca/>
- Trusted Digital Repositories: <https://www.icpsr.umich.edu/web/pages/datamanagement/preservation/trust.html>; <https://www.dpconline.org/handbook/institutional-strategies/audit-and-certification> ; <https://www.dpconline.org/handbook/institutional-strategies/standards-and-best-practice> ; <https://old.ndk.cz/archivace/certifikace-dlouhodobeho-uloziste> ; https://wikisofia.cz/wiki/D%C5%AFv%C4%9Bryhodn%C3%BD_repozit%C3%A1%C5%99:_audit_a_certifikace
- Webová stránka: <https://www.integrate.io/glossary/what-is-data-mart/>
- What Is the Difference between Data Warehouse and Data Mars, Data Swamp, Data Lake, and Data Cube? <https://intersog.com/blog/what-is-the-difference-between-data-lakes-data-marts-data-swamps-and-data-cubes/>

Obrázky - zdroje (ak nie je uvedené priamo pri obrázku):

Snímka 3: Scrolls: Photo by [Wilhelm Gunkel](#) on [Unsplash](#)

Archive Shelves at Sächsisches Staatsarchiv in Dresden, Saxony, Germany: Photo by [C.M](#) on [Unsplash](#)

Computer with a code. Photo by [Markus Spiske](#) on [Unsplash](#)

Snímka 7: Free computer server room image, public domain CC0 photo.

National Archives, UK. This file is licensed under the Creative Commons Attribution 3.0 Unported license.

National Wildlife Property Repository. <https://www.flickr.com/photos/usfwsmtprairie/33462063285>

Licencia: <https://creativecommons.org/licenses/by/2.0/>

Snímka 8: <https://bond.libguides.com/open-access-and-scholarly-publishing/publishing>

Snímka 5, 6, 8, 9, 13, 14, 15: logá a obrázky – scan - webové stránky inštitúcií

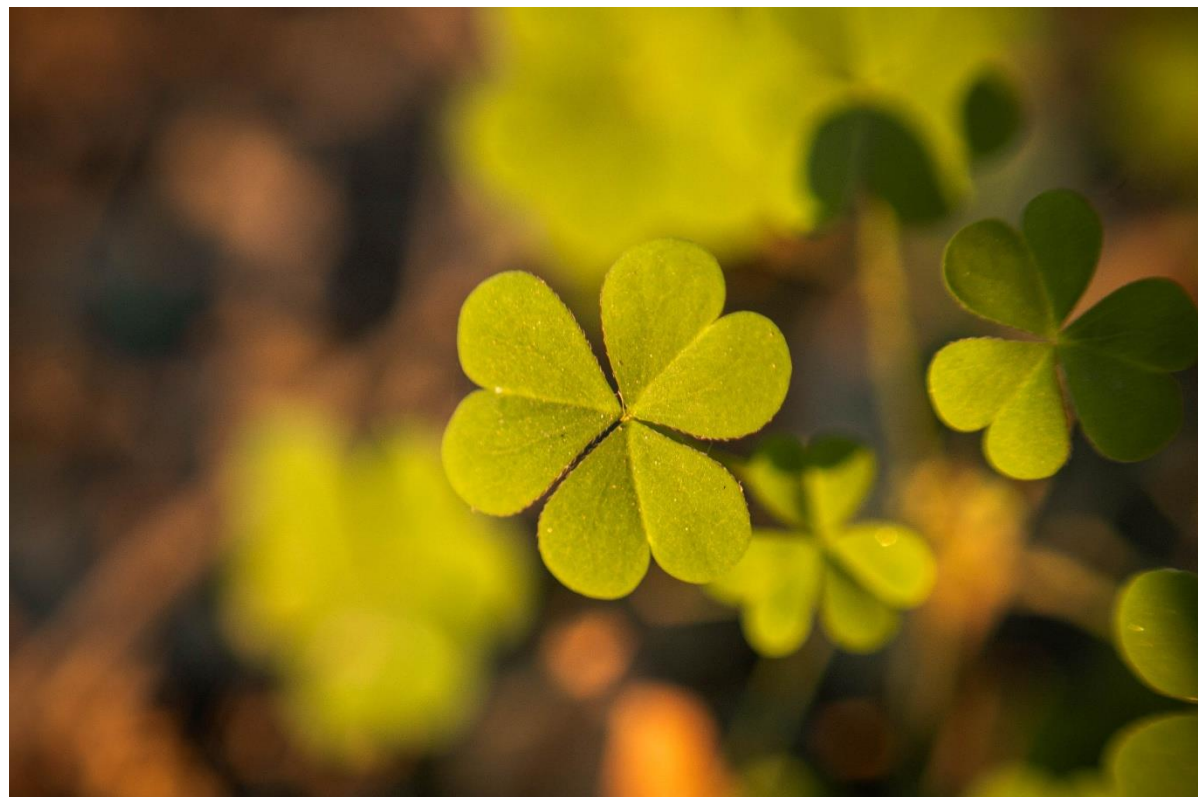
Ďakujem za pozornosť.

Kontakt:

gabriela.fisova@cvtisr.sk

otvorenaveda@cvtisr.sk

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



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

Photo by [Shazaf Zafar](#) on [Unsplash](#)