

Archivácia

Archivácia a manažment výskumných dát

Gabriela Fišová

Centrum vedecko-technických informácií SR

STU, 27. 2. 2024



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

Prečo archivovať – zabezpečenie kontinuity vedeckého záznamu

- Vedecké články a dátové súbory sú vyhľadávané a citované (= POTREBNÉ) aj **storočia po svojom uverejnení**.

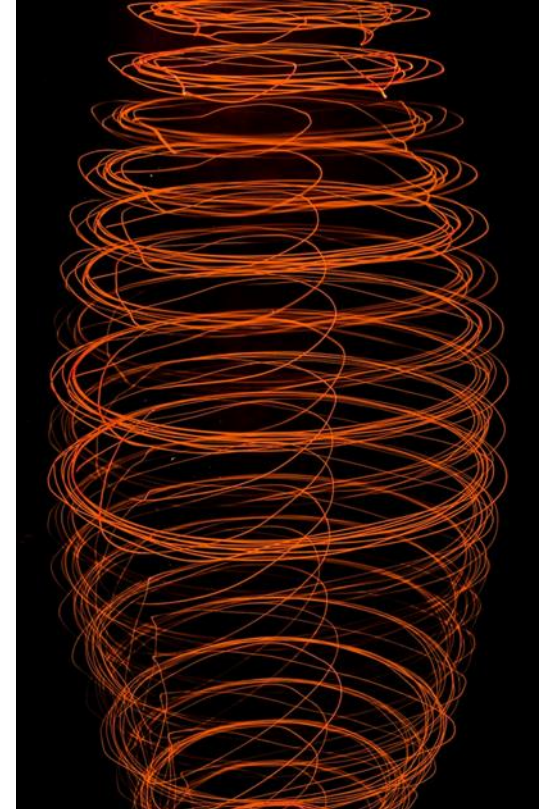
Samozrejme nie všetky, ale nikto nevie vopred určiť, ktoré budú „tie šťastné“.

- 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: digitálny obsah - **Error 404** ... ALEBO...

**Kam s tým všetkým
digitálnym
obsahom?**



Archivácia

OCHRANA (PRESERVATION)

Súbor postupov a činností zameraných na:

- **uchovanie dát** po určitú dobu a/alebo
- **zhromažďovanie dát na budúce použitie,**
- súčasťou je **archivácia** a/alebo **odosielanie dát do repozitára** (úložiska).

ARCHIVÁCIA

Dlhodobé uchovávanie a ochrana dát/dokumentov,
metódy/postupy na ich čítania a interpretácie.



Archivácia

Seriózní vydavatelia sa zapájajú do archivačných iniciatív



JSTOR

(Journal storage, <https://www.jstor.org/>), od r. 1995

- 12 mil. článkov v časopisoch, 100000 kníh, milióny obrázkov (images) a primárne zdroje v 75 disciplínach
- prístup aj k otvorenému obsahu, <https://about.jstor.org/oa-and-free/>



Portico

(<https://www.portico.org/>) , od r. 2005

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

Archivácia



LOCKSS

(*Lots Of Copies Keep Stuff Safe*, www.lockss.org)
Stanford University, od r. 1998

System, 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/>, od r.2006)

- **tmavý archív** (dark archive)
- dlhodobé uchovávanie
- chráni obsah pred stratou, poškodením aj pre budúce generácie

Archivácia



LOCKSS



CLOCKSS



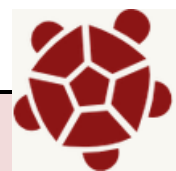
PORTICO

Conditions for Trigger event			
	LOCKSS	CLOCKSS	PORTICO
When library cancels subscription with publishers and needs access to back issues to which they had subscription	Yes	No	Yes, but if library discontinues Portico participation, then they will no longer be able to get post-cancellation access to content through Portico.
E-Journal and its past issues are no longer available from the publisher	Yes	Yes, but the title would be openly made accessible to all	Yes, the title would be made accessible to all active participants without considering whether they were previously subscribing the content or not
When publisher ceased operation and e-content is no longer available	Yes	Yes, but the title would be openly made accessible to all	Yes, the title would be made accessible to all active participants without considering whether they were previously subscribing the content or not
Natural disasters/ Catastrophic failure	Yes	Yes, but on condition if the publisher is unable to provide service due to the said reason	Yes, but on condition if the publisher is unable to provide service due to the said reason
Temporary failure of publisher's operation/ servers	Yes	No	No

Archivácia






PORTICO



LOCKSS



CLOCKSS

 LOCKSS	 CLOCKSS	 PORTICO
Open source software	Open source software	Proprietary software
Distributed, peer to peer platform with error detection	Distributed, peer to peer platform with error detection	Centralized, hosted platform
Small workstation required, it can be run of a CD	Specific server hardware required	No equipment required from client side
Light archive (even for short term preservation)	Dark archive (long term preservation)	Dark archive (long term preservation)
Preserves content in publisher's original format	Preserves content in publisher's original format	Use standard format for preservation

Archivácia

KB } nationale
bibliotheek

[e-Depot](#) (Holandsko)



[OCLC](#) (British Library)



[HathiTrust](#) (konzorcium univerzít)



Slovensko: projekt [DIKDA](#) (SNK)

Repozitáře



Repozitáre



Digitálny repozitár

Robert Kahn a Robert Wilensky, 1995

- 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)

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

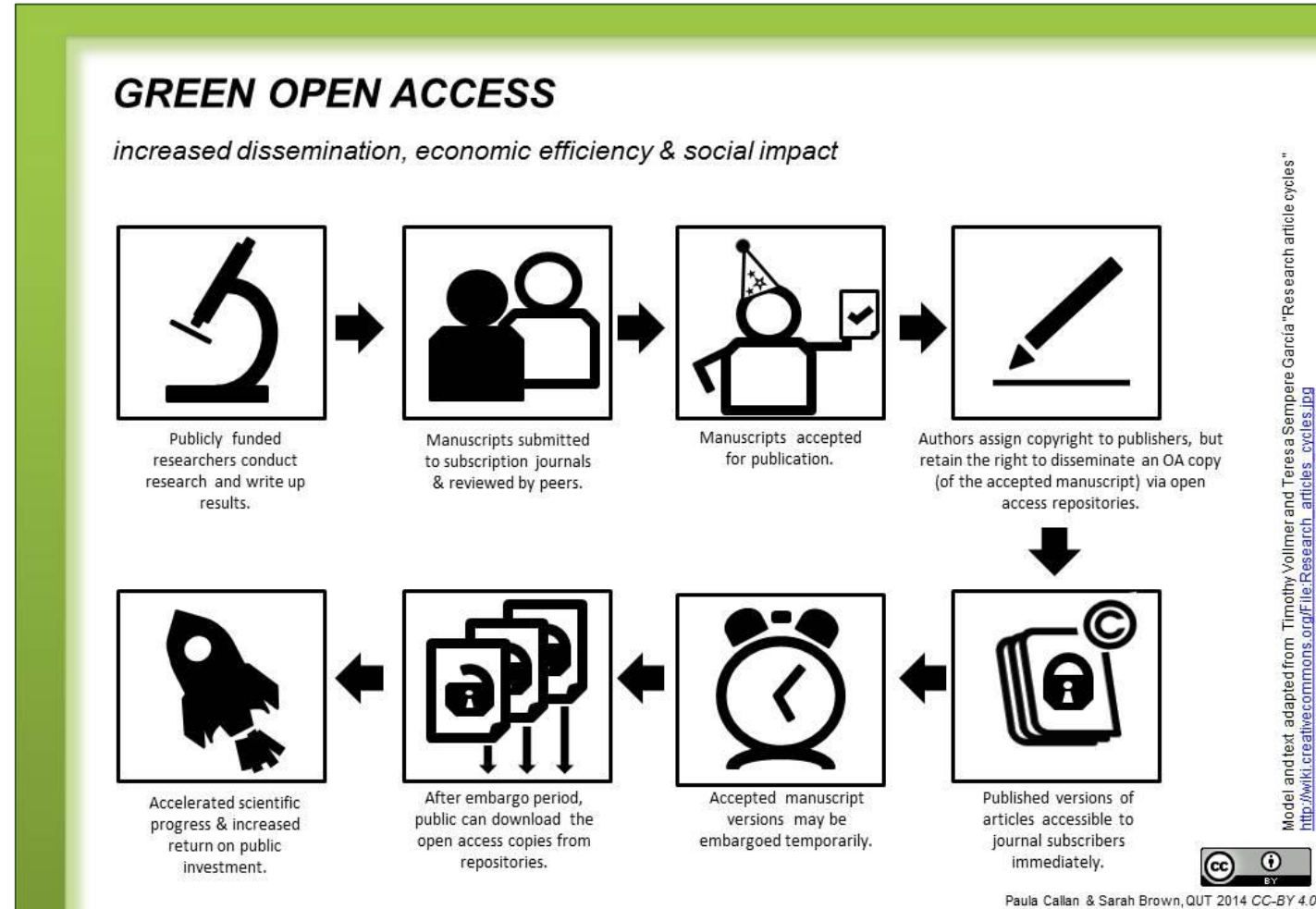
DIGITÁLNY ARCHÍV (Digital Archive)

Organizovaná zbierka digitálnych dokumentov (*born digital* alebo digitalizovaných) zhromažďovaná s cieľom ich dlhodobého uchovávanía a ochrany.

Repozitáre

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:

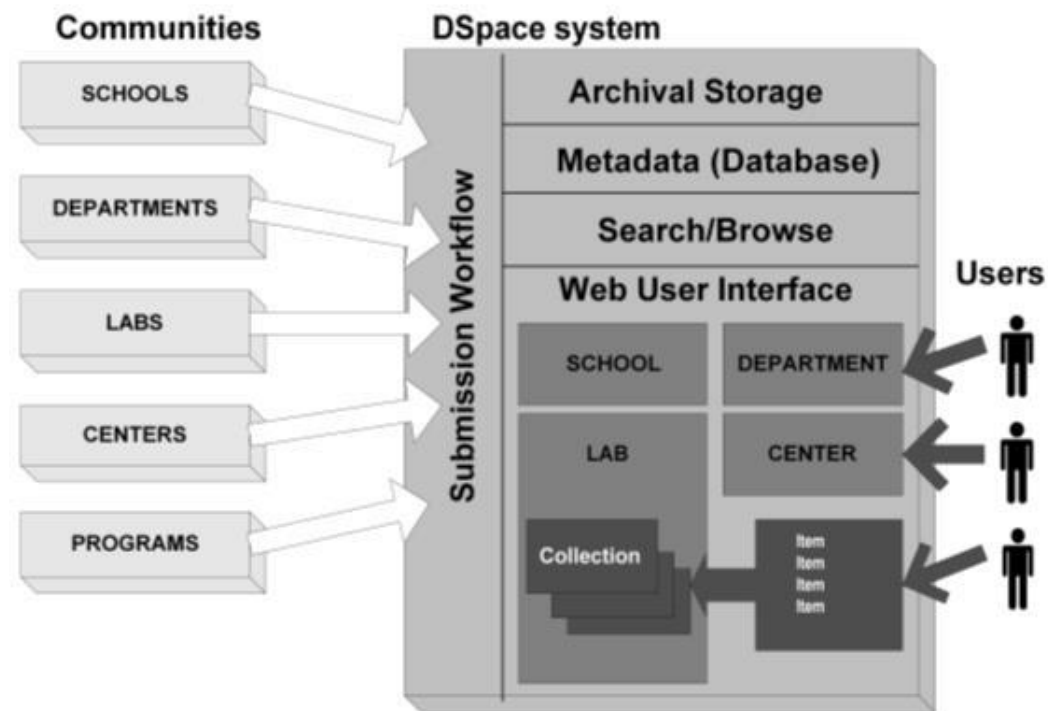
- jedinečný identifikátor (handle)
- vlastnosti, metadáta (properties)
- obsah v bitoch (content)
- digitálny podpis (signature), zabezpečenie autenticity obsahu
- záznamy o použití, vyhľadávaní objektu (transaction log)

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

Hostingové riešenia: napr. SimpleDL, DSpaceDirect.

FUNKCIE REPOZITÁRA:

- **spracovanie** dokumentov/dát
- tvorba **metadát**
- **uloženie a archivácia** digitálneho obsahu
- zabezpečenie **integritu a autenticity**
- **sprístupnenie** dokumentov/dát



Repozitáre

Predmetové OA repozitáre (Subject-based/Disciplinary)

TYPY

Predmetové
(Odborové/tematické)

arXiv.org



Inštitucionálne (*Autoarchivácia + Zelená cesta*)

DR Žilinskej univerzity, repozitár UMB,
Inštitucionálny repozitár SAV, DR Univerzity Komenského



Národné



Multiodborové (univerzálne/všeobecné)



Repozitáre

TYPY

Dátové

(GIGA)ⁿ DB



Softvérové



Iné

protocols.io – úložisko s otvoreným prístupom na zaznamenávanie a zdieľanie podrobných aktuálnych výskumných metód a protokolov.



[Data Repository Guidance - Nature](#)

Predmetové, všeobecné...

Repozitáre

DÁTOVÉ REPOZITÁRE

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

Typy dátových repositárov

Data Warehouse

Data Lake

Data Marts

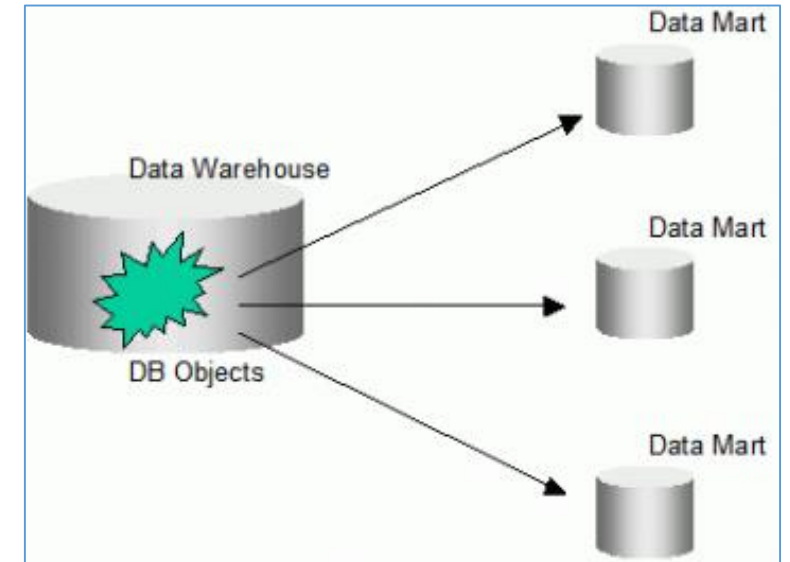
Data Cubes

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 uživatelé	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

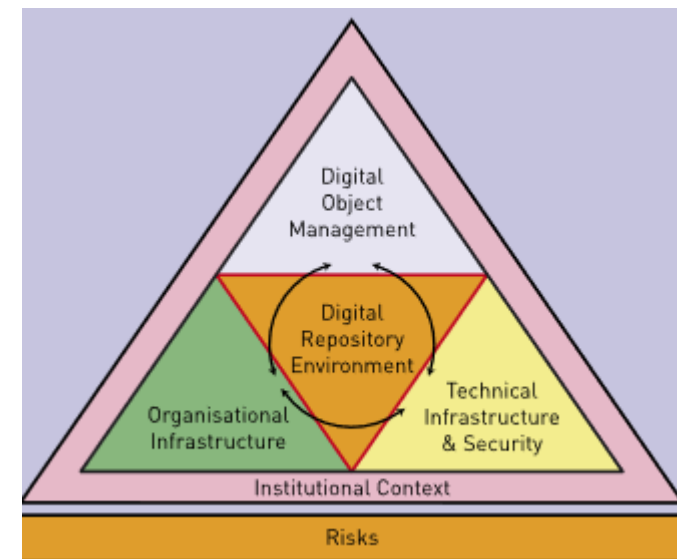
PLATTER (A Planning Tool for Trusted Electronic Repositories)

- 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 českom 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

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

Repozitáre

Ako nájsť vhodný OA repozitár?

- 1) Pri hľadaní repozitára/úložiska pre výskumné dáta si treba vždy overiť, či **existuje nejaký odborový repozitár/repozitár komunity**, kde by sa dáta mohli archivovať.
- 2) Bez ohľadu na vybraný repozitár/úložisko si treba vždy overiť, či je
 - **dlhodobo udržateľné;**
 - **ukladá dáta bezpečným spôsobom;**
 - **zabezpečí, že údaje budú naďalej vyhľadateľné** (pomocou trvalého identifikátora), **prístupné a opakovane použiteľné;**
 - **popisuje údaje štandardným spôsobom** s použitím prijatých štandardov metadát;
 - **umožňuje vkladateľovi určiť licenciu**, ktorou sa riadi prístup k dátam a ich opakované použitie.

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/>)

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/>)

Open Directory of Open Access Repositories,

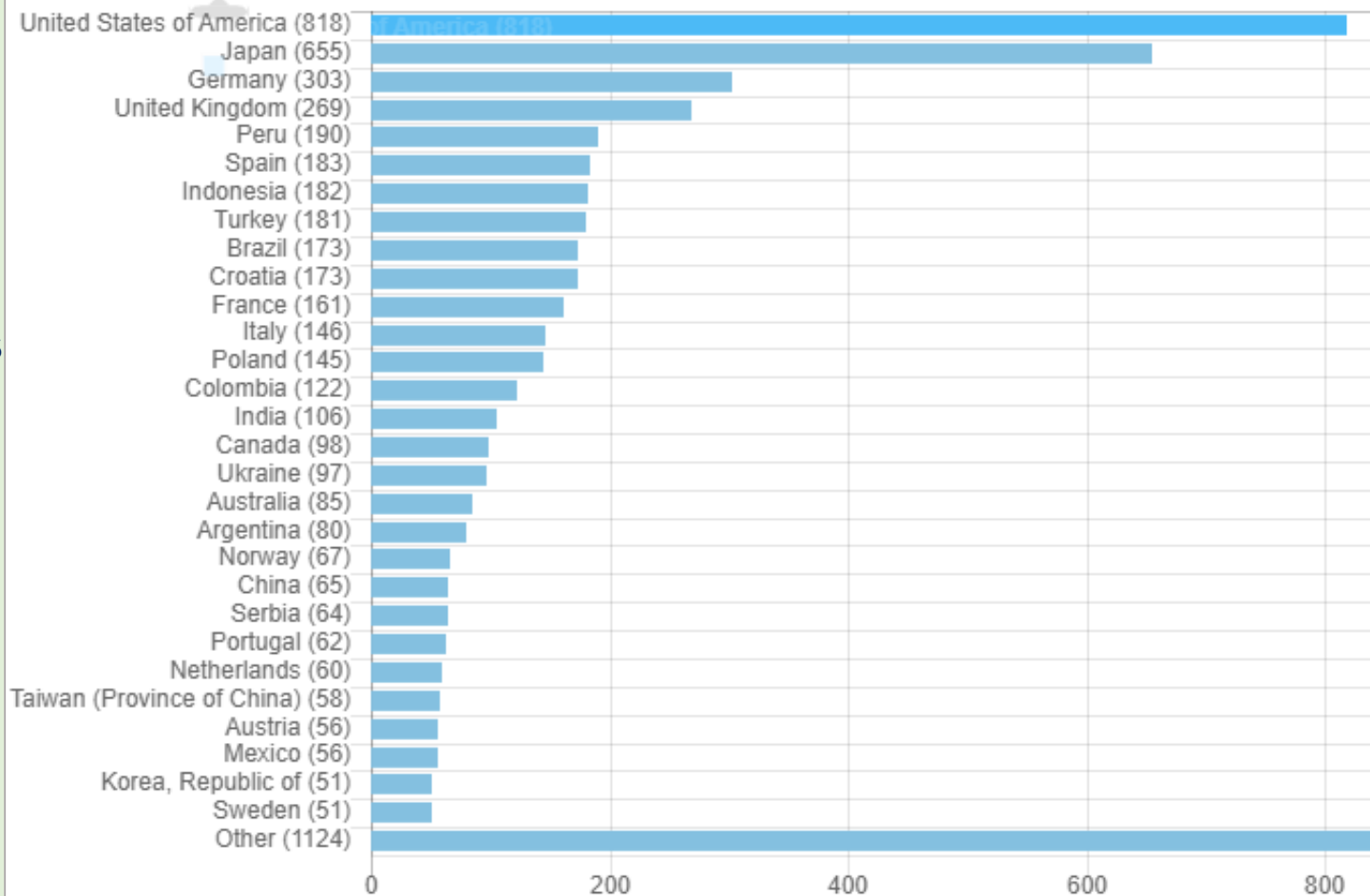
[OpenDOAR](https://v2.sherpa.ac.uk/opensoar/) (<https://v2.sherpa.ac.uk/opensoar/>)

Sherpa Services

[Sherpa Services](#)

(prístup aj k OpenDOAR)

Repositories by Country



Zhrnutie na záver

Prečo archivovať dáta / články / knihy...?

Čo je archivácia?

Čo je tmavý archív?

Aké systémy dlhodobého uchovávania 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...)

Dátové repozitáre

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>
- <https://www.lismcqspractice.com/2020/05/digital-preservation-lockss-clockss.html>
- **Storage, catalogue, repository and archive - what's the difference?** <https://www.umu.se/en/library/research-data/specialised-topics/storage-catalogue-repository-and-archive/>

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:_audity_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 9: 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/usfwsmtnprairie/33462063285> Licencia: <https://creativecommons.org/licenses/by/2.0/>

Snímka 11: https://commons.wikimedia.org/wiki/File:Repository_oaicons.png This file is made available under the [Creative Commons CC0 1.0 Universal Public Domain Dedication](#).

Snímka 13: Green Model: <https://bond.libguides.com/open-access-and-scholarly-publishing/publishing>

Snímka 6, 7, 8, 9, 10, 15, 16, 19, 20, 21: logá a obrázky – printscreeny - webové stránky inštitúcií

Ďakujem za pozornosť.

Kontakt:

gabriela.fisova@cvtisr.sk

otvorenaveda@cvtisr.sk

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

