

Archivácia a repozitáre

Elektronické informačné zdroje pre vedu - Publikačný poradca

Mgr. Gabriela Fišová

Centrum vedecko-technických informácií SR



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



Odpovedať na otázky:

Archivácia – pamäť vedeckého záznamu

Čo sa stane s obsahom časopisu, keď vydavateľ skrachuje?

Archivačné iniciatívy

Repozitáre – všeobecná charakteristika, typy, príklady

Repozitár – obsah, funkcie, softvér, dôveryhodnosť

Výber repozitára – registre repozitárov, príklady

Dátové repozitáre – všeobecná charakteristika, typy

Ukážky repozitárov + linky

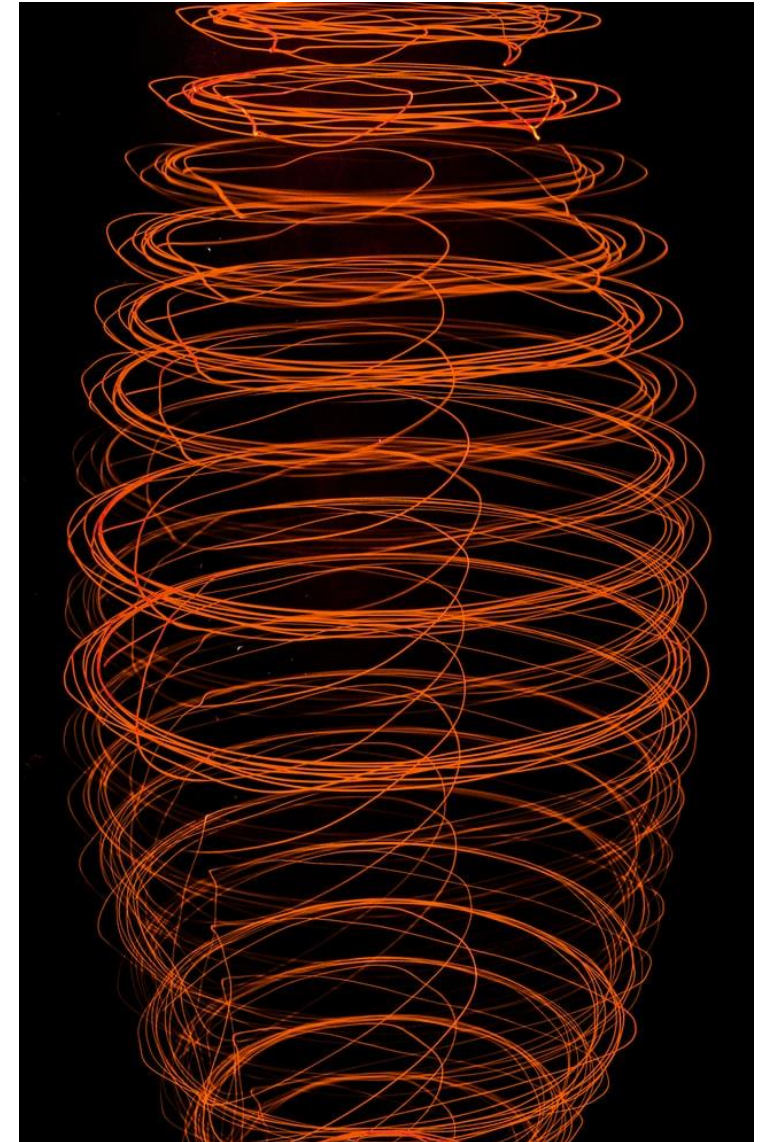
Archivácia – pamäť vedeckého záznamu

- 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...



Archivácia

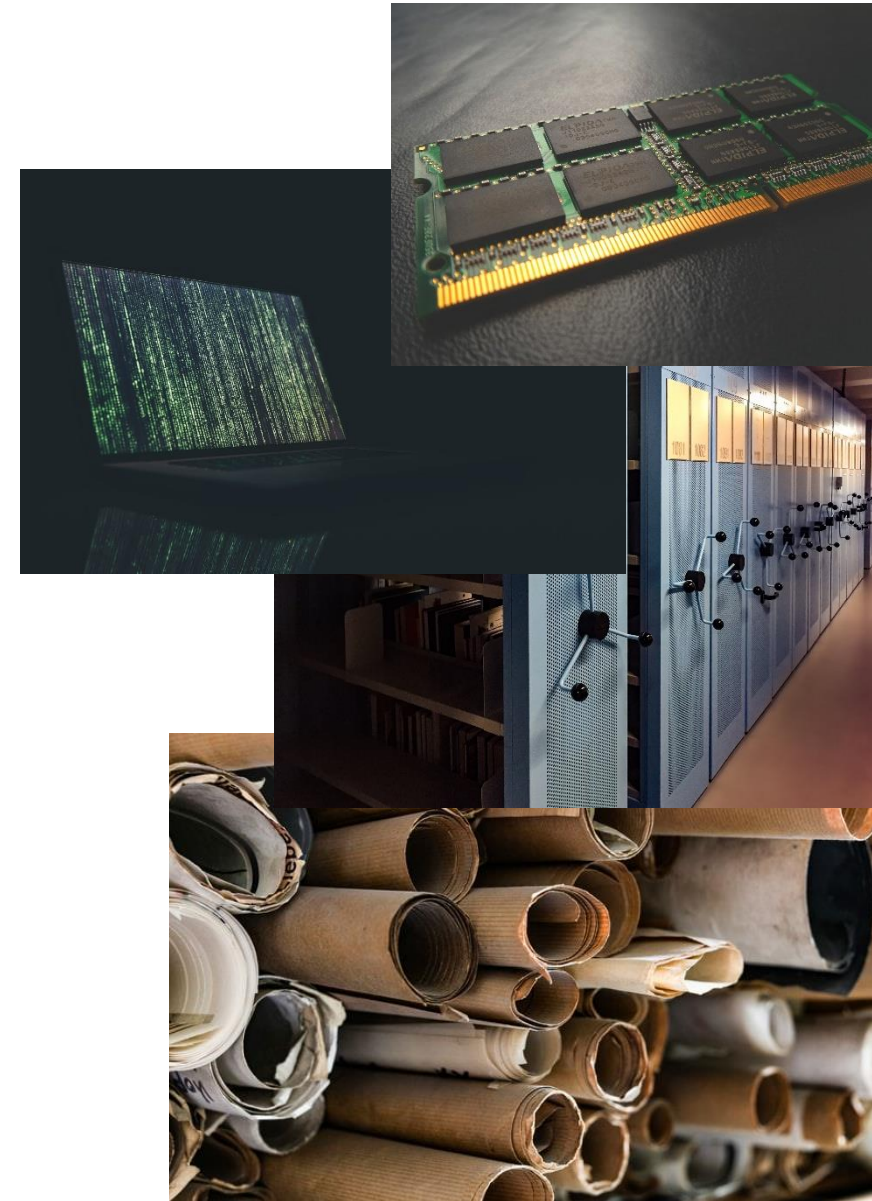
Archivácia – proces podporujúci dlhodobé **uchovávanie** a **ochranu** dát/dokumentov a **metód/postupov** ich **čítania** a **interpretácie**

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).

Seriózni vydavatelia sa zapájajú do archivačných iniciatív

 **Portico** (<https://www.portico.org/>) – **tmavý archív** (*dark archive*) 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/>) – 12 mil. článkov v časopisoch, 100000 kníh, milióny obrázkov **JSTOR** (images) a primárne zdroje v 75 disciplínach. Prístup aj k otvorenému obsahu – 500000 článkov vo verejnej sfére (public domain)...





- 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 na báze softvéru s otvoreným zdrojovým kódom, ktorý kontroluje viaceré kópie dokumentu voči sebe navzájom, aby odhalil náhodne vzniknuté chyby (alebo prípadné pokusy o zmenu či potlačenie materiálu, ktorý bol vydaný. Keby bola na internete len jedna kópia pod jednou administráciou, bola by oveľa zraniteľnejšia voči diverzii či cenzúre).
- [CLOCKSS](#) (*Controlled LOCKSS*, <https://clockss.org/>) –túto službu si platia najmä veľkí vydavatelia (u ktorých je malá pravdepodobnosť zániku, preto zatiaľ zverejnila len malú časť archívov)



Pojem **digitálny repozitár** zaviedli v roku 1995 [Robert Kahn a Robert Wilensky](#)

- definovali teoretické princípy otvoreného univerzálneho informačného systému, známeho pod názvom Kahn-Wilensky Framework (KWF)
- definovali komponenty otvoreného systému na uchovávanie, prístup a manažment (organizáciu, riadenie) informácií, **pričom architektúru systému oddeľujú od uchovávaného obsahu**, čím je ich prístup univerzálny

Digitálny archív (Digital Archive) - organizovaná zbierka digitálnych dokumentov (*born digital alebo digitalizovaných*) zhromažďovaná s cieľom ich dlhodobého uchovávania a ochrany. Súčasťou môže byť sprístupňovanie cez sieťové rozhranie.

Digitálny repozitár/úložisko (Digital Repository/Storage) - systém na prevádzku súboru služieb, ktoré sa zaoberajú získavaním, správou a sprístupňovaním digitálneho obsahu. *Autoarchivácia + Zelená cesta*

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



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

Typy repozitárov

Predmetové/Odborové (tematické) - zameriavajú sa na konkrétnu vednú oblasť (napr. **arXiv**, oblasť fyziky, od r. 1991)

Inštitucionálne - 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. *Autoarchivácia + Zelená cesta*

Národné - obsahujú publikačné výstupy krajiny

Multiodborové (univerzálne/všeobecné) - napr. Zenodo

Dátové, softvérové,

arXiv.org



INŠTITUCIONÁLNY REPOZITÁR SAV

Repozitár - obsah

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

Digitálny objekt obsahuje:

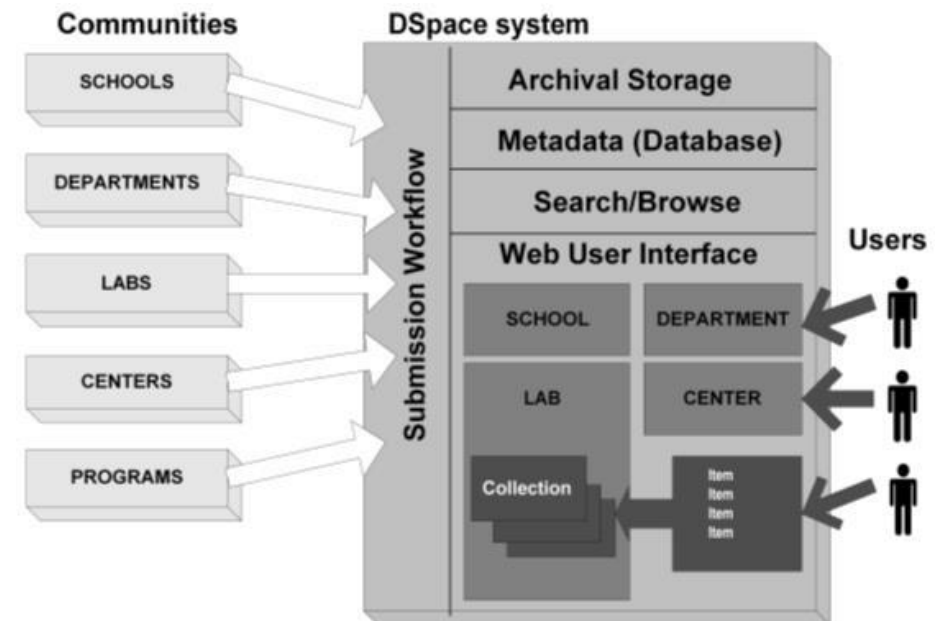
- Jedinečný identifikátor (**handle**)
- Vlastnosti, metadáta (properties)
- Obsah v bitoch (content)
- Digitálny podpis (signature), zabezpečenie autenticity obsahu
- Záznamy o použití, vyhľadávaní objektu (transaction log)

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
- Sprístupnenie dokumentov



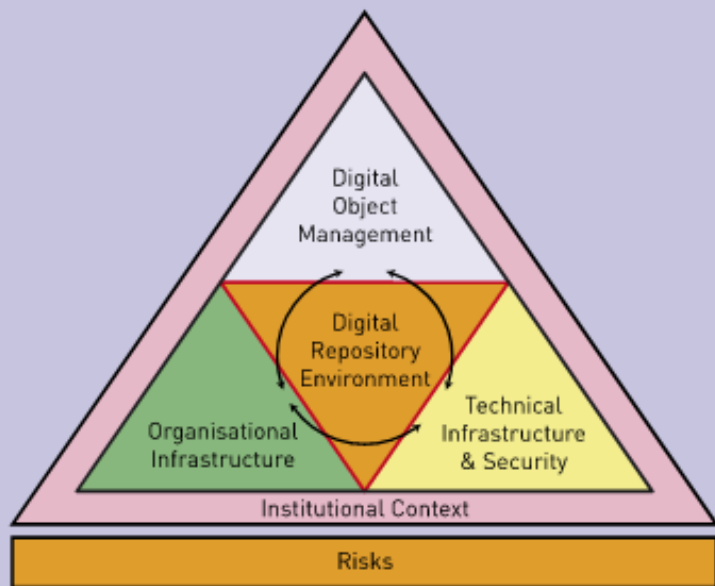
Repozitár - dôveryhodnosť

DRAMBORA (Digital Repository Audit Method Base on Risk Assessment)

- ponúka kvantifikovateľný pohľad na závažnosť rizík, ktorým repozitáre čelia, a efektívne prostriedky na reportovanie týchto rizík

PLATTER (Planning Tool for Trusted Electronic Repositories)

- plánovanie aktivít repozitárov v kľúčových oblastiach, tak aby splnili svoje ciele
- dopĺňa existujúce iniciatívy pre audit a certifikáciu – napr. DRAMBORA



<http://www.repositoryaudit.eu/>

DRAMBORA interactive

Digital Repository Audit Method
Based on Risk Assessment



Register for DRAMBORA

Username:

Password:

Remember: ☐ [Log In](#)

[Forgot Password?](#)

[Home](#)

[Latest News](#)

[Get Expert Help](#)

[Having Problems?](#)

[Download Offline Version](#)

[Submit Feedback](#)

[DRAMBORA Training](#)

[About](#)

[Objectives](#)

[Benefits](#)

[The DRAMBORA Team](#)

[Collaborations](#)

[Dissemination](#)

[DRAMBORA Users](#)

Welcome to DRAMBORA Interactive: Log in or Register to Use the Toolkit

Welcome to DRAMBORA Interactive



Please register, log in or read [about the toolkit](#), get an overview of its [objectives](#), or learn of the [benefits](#) that it can offer you and your repository.



The [Digital Curation Centre \(DCC\)](#) and [DigitalPreservationEurope \(DPE\)](#) are delighted to announce the release of the Digital Repository Audit Method Based on Risk Assessment

(DRAMBORA) toolkit. This toolkit is intended to facilitate internal audit by providing repository administrators with a means to assess their capabilities, identify their weaknesses, and recognise their strengths. Digital repositories are still in their infancy and this model is designed to be responsive to the rapidly developing landscape. The development of the toolkit follows a concentrated period of repository pilot audits undertaken by the DCC, conducted at a diverse range of organisations including national libraries, scientific data centres and cultural and heritage data archives.

For more information, why not download the DRAMBORA Flyer in [PDF](#) or [JPEG](#) formats?

News

[DRAMBORA Translated into Japanese](#)
13th Sep 2010

[DRAMBORA on LinkedIn](#)
30th Nov 2009

[DRAMBORA is part of the SHAMAN Assessment Framework](#)
1st Jun 2009

Download

Please [register](#) to download the offline copy of the DRAMBORA Toolkit

Comment

Submit feedback via our [feedback form](#) or [email us](#).

Get Help

Get help in the DRAMBORA evaluation process by volunteering to host a [DRAMBORA facilitated assessment](#).

DRAMBORA Consortium Sponsors:



Dôveryhodnosť repozitárov – referenčný model OASIS

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

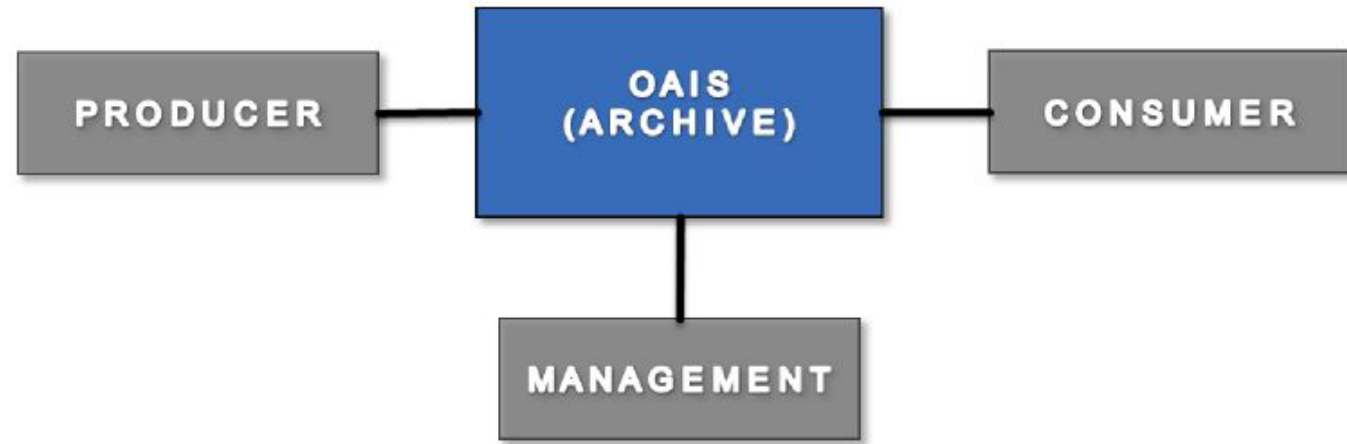
Repozitár by mal spĺňať **10 podmienok**:

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 **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.

Dôveryhodnosť - certifikácia

Štandard: OASIS

Externý model OASIS



Zdroj: https://wikisofia.cz/w/images/3/36/Extern%C3%AD_entity.png

Funkčný model OASIS

Príjem – príjem dátového balíka od producenta, príprava na uloženie

Archívne úložisko – služby a funkcie ukladania/vkladania, údržba, vyhľadávanie

Plánovanie ochrany – zaistenie dlhodobej dostupnosti a zrozumiteľnosti informácií používateľom

Správa dát – príjem, údržba, prístup k popisným informáciám a administratívnym dátam

Prístup – poskytnutie prístupu do digitálneho archívu používateľom

Administrácia – služby a funkcionality pre zabezpečenie celkovej prevádzky archívu

Repozitáre

Ako nájsť vhodný repozitár?

Medzinárodné registre repozitárov

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/opendoar/)

(<https://v2.sherpa.ac.uk/opendoar/>)

Registry of Research Data Repositories,
[re3data](https://www.re3data.org/), (<https://www.re3data.org/>)

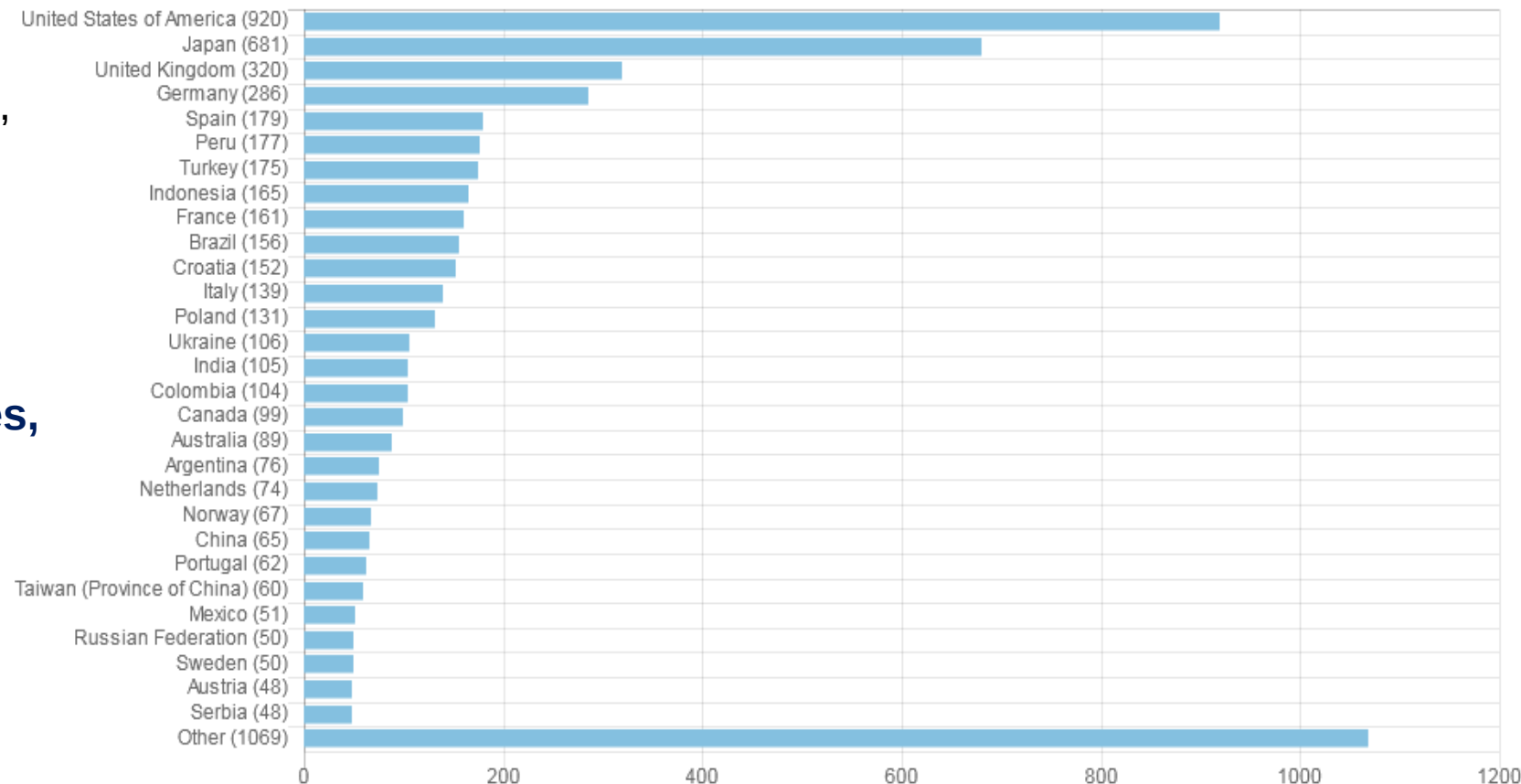
Registry of Open Access Repository
Mandates, [ROARMAP](http://roarmap.eprints.org/), (politiky,

<http://roarmap.eprints.org/>)

[SherpaRomeo](https://v2.sherpa.ac.uk/romeo/)

(<https://v2.sherpa.ac.uk/romeo/>)

-umožňuje overiť, ktorá verzia článku sa
môže uložiť do repozitára



OpenDOAR Statistics, Repositories by Country

https://v2.sherpa.ac.uk/view/repository_visualisations/1.html

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

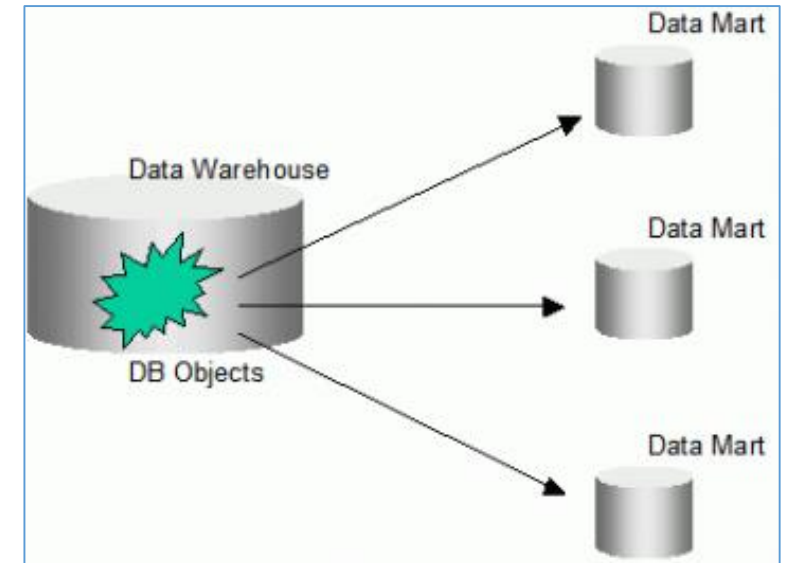
Data Cubes

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 uživatelé	Datoví vědci



<https://www.webdepozit.sk/>

Budovanie informačného systému na zber (webharvesting), spracovanie, sprístupnenie a dlhodobú archiváciu slovacikálnych webových prameňov a pôvodného elektronického (e-Born) obsahu. Realizuje zber webových stránok formou rôznych typov kampaní a archiváciu elektronických seriálov a monografií.





Vyhľadávanie vo webovom archíve

☒ www pramene ☐ elektronické publikácie



SYSTÉM DIPARCHÍVY A KATALÓGY DIPWWW PRAMENE ELEKTRONICKÉ PUBLIKÁCIEPROJEKT DIPKONTAKTYENGLISH

Spoločne chránime **digitálne** dokumentárne dedičstvo



Archivujte s nami

Navrhujem archivovať webové sídlo

Navrhujem archivovať elektronickú publikáciu

Webdepozit v číslach



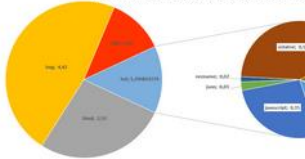
Voľné 768 TB
Využitá 32 TB
Celkovo 800 TB

Profil archívu

Slovenský webpriestor

V rámci **celoplošného zberu 2020** sme z katalógu DDP zbierali **594 434 domén**.

Počet registrovaných SK domén je stále rastúci. Ich počet možno zistiť na stránke správcovskej spoločnosti SK-NIC.



Zozbieraný objem Celoplošného zberu 2018 v TB

Štatistika domén

Informácie

Novinky

WWW kampane (zbery)

Elektronické publikácie

Vaše otázky

Národný projekt Digitálna knižnica a digitálny archív sa zameriava na masovú digitalizáciu knižničných a archívnych fondov a na ochranu písomného kultúrneho dedičstva (papieru) pred nevyhnutnou degradáciou.

Kategórie

<https://dikda.sk/>

<http://archiv.vm.stuba.sk/vademecum/>

STU SLOVENSKÁ TECHNICKÁ
UNIVERZITA V BRATISLAVE

ÚVOD PRECHÁDZAŤ A HĽADAŤ

Slovenská technická univerzita v Bratislave

+ [Fondy, zbierky a knižnica](#) (624)

Digitálny archív Archívu Slovenskej technickej univerzity v Bratislave bol spustený vo februári 2017 v roku 80. výročia vzniku Slovenskej technickej univerzity. V súčasnosti prezentuje výsledky prvej etapy digitalizácie fondov a zbierok Archívu Slovenskej technickej univerzity v Bratislave, do ktorej boli zahrnuté systémové okruhy najvzácnejších a najviac ohrozených archívnych dokumentov z rokov 1938 až 1980.

🔍

📄 Vyhľadáte v kategórii: **Slovenská technická univerzita v Bratislave**

📄 Typ dát:

Digitizované archívne dokumenty

📄 Umiestnenie hľadaného výrazu:

Kdekoľvek Na začiatku Na konci

Celé slovo

📄 Viacslovný výraz:

a zároveň aspoň jeden ako fráza

Prečo archivovať dáta / články / knihy...?

Čo je archivácia?

Čo je to tmavý archív?

Aké systémy dlhodobého uchovávania majú vydavatelia k dispozícii?

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

Repozitár - obsah

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

Výber repozitára – registre repozitárov

Dátové repozitáre – charakteristika, typy (data warehouse, data lake...)

Ukážky repozitárov + linky

Zaujímavé 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://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

Použité zdroje:

Zdroje uvedené na jednotlivých snímkach a prezentácie:

- Archivácia a manažment výskumných dát, Zuzana Stožická. CVTI SR. <https://eiz.cvtisr.sk/kurzy/eiz-pre-vedu-publikacny-poradca/>
- Manažment výskumných dát. Archivácia a repozitáre. Ľubica Jamborová. CVTI SR. <https://eiz.cvtisr.sk/kurzy/eiz-pre-vedu-publikacny-poradca/>
- 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 4: 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](#)

RAMs: Photo by [Possessed Photography](#) on [Unsplash](#)

Snímka 4, 5, 7, 10, 16, 17: logá a obrázky – scan - webové stránky inštitúcií

Snímka 19: Photo by [Shazaf Zafar](#) on [Unsplash](#)

Ďakujem za pozornosť

