Adam Bodnár, 2. ročník

25. november 2019 ,Brno

Zenodo - sharing Research Data across Europe, CERN

<https://zenodo.org/>

Zenodo je projekt založený a vyvíjaný vedcami, ktorý podporuje myšlienku hnutia Open science. A teda voľné šírenie vedeckých publikácií, dát, poprípade zdieľanie výsledkov výskumov. V snahe sprístupniť vedecké výstupy, hlavne okruhu vedcov, ale aj širokej verejnosti.

Pod záštitou OpenAire project, v spolupráci s CERN-om, priekopníkom v oblasti open source a open access, bol projekt spustený v máji 2013 pod názvom OpenAire orphan records repository. Dva roky po tom bol znova spustený ako Zenodo. Zenodo predstavuje univerzálny repozitár pre všetky druhy dát získaných z výskumov, bez ohľadu na veľkosť či formát. Umožňuje dlhodobé uchovanie, vyhľadávanie, zdieľanie alebo vzájomné hodnotenie vedeckých prác. Vďaka tomu je v dnešnej dobe využívaný množstvom výskumníkov, vedcov a odborníkov z rôznorodých odvetví po celom svete.

Ďalšou z výhod môže byť aj jednoduchosť publikácie prác. Po prihlásení, či už priamo na stránkach Zenodo, alebo po prepojení cez GitHub, poprípade ORCID, organizáciu snažiacu sa zabezpečiť aby každý kto prispieva rozvoju vedy mal unikátnu identifikáciu, stačí súbory nahrať. Kvôli prehľadnosti a zároveň lepšiemu dohľadaniu je potrebné dodať popis. Nakoniec stačí už len zvoliť variantu prístupu a to: otvorený, uzavretý na dobu určitú alebo úplne uzavretý. Preto môže byť veľmi užitočným nástrojom aj pre aktuálne prebiehajúce výskumy.

Aktuálne je v repozitáry nahratých približne 1 419 250 súborov. Z toho 1 391 297 s otovreným prístup a len 25 325 s prístupom uzavretým. Nadpolovičnú väčšinu tovria publikácie (viac ako 800 000) nasledované obrázkami (viac ako 500 000), prekvapivo netriviálne zastúpenie má nahratý software(viac ako 300 000 príspekov). Nachádza sa tu okolo 3850 komunít, vytvorených a udržiavaných užívaťeľmi, zameraných na rôznorodé oblasti vedy.

Všetky nahrané dokumenty je možné po uvedení dôvodu z repozitára odobrať. Dokumenty sú verzované, čo znamená ľahký prístup k predošlým verziám napr. v prípade deformácie dát. Originály nahratých súborov sú uložené v dátových centrách CERN-u, ku ktorým priebežne vznikajú kópie na prevenciu straty informácii. Čo sa týka ochrany osobných údajov, tieto údaje nie sú žiadnym spôsobom zbierané, ani zdieľané tretím stranám. Na rozdiel od toho Zenodo môže poskytovať neosobné informácie, ako súbory cookies či IP adresy, svojim partnerom.

Nezanedbateľnými faktormi sú určite aj podpora citácie publikovaných prác, možnosť odkazovania na jednotlivé publikácie, ako aj hodnotenie od ostatných užívateľov. Toto všetko tvorí ideálne prostredie pre získanie kreditu a postavenia v rámci vedeckej spoločnosti. Súčasne oddeľuje dobre spracované a kvalitné informácie od tých menej významných, čím uľahčuje vyhľadávanie.

Prístup, ktorý tvorcovia Zenoda zaujali k otázke, aké formáty dát budú podporované, bol veľmi priamočiary. Na formát sa nekladú žiadne nároky. Toto rozhodnutie bolo bezpochyby užívateľsky prívetivé, keďže sa stráca potreba často náročnej konverzie dát medzi rôznymi formátmi. Taktiež výrazne zväčšuje škálu potencionálne nahratých digitálnych objektov, čím poskytuje popis aj toho najmenšieho detailu, ktorý viedol k úspešnému ukončeniu experimentu či vydania vedeckej publikácie. Od samotnej publikácie v textovej forme cez obrázok schémi zapojenia batérie až po audio nahrávku autorových poznámok. Toto všetko môže prispieť jasnejšiemu pochopeniu myšlienok autora/autorov.

Na druhej strane takáto voľnosť prináša komplikácie v oblasti uchovania veľkého množstva niekedy diametrálne odlišných súborov a ich spätného dohľadania. Konkrétne aké takzvané metadáta, nesúce informácie o primárnych dátach, budú ukladané na zaistenie rýchleho a spoľahlivého vyhľadávania. Zenodo sa riadi smernicami FAIR data principles, ktoré popisujú štyri hlavné body, ktorými sa majú metadáta zaistiť.

A to:

1. Nájditeľnosť

* Zenodo prideľuje každému nahratému objektu DOI
* udržuje bohaté opisujúce metadáta v súlade s DataCite's Metadata Schema
* metadáta každého objektu sú indexované a dohľadateľné hneď po publikácii

1. Dostupnosť

* metadáta sú získateľné pomocou OAI-PMH protokolu ako aj REST API
* oba sú to univerzálne open access protokoly
* všetky metadáta sú verejne prístupné
* udržiavané dlhodobo, v súčasnosti minimálne 20 rokov

1. Interoperabilita

* JSON Schema na reprezentáciu metadát
* možnosť prevodu do Dublin Core, MARCXML

1. Opätovná použiteľnosť

* dáta sú voľne dostupné, môžu byť modifikované a zdieľané, ak autor neurčí inak
* vďaka súladu s DataCite's Metadata Schema nie sú metadáta iba Zenodo špecifické.

Spracovanie tak výrazného množstva informácií vyžaduje technickú podporu. Súbory sú ukladané pomocou služby EOS, ktorú poskytuje CERN. EOS je diskový úložný priestor s nízkou odozvou a kapacitou presahujúcou 150 petabajtov. Najväčšiu časť zaberajú dáta z Veľkého Hadronovho urýchľovača. Repozitáre Zenoda sú rozmiestnené na viacerých diskoch, celkovo zaberajúc 18 petabajtov.

Z pohľadu užívateľov, ktorí využívajú repozitár na úschovu svojich dlhoročných, často finančne náročných výskumov či experimentov, je dôležitou vlastnosťou bezpečnosť. Ako sa Zenodo vysporiadava s potencionálnou stratou alebo korupciou už bolo popísané vyššie. Akými ochrannými mechanizmami ale disponuje v prípade pokusu o neoprávnený prístup k súborom, ktoré nie sú verejne prístupné?

V prvom rade pri pokuse o „útok z vonku“. Prístup k serverom Zenoda má len špeciálne školený personál. Najaktuálnejšiu ochranu zabezpečuje systém Puppet, ktorý automaticky aktualizuje všetky bezpečnostné prvky. Pre ukladanie hesiel užívateľov sa používa silný šifrovací algoritmus.

V druhom rade pri „útoku z vnútra“. Fyzický prístup k dátam má len veľmi malá časť zamestnancov CERN-u. Dôležité je spomenúť, že obsah súborov nie je šifrovaný. Preto je na prístup k dátam nutné povolenie, zväčša za účelom opravy problémov.

Na záver je nepochybne zaujímavé, aké pokroky sa chystajú smerom do budúcnosti. Jednou z vecí je zlepšenie citácie z externých repozitárov ako napr. NASA ADS. Ďalej sa pracuje na drobných vylepšeniach vyhľadávania a spracovania údajov. Všetko v snahe čo najviac uľahčiť prácu užívateľom.

Otvorený prístup je v tak dynamickej sfére akou je veda nesmierne dôležitý. Zenodo je projekt, ktorý uľachčuje prácu vedcom vo veľa smeroch. Uchovaním ich dokumentov ich zbavuje starosti z prípadnej straty informácii. Hodnotiacim systémov tlačí na tvorbu kvalitných publikácii a nie len, takých, ktoré by sme mohli označiť „do počtu“. Zároveň ich tým zviditeľňuje a zvyšuje ich reputáciu. Umožnuje im tvoriť komunity kde môžu spoznávať ľudí z celého sveta a nadväzovať tak kontakty, ktoré by inak nezískali. Aj vďaka tomu ho môžme považovať za úspešný projekt, ktorý stojí za vyskúšanie.

**Použité zdroje:**

<https://about.zenodo.org/>

<https://www.openaire.eu/zenodo-guide>

<https://orcid.org/about>

<https://puppet.com/products/how-puppet-works>

<https://cern.service-now.com/service-portal/service-element.do?name=eos-service>

<https://sciencenode.org/feature/cern-and-openaireplus-launch-new-european-research-repository.php>

<https://zenodo.org/search?page=1&size=20&sort=mostviewed>

**Použité skratky:**

**CERN -** European Organization for Nuclear Research, <https://home.cern/>

**ORCID -** Open Researcher and Contributor ID, <https://orcid.org/about>

**DOI -** Digital Object Identifier, <https://www.doi.org/faq.html>

**OAI-PMH -** The Open Archives Initiative Protocol for Metadata Harvesting, <https://www.openarchives.org/pmh/>

**REST API -** REpresentational State Transfer application program interface, <https://restfulapi.net/>

**JSON Schema** - Javascript Object Notation schema, <https://json-schema.org/>

**Dublin Core –** štandart pre metadatový popis , <https://www.dublincore.org/about/>

**MARCXML -** Machine-readable cataloging extensible markup language, <http://www.loc.gov/standards/marcxml/>

**DataCite's Metadata Schema –** zoznam vlastnosí metadát, slúži na konzistentnú identifikáciu v prípade napr. citácie, <https://datacite.org/mission.html>

**EOS –** storage device,<https://cern.service-now.com/service-portal/service-element.do?name=eos-service>

**NASA ADS-** National Aeronautics and Space Administration Astrophysics Data System, <https://ui.adsabs.harvard.edu/>

**Metadata v Dublin Core:**

<dc:title> Zenodo - sharing Research Data across Europe</dc:title>

<dc:creator>Adam Bodnár</dc:creator>

<dc:date>25.11.2019</dc:date>

<dc:type>essay</dc:type>

<dc:format>.docx, 25KB</dc:format>

<dc:language>sk</dc:language>