

OPEN SCIENCE V AKADEMII VĚD ČR

Článek popíše připravenost Knihovny AV ČR a aktivity Akademie věd ve stále častěji zmiňované oblasti Open Science. Budou zmíněny možnosti, které nabízí datový repozitář ASEP, příprava dat, proces jejich ukládání a zobrazování. U přípravy dat bude představen Data management plan, který je nedílnou součástí čerpání financí z evropských fondů. Další důležitou složkou cirkulace dat jsou tzv. FAIR principy, které jsou úzce spjaty s EOSC. Tyto principy budou v článku představeny, dále budou okrajově zmíněny možnosti certifikace datových repozitářů a nakonec budou čtenáři seznámeni s aktivitou České republiky v EOSC. Na úplný závěr bude představen užitečný Open Science nástroj „re3data.org“, jedná se o registr datových repozitářů.

Termín Open Science (neboli Otevřená věda) je v poslední době diskutován čím dál častěji a dostává se do popředí nejen na akademické půdě. V oblasti knihovnictví a informačního sektoru je kladen důraz hlavně na přístup k otevřeným datům (Open Data) a otevřeným publikacím (Open Access). Skupina Open Science však zastřešuje i další problematiky jako Citizen Science, Open Peer Review, Open Clouds, Open Education, aj.

Podle Evropské komise lze termín Open Science definovat jako nový přístup k vědeckým procesům založený na spolupráci a nových způsobech šíření znalostí prostřednictvím digitálních technologií. Organizace pro hospodářskou spolupráci a rozvoj (OECD) definuje tento termín jako veřejné zpřístupnění primárních výstupů výzkumu, který je financován z veřejných zdrojů – tedy publikací a vědeckých dat – v digitální formě s minimálním omezením nebo bez něj [1].

Na základě datové politiky jsou vědci čerpající evropskou finanční podporu (např. z programu H2020) povinni zveřejňovat vědecká data a z nich uskutečněné publikační výsledky. Z této povinnosti je možné se v ojedinělých případech vyvázat (tzv. opt-out, týká se např. ochrany osobních údajů, know-how a dalších). Data lze zveřejnit prostřednictvím obecných, oborových nebo institucionálních repozitářů.

DATOVÝ REPOZITÁŘ ASEP

V Akademii věd ČR byla v roce 2010 ustanovena „Politika otevřeného přístupu k vědeckým informacím“ [2]. Knihovna Akademie věd ČR (dále jen „Knihovna AV ČR“) je od roku 1993 administrátorem systému ASEP, jehož základem je bibliografická databáze, v níž jsou uloženy záznamy všech zásadních výsledků základního výzkumu, které vznikly na pracovištích Akademie věd České republiky (AV ČR) [3]. Od roku 2012 je možné k jednotlivým vědeckým výstupům připojovat plné texty v různých režimech zpřístupnění. Vybrané záznamy

z repozitáře jsou sklíženy do databáze OpenAIRE pomocí protokolu OAI-PMH. Knihovna AV ČR monitoruje situaci vývoje a je si vědoma nutnosti nabídnout vědcům služby neustále se rozvíjícího výzkumného procesu. Systém ASEP byl v roce 2017 rozšířen o nadstavbu umožňující ukládat datové záznamy.

Životní cyklus dat (viz. Obr. 1), podle e-learningové open science platformy FOSTER, je definován takto:

- Tvorba, příprava a získávání dat
- Zpracování dat
- Analýza dat
- Dlouhodobé uchování dat
- Zpřístupnění dat
- Opětovné využití dat

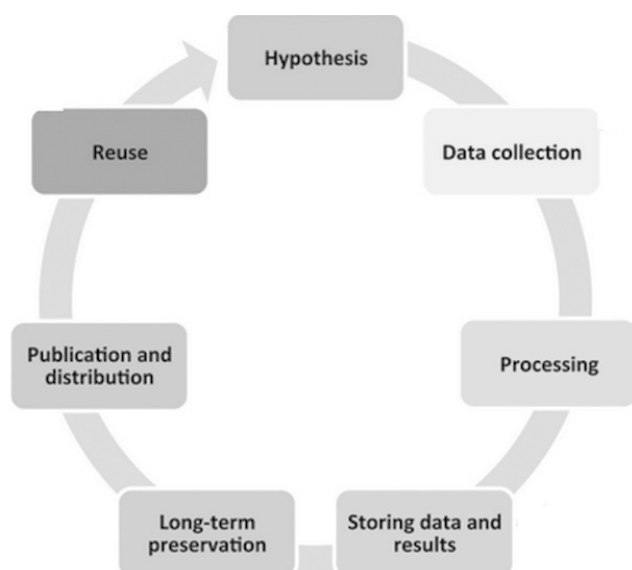
Knihovna AV ČR vstupuje do tohoto cyklu v jeho druhé polovině, kdy nabízí vědcům AV ČR prostor pro uložení dat, jejich publikování a samozřejmě dalším zájemcům znovupoužití dat. Jedná se o repozitář multioborový a je možné vložit různé datové soubory, tzv. datasets. Přes webové formuláře zapisují autoři metadata, která splňují mezinárodní standardy. Systém defaultně nabízí výběr z licencí Creative Commons. Pokud nejsou dostačující, má autor možnost přiložit svou vlastní licenci. Vložit samotná data lze dvojím způsobem:

1. dataset nahrát přímo do ASEP k záznamu
2. k záznamu v ASEP přiložit perzistentní odkaz

Počet bibliografických záznamů	293 068
Počet druhů výsledků	30
Roční přírůstek záznamů	cca 12 000
Počet plných textů	18 242
Veřejně přístupné	9 922

Tab. 1: Počty záznamů a druhů dokumentů uložených v repozitáři ASEP, stav ke dni 25. 10. 2019.

Zdroj: <https://asep.lib.cas.cz/ar1-cav/cs/index/>



Obr. 1: Životní cyklus dat. Zdroj: <https://www.fosteropenscience.eu/content/what-open-science-introduction>

1.
0506600 - UIVT-O (2019) **DATA Vědecká data**
Jeřábek, M. - Šrotýř, M. - Tvrdý, K. - Brabec, Marek - Kasanický, Ivan - Konár, Ondřej - Malý, Marek - Pelikán, Emil - Mlynářová, T. - Zaoral, O. - Termer, J. - Rak, J. - Nosek, V. - Vajda, P. - Sklenář, P.
 PrePark.
 [PrePark.]

PrePark je softwarový statistický nástroj, který na základě dat z mýtného systému predikuje obsazenost jednotlivých parkovacích míst na dálnicích v ČR.
 Klíčová slova: toll system * prediction * availability of parking spaces * highway
 Grant CEP: GA TA ČR TA02031411
 Institucionální podpora: RVO:67985807
 Obor OECD: Transport engineering
 Trvalý link: <http://hdl.handle.net/11104/0297820>

Data v jiném repozitáři: [Úložiště ÚI](#)

Publikace ASEP:
 PrePark

Obr. 2: Propojení datového a bibliografického záznamu v repozitáři ASEP

do úložiště, kde je dataset uložen (jakékoliv jiné úložiště, např. CERN, Zenodo, aj.)

U datasetu lze nastavit různé přístupy – otevřený režim, s embargem nebo na vyžádání. Maximální velikost jednoho přiloženého datasetu jsou 2GB, k jednomu záznamu mohou být přiloženy soubory o velikost až 20GB. Větší objem dat je možné vložit po domluvě s administrátorem. Doporučujeme vybírat standardní formáty, které zaručují čitelnost v delším časovém horizontu. Obr. 2 znázorňuje, jak repozitář umožňuje propojení mezi bibliografickým záznamem a datovým záznamem, na jehož podkladech výstup vznikl. V online katalogu jsou vyhledatelné oba záznamy a vzájem-

ně na sebe odkazují.

DATA MANAGEMENT PLAN

Knihovna AV ČR vědcům nabízí kromě prostoru k ukládání dat i další uživatelskou podporu jako např. informace o Data Management Plan (dále jen „DMP“), který je nedílnou součástí čerpání evropské podpory z projektů Horizon 2020. DMP popisuje životní cyklus dat. Je to součást výzkumu, která obsahuje informace, které zajistí, že data budou postavena na FAIR principech. Jedná se o živý dokument. První, stručná verze by měla vzniknout během prvních 6 měsíců. Dále by měl být pravidelně aktualizován, jakmile se objeví závažné změny v průběhu projektu (např. nová data, změna v konsorciálních podmínkách atd.) a závěrečná zpráva by měla vzniknout po ukončení projektu [4]. V Knihovně AV ČR jsme pro vědce vytvořili souhrnný přehled k problematice DMP, která je k nahlédnutí na stránkách podpory ASEP [5]. A zároveň jsme vytvořili grafický přehled otázek, které by si měli vědci zodpovědět předtím, než začnou DMP vytvářet (viz. Obr. 3). K samotnému vytvoření DMP může sloužit poměrně zdařilý online nástroj DMP online [6]. Po zaregistrování intuitivně naviguje vědce krok za krokem pomocí otázek a vyplňování příslušných polí. Pod záložkou Funder requirements jsou k dispozici šablony různých organizací, které DMP vyžadují (včetně Evropské komise).

FAIR PRINCIPY

FAIR principy jsou nedílnou součástí open science prostředí a jsou často diskutovány v rámci Evropské komise a EOSC. V roce 2016 vyšel v časopise Nature článek s názvem „The FAIR Guiding Principles for scientific data management and stewardship“ [7]. Lze jej považovat za základ FAIR principů, jejichž dodržování umožňuje vzniknout důvěryhodnému prostředí pro ukládání a znovupoužití dat. Ve zkratce jde o to, aby data byla:

- F – FINDABLE (vyhledatelná, např. pomocí metadata)
- A – ACCESSIBLE (dostupná, např. prostřednictvím persistentních identifikátorů, DMP, aj.)
- I – INTEROPERABLE (interoperabilní, např. skrz normy/standards nebo společná metadata)
- R – REUSABLE (znovupoužitelná, např. prostřednictvím právních předpisů/smluv, např. Creative Commons)



Obr. 3: Okruh otázek k DMP

Expertní skupina Evropské komise sepsala a vydala příručku „Turning FAIR into reality“ [8]. Tento dokument může být užitečnou pomůckou při žádosti o evropské granty.

CERTIFIKACE DATOVÝCH ÚLOŽIŠŤ

V rámci EOSC-u se jeden z projektů týká konkrétně certifikace datových repozitářů. Tato problematika je také v prostředí open science velmi často zmiňována. Jedná se o hodnocení důvěryhodnosti datového úložiště, což může být jedno z důležitých kritérií pro vědce. Existuje více druhů certifikací, které jsou různě odstupňované dle požadavků na ochranu. Jedním z nejznámějších je Core Trust Seal [9], který se spojil s již zaniklým Data Seal of Approval. Získání certifikátu může být časově náročné, recenzenti detailně zkoumají možnosti repozitáře a splnění daných kritérií. Tento proces zároveň slouží jako nástroj pro vlastní ověření správné funkčnosti datového úložiště. Jedná se o placenou službu. V roce 2019 vlastní tento certifikát v České republice repozitář LINDAT/CLARIN, Český sociálně datový archiv a Mapová sbírka Přírodovědecké fakulty UK. Knižovna AV ČR by ráda v budoucnu pro svůj datový repozitář ASEP také o tento certifikát zažádala.

REGISTR DATOVÝCH REPOZITÁŘŮ

Velmi užitečným nástrojem v oblasti open science je Registr repozitářů vědeckých dat „re3data.org“ [10]. Slouží primárně vědcům, kteří nemají k dispozici institucionální úložiště a hledají vhodné místo kam uložit svá data. Dále může velmi dobře sloužit knihovníkům, poskytovatelům podpory, vydavatelům a jiným skupinám. Jedná se o registr napříč všemi vědeckými disciplínami, je v něm možné filtrovat podle předmětu,

země, certifikace, přístupu k datům a mnoha dalších kritérií. Podmínky registrace jsou následující:

- musí se jednat o instituci (právníckou osobu) jako např. knihovnu, univerzitu
- musí poskytnout podmínky přístupu k úložišti a podmínky používání
- musí být zaměřen na výzkumná data

Stejně jako v předchozím případě, prochází repozitář určitou selekcí a musí splňovat daná kritéria. Knižovna AV ČR by ráda o zařazení do tohoto registru požádala za nějaký čas, až bude datový repozitář více prakticky využíván.

Odkazy

- [1] <https://www.fosteropenscience.eu/content/what-open-science-introduction>
- [2] https://www.avcr.cz/opencms/export/sites/avcr.cz/.content/galerie-souboru/Politika_otevreného_přístupu_v_AVCR.pdf
- [3] Chmelařová, Zdeňka, Doleželová, Jana. Data deposit into the ASEP repository. In: Vyčítalová, H., ed. Conference on grey literature and repositories. Proceedings. Praha: Národní technická knihovna, 2017, s. 67-76. ISSN 2336-5021.
http://repozitar.techlib.cz/record/1218/files/idr-1218_1.pdf
- [4] https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm
- [5] https://www.lib.cas.cz/podpora/data/asep/pro-autory/DMP_2018.pdf
- [6] <https://dmponline.dcc.ac.uk/>
- [7] <https://www.nature.com/articles/sdata201618>
- [8] <https://op.europa.eu/en/publication-detail/-/publication/7769a148-f1f6-11e8-9982-01aa75ed71a1/language-en/format-PDF/source-80611283>
- [9] <https://www.coretrustseal.org/>
- [10] <https://www.re3data.org/>

Mgr. Karolina Podloucká
podloucka@lib.cas.cz
(Knižnica AV ČR, v. v. i.)