

Zpráva ze služební cesty

č.j. NTK/1679/2026/9

The Big Migration – On the Road to the Next Generation of Scientific Repositories

Varšava, Polsko

12.-14.5.2026

Účastníci:

Barbora Vacková, 82 – Centrum repozitářů a správy metadat

Datum zpracování: květen 2026



Účel cesty

Konference The Big Migration – On the Road to the Next Generation of Scientific Repositories byla pořádána společností PCG Akademie, certifikovaným dodavatelem repozitářového softwaru DSpace, a University of Warsaw, největší polskou univerzitou. Tematickým záběrem byla migrace a vývoj v rámci DSpace. Konference byla zejména zaměřená na region střední a východní Evropy, ovšem účast byla celoevropská.

Hlavním účelem cesty bylo zmapovat středoevropské zkušenosti s migrací a updaty softwaru Dspace, získat přehled o institucích, které Dspace provozují na samostatně bez dodavatele a které naopak využívají služeb poskytovatelů a také řešení současných palčivých otázek provozu repozitářů, v první řadě především nápor botů a ostatních scraperů.

Časový průběh služební cesty

12.5.2026 odjezd z Prahy do Varšavy

12. a 13.5.2026 účast na konferenci The Big Migration

13. a 14.5.2026 odjezd z Varšavy do Prahy

Podrobnosti a shrnutí programu konference

Program konference je dostupný na webu <https://pcgacademia.pl/en/events/conferences/big-migration/>.

Úterý 12.5.2026 – registrace, úvod a dopolední část

Po registracích proběhlo úvodní slovo prorektora pro výzkum Varšavské univerzity a předsedy rady společnosti PCG Academia. Poté následoval high-level knowledge transfer od Anny Książczak-Gronowské, supervizorky REIN repozitáře Varšavské univerzity. V rámci migrace integrovali institucionální repozitář s repozitářem výzkumných dat a kontrolovanými slovníky (knowledge management platform) Jaskier. Tento overview zkušeností s migrací se týkal především spolupráce s PCG Academia, která migraci prováděla. Po této přednášce následoval Pablo de Castro z Glasgow, který mluvil o rozdílech, a především zvyšující se míře podobnosti repozitářů a CRIS systémů. Toto téma je důležité především z hlediska plánovaného spojení dvou distribucí softwaru DSpace, DSpace zastřešovaný společností Lyris a DSpace-CRIS od italské 4science. Toto spojení je naplánováno přes dvě budoucí majoritní verze, letošní verzí 10 a na příští rok naplánovanou verzí DSpace 11.

Po pauze s ukázkami jednoduchých workflow v DSpace UI rozhraní následoval blok krátkých prezentací o různých zkušenostech s migracemi DSpace, příkladem ze softwaru PURE od Elsevier. Přednášky byly spíše managementově orientované a obsahovaly jen málo technických detailů. Všechny přednášky se nesly v duchu agregace dat z různých zdrojů a jejich transformací na nový jednotný formát.

Úterý 12.5.2026 – odpolední část

V odpolední části probíhaly prezentace různých podpůrných služeb pro repozitáře a Open Science pro dodržení best practices. První prezentace byla o ORCIDu a jeho integraci se softwarem DSpace, kde přednášející kladla důraz na užívání celého ekosystému, který ORCID nabízí – zejména využívání API rozhraní pro zadávání ORCIDů a ne manuálního zadávání.

Poté byla prezentace certifikace CoreTrustSeal, která posluchače provedla procesem získání certifikace. V rámci diskuze po prezentaci bylo budoucím žadatelům doporučeno prohlížet staré schválené žádosti, které CoreTrustSeal archivuje na svých stránkách, a inspirovat se technickým popisem a policies repozitářů, které certifikaci již získaly.

Nakonec se představil Petr Knoth s CORE, který shromažďuje databázi odborných publikací a textů z celého světa – jak Open Access, tak uzavřené soubory. CORE dále poskytuje mnoho různých služeb pro repozitáře, od agregovaného metasearch engine přes health checks dat jednotlivých repozitářů po dashboard a další integrované nástroje. V rámci prezentace zaznělo i doporučení na přístup k AI crawlerů a to limitovat počet indexačních nástrojů, které využívají repozitářovou infrastrukturu, ale neomezit úplně jejich přístup, neboť jsou repozitáře v dnešní době důležitý zdroj informací. Zároveň přislíbil podrobnější prezentaci jejich CORE Dashboardu na další den.

Následovala závěrečná řeč od zástupce PCG Academie.

Středa 13.5.2026 – první část

V první části proběhlo představení regionálních skupin DSpace Německa, Švýcarska, Polska, Itálie a Česka a Slovenska. CZ&SK skupinu reprezentoval Jozef Mišutka z firmy Dataquest, která českou skupinu na začátku tohoto roku založila a uspořádala první oficiální regionální setkání této skupiny. Jednotliví zástupci především vyzdvihovali výhody zapojení se do těchto skupin a příklady problémů, které takto spojení dokázali přednést před DSpace Steering Comitee.

Poté zástupci CORE Petr Knoth a Catherine Kuliavets představili novou službu CORE Dashboard. Je dostupná pro členy CORE a umožňuje široký záběr funkcionalit pro správu repozitářů a interface pro statistiky. Dále nabízí nástroje pro obohacení záznamů (dohledáním jiných dostupných informací), hodnocení kvality metadat dle Research Outputs Metadata Schema (RIOXX), kontrolu duplicit záznamů v repozitáři a open access a FAIR compliance kontrolu výsledků v repozitáři. Z dashboardu lze ovládat indexování a kontroly indexování repozitářů do CORE rozcestníku. V případě problému dashboard vygeneruje přehled chybových záznamů i doporučení, jak chyby opravit. Pro široký dopad repozitářů NTK je možné zvážit využití CORE dashboardu jako dalšího nástroje pro správu repozitářů, popřípadě CORE dashboardem doplnit dashboard, například od OpenAIRE.

Následovala prezentace italského zástupce, který mluvil o vlastním vytvořeném dashboardu na základě dat z DSpace repozitáře.

Středa 13.5.2026 – druhá část a závěr

Po pauze následovaly poslední tři prezentace a závěrečná řeč. První z prezentací se týkala integrace AI a repozitářů. Zvláštní pozornost byla věnována zpracovatelnosti obsahu AI. Tým z PCG Academie navrhoval vytvoření třetí vrstvy s agenty nad repozitáři (první – datová vrstva,

druhá – sémantická vrstva), která by prezentovala machine-reading a NLP optimalizované rozhraní. PCG Academia přistupuje k repozitáři s AI jako k repozitáři, který si je „vědom svého vlastního obsahu“. Součástí integrace AI je například agent v pozadí, který kontroluje a doplňuje údaje v rámci kurátorského workflow. Změnu ale vždy musí schválit lidský kurátor. Další agent může analyzovat obsah datasetů a provádět obsahovou analýzu obsahu repozitáři a poskytovat nadstavbu nad defaultním vyhledáváním klíčovými slovy v repozitáři. přednáška se také mírně dotkla problematiky vnějších AI botů a content scraperů, o kterých se poté rozvinula náruživá diskuze, která se přelila i do oběda. Dle očekávání, všechny repozitáře musí obrovský nárůst požadavků a síťového provozu řešit a jedná se tak o jedno z klíčových témat repozitářů pro nejbližší období. Byla tak škoda, že k této problematice nebylo řečeno více. Naštěstí se ale v diskusi vyjádřili klíčoví odborníci a lidé, kteří mají s manageováním provozu velké zkušenosti. Důraz byl kladen především na zachování přístupnosti obsahu bez toho, aby repozitáře musely nést celé náklady za přidanou hodnotu, kterou boti z otevřených dat a publikací získávají. Klíčové také bylo zachování dostupnosti a funkčnosti pro lidské uživatele, například vytvořením endpointů speciálně pro AI boty.

Druhá přednáška byla od týmu z Litevské Akademie kultury. Kristine Bramberga a Austra Ciša sdílely své zkušenosti s návrhem a tvorbou repozitáře pro kulturní produkci, která je zároveň závěrečnou prací studentů Akademie. Repozitář byl vyvíjen ve spolupráci s PCG Academia. Problematika reprezentace komplexních vztahů při vzniku artefaktů (uměleckého díla, související analýzy, scénáře apod.) v repozitářích uložených byla na problematice kulturní instituce především zřejmá.

Poslední prezentace se týkala polského Národního filmového archivu Audiovizuálního institutu (FINA). Představovala především problematiku velmi rozsáhlých historických kolekcí z managementového pohledu.

Následovala závěrečná řeč viceprezidenta rady PCG Academie. V ní podtrhoval především existující rozvinuté a bohaté repozitáře střední Evropy a jejich budoucí potenciál řídit vývoj celé platformy DSpace.

Závěry a shrnutí

Konference byla přínosná z několika hledisek: umožnila účastníkům zmapovat oblast provozovatelů a poskytovatelů repozitářových služeb ve střední Evropě a získat mírný přehled o problémech, na které při jejich údržbě a provozu narážejí. Zdaleka nejzásadnější byla diskuze o výměna znalostí a názorů týkající se současné problematiky botů. Jak již bylo zmíněno, toto bude v dohledné době velkým tématem pro všechny repozitáře, od malých, jako je náš Institucionální repozitář, Národní repozitář středního rozsahu, až po ty největší. Je dobrou zprávou, že si to uvědomují nejen provozovatelé repozitářů, ale i poskytovatelé repozitářových služeb. Můžeme tedy doufat, že budou probíhat práce na společných, jednodušších, stabilnějších a zároveň responsivnějších řešeních.

Další důležitou přednáškou byla přednáška o repozitáři Litevské Akademie kultury. Reprezentace vztahů a komplexní skutečnosti, kterou se repozitáře snaží uchovat a zaznamenat, je v samém základu problematiky repozitářů. Ze zkušeností, které při reprezentaci takto komplexních celků získají kulturní instituce, budou moci čerpat i ostatní neumělecky zaměřené instituce. Posledním důležitým zjištěním bylo, že i když většina institucí provozujících repozitáře využívá služeb externích poskytovatelů, zdaleka nejsme jediní, kdo repozitář udržuje především svépomocí, a že i přes to lze poskytovat kompetitivní a poměrně podobně spolehlivé služby v oblasti repozitářů.

Zdroje

Program a veškeré informace o konferenci jsou dostupné na webu <https://pcgacademia.pl/en/events/conferences/big-migration/>.

Závěrečná zpráva z konference od pořadatele a nahrávky prezentací a fotodokumentace jsou dostupné z <https://pcgacademia.pl/en/summary-of-the-big-migration-conference-international-experts-discuss-the-future-of-dspace-migrations/>.

Fotodokumentace

