

Zpráva ze služební cesty

20th International Digital Curation Conference

Záhřeb, Chorvatsko

16.-18.2.2026

Účastníci:

Barbora Vacková, 82 - Centrum repozitářů a správy metadat
Číslo jednací: NTK/2554/2026

Datum zpracování: březen 2026



Účel cesty

Digital Curation Centre (DCC) je mezinárodně uznávané odborné centrum v oblasti digitálního kurátorství, správy výzkumných dat a otevřené vědy se sídlem ve Skotsku. Bylo založeno v roce 2004 a jeho posláním je pomáhat organizacím po celém světě co nejlépe využívat data ve výzkumu, vývoji a provozních činnostech ve všech kontextech – akademickém, průmyslovém, vládním a dalších. Každoročně pořádaná konference na podporu kurátorství se letos konala v chorvatském Záhřebu ve dnech 16. – 18.2.2026. Účast zaměstnanců NTK na této konferenci je klíčová z hlediska čerpání znalostí v oboru digitálního kurátorství a správy vědeckých dat. Velkým přínosem této konference je fakt, že se nejedná pouze o teoretické přednášky, ale nabízí mnoho případových studií (tzv. use case) z oboru, kde se můžeme dozvědět praktické řešení problémů a výzev od kolegů, kteří již některou problematikou prošli nebo ji řešili. Jedná se o skvělou příležitost k výměně názorů a navázání nových kontaktů.

Časový průběh služební cesty

16.2.2026	odlet z Prahy do Záhřebu
17.2. - 18.2.2026	účast na konferenci IDCC 2026
18.2.2026	odlet ze Záhřebu do Prahy

Podrobnosti a shrnutí programu konference

Program konference je dostupný zde: <https://dcc.ac.uk/events/idcc26/programme>

Prezentace z konference jsou dostupné zde: [IDCC26 Zenodo community](#)

Detailní informace k jednotlivým příspěvkům, které jsme navštívily:

Evolving with the Community: Updates to the CoreTrustSeal Requirements for 2026 – 2028

Jared Lyle a Tuomas Alateraa

Prezentace dostupná zde: <https://doi.org/10.5281/zenodo.18682657>

Tato prezentace shrnuje nadcházející změny v certifikaci CoreTrustSeal. Jedná se o jednu z nejuznávanějších a nejrozšířenějších certifikací pro digitální repozitáře. Základní rámec certifikace se nemění. Stále se jedná o komunitně řízený proces se 16 hlavními požadavky, které se zaměřují na

důvěryhodnost, udržitelnost a dlouhodobý přístup k datům. Certifikace zůstává platná po dobu 3 let a proces zahrnuje sebehodnocení následované peer-review.

Mezi klíčové změny pro období 2026 – 2028 patří následující body:

- **Důraz na aktivní archivaci (Active Preservation):** Samotné bezpečné uložení dat (storage) již není považováno za dostatečné. Nové požadavky jasněji definují úroveň péče, přičemž nejvyšší prioritou je aktivní správa dat po celou dobu jejich životního cyklu.
- **Odpovědnost při outsourcingu:** Repoziťáře mohou využívat externí služby, ale odpovědnost za data nelze přenést. Musí být jasně doloženo, kdo co dělá, na základě jakých smluv a jak probíhá dohled nad těmito službami.
- **Praxe namísto pouhých politik:** Certifikátoři budou klást menší důraz na „ambiciózní“ prohlášení v dokumentech a budou vyžadovat konkrétní důkazy o probíhajících postupech, monitorování a revizích.
- **Archivace neznamená „uchovat vše navždy“:** Nově je výslovně uznáno, že repoziťáře mohou provádět přehodnocení (re-appraisal) a případně data odstraňovat, pokud jsou tyto kroky zdokumentovány a mají jasná pravidla pro dopady na citace a trvalé identifikátory (PIDs).
- **Certifikace jako proces zlepšování:** Při obnově certifikace se nečeká dokonalost, ale reflexe pokroku, identifikace mezer a plánování budoucích vylepšení.

Cílem těchto úprav je zajistit větší konzistenci pro žadatele i hodnotitele a lépe odrážet realitu moderní správy digitálních dat.

Trust, Types & Transparency: TIC-TAC-TOE

Hervé Lhours, Mari Kleemola, Oliver Parkes a Fabio Liberante

Prezentace dostupná zde: <https://doi.org/10.5281/zenodo.18942172>

Tato prezentace se zaměřila na rozvoj Evropské sítě důvěryhodných repoziťářů v rámci projektu **FIDELIS**. Hlavním cílem projektu FIDELIS je vytvořit a provozovat síť důvěryhodných repoziťářů, která podpoří rozvoj ekosystému EOSC. Klíčovým nástrojem je **TTRAM** (*Transparent, Trustworthy Repository Attributes Matrix*). Jedná se o matici 30 aktivit a funkcí (v oblastech správy objektů, infrastruktury, technologií a bezpečnosti), která čerpá ze stávajících standardů a pomáhá repoziťářům sdílet strukturované informace o jejich praxi.

Autoři navrhuji nový rámec pro chápání důvěryhodnosti repoziťářů, který přesahuje jednoduché binární hodnocení, tzv. **Koncept TIC-TAC-TOE**:

- **TIC (Trust In Context):** Důvěra není statický výsledek, ale závisí na kontextu objektů, aktérů a očekávaných služeb. Různé authority (např. CoreTrustSeal vs. Evropská komise) mají na repoziťáře odlišné požadavky.
- **TAC (Types And Categories):** Repoziťáře by neměly být kategorizovány pouze podle typu instituce, ale podle toho, jaké typy objektů přijímají, jaké uživatele obsluhují a jakou úroveň kurace (LoRCAP) poskytují.

- **TOE (Transparent Objects and Entities):** Důvěra musí být podložena transparentními informacemi o organizacích, katalogích a digitálních objektech. Pro tyto subjekty se používá širší označení DOVE (*Digital Object Venue Entity*).

Prezentace uzavírá tím, že další práce se zaměří na hlubší pochopení entit DOVE, modelů autorit a charakteristik uživatelů a vkladatelů, aby bylo možné lépe doporučit, které informace mají repozitáře zveřejňovat pro prokázání své důvěryhodnosti. Vzhledem k tomu, že NTK se v Centru repozitářů a správy metadat dlouhodobě zabývá problematikou důvěryhodných repozitářů, budeme rozvoj tohoto konceptu nadále sledovat.

Developing Trusted Preservation Services for Sensitive Research Data: A Case Study

Mari Elisa Kuusniemi

Prezentace dostupná zde: <https://doi.org/10.5281/zenodo.18942175>

Tato případová studie popisuje proces budování bezpečných služeb pro uchovávání citlivých vědeckých dat prostřednictvím dvou hlavních platform: **Databank** (kurátorské úložiště na 5–15 let) a **Data archive** (dlouhodobá archivace na 25 a více let).

Jako první krok si musela univerzita ujasnit svůj přístup k dlouhodobému ukládání citlivých dat a identifikovat mezery mezi požadavky na informační bezpečnost a ochranu osobních údajů. Důležitým faktorem byla příprava na novou legislativu o archivaci dat, která vstoupí v platnost v létě 2026.

Dále bylo potřeba vybudovat expertní síť. Byla posílena spolupráce s právními poradci (GDPR), etickými komisemi a IT specialisty, což umožňuje konzultovat složité případy v reálném čase. A tým samotný si musel zlepšit dovednosti v posuzování citlivosti dat a rizik, mimo jiné prostřednictvím interních skupin a tréninkového programu FAIR-IMPACT.

Jako třetí krok je označováno vytvoření pracovních postupů a procesů. V rámci vývoje byly zavedeny nové administrativní a kontrolní mechanismy: **Data Access Protocol (DAP)** - protokol pro řízení přístupu k datům s omezeným režimem. A **DPIA (Data Protection Impact Assessment)** - posouzení vlivu na ochranu osobních údajů v souladu s GDPR pro obě platformy. Byla nutná aktualizace podmínek užití pro podporu ukládání a opětovného využití citlivých dat.

Dalším krokem bylo zhodnocení technického zabezpečení kdy bylo zjištěno, že zatímco Data archive již požadavky splňoval, pro Databank muselo být adekvátní zabezpečení nově navrženo a implementováno. Prezentace zmiňuje i praktické překážky, jako je náročnost šifrování velmi objemných datových sad, což komplikuje jejich přenos a následné využití.

Jako závěr a doporučení této případové studie vyplývá, že vývoj služeb pro citlivá data je komplexní proces, nikoliv jednorázový úkol. Je nutné vyčlenit značné množství času na přípravu administrativní

dokumentace. Klíčovými faktory úspěchu jsou spolupráce napříč obory a neustálé zlepšování služeb pro dosažení udržitelných a legislativně čistých řešení.

FAIRVault, An Inter-University Federated Dataverse Pilot In Flanders

Nicky Daniels, Kevin Leonard, Laura Standaert, Myriam Mertens, Jone Paesmans, Sybren Slegers, Linde Tuybens, Thomas Van de Velde a Thomas van Erven

Prezentace dostupná zde: <https://doi.org/10.5281/zenodo.18942177>

Tento příspěvek přinesl zajímavý příklad meziuniverzitní kooperace zaměřený na vytvoření federovaného úložiště pro výzkumná data. Představuje společný projekt čtyř vlámských univerzit (UGent, VUB, UAntwerpen a UHasselt), který probíhá s podporou Vlámské rady pro otevřenou vědu. Momentálně se dokončuje druhá fáze a začíná pilotní testování.

Cílem projektu je vyvinout řešení pro uchovávání a sdílení dat, které odpovídá právním a etickým rámcům („otevřené, jak je to jen možné, uzavřené, jak je to nutné“) v situaci, kdy stávající repozitáře nevyhovují všem potřebám institucí. Projekt je založen na softwaru **Dataverse** a využívá model „Bring your own storage“, kde si každá univerzita připojuje vlastní lokální úložiště. FAIRVault je navržen tak, aby umožňoval bezpečné ukládání citlivých údajů. K tomu využívá systém datových značek (Data Tags), které barevně odlišují úroveň přístupu od zcela otevřených (modrá) až po přísně omezená data vyžadující smlouvu o přenosu (červená). Do repozitáře mohou vkládat data pouze výzkumníci z partnerských institucí. Všechny vklady procházejí manuální kurací odborníky z domovské univerzity daného výzkumníka, aby byla zajištěna kvalita a soulad s pravidly. Repozitář podporuje FAIR principy přidělováním DOI, rozsáhlými metadaty a automatickou konverzí formátů souborů. Důvěryhodnost je budována skrze expertní posuzování dat a jasně definované politiky uchovávání.

Standardization Vs. Preservation? Supporting Interoperability by Enhancing Thematic Metadata at Social Science Archives

Roza Eszter Vajda, Judit Gárdos a Tímea Venczel

Prezentace dostupná zde: <https://doi.org/10.5281/zenodo.18879682>

Autoři tohoto příspěvku jsou z Centre for Social Sciences v Maďarsku a řeší napětí mezi přirozenou rozmanitostí metadat v archivech a „ideálem“ FAIR principů, konkrétně potřebou zvýšit interoperabilitu (schopnost systémů spolupracovat) skrze kvalitnější tematická metadata. Představili projekt **ONTOLISST**, který kombinuje výzkum ontologií a vývoj nástrojů pro analýzu a potenciální změnu dokumentační praxe v archivech. Výzkum zahrnoval 10 archivů sociálních věd (např. GESIS, ICPSR, SNDS), kde probíhaly průzkumy metadat a rozhovory. Výzkumem bylo zjištěno, že interoperabilita zatím není pro archivy

nejvyšší prioritou, tematická metadata jsou často povrchní a používání řízených slovníků je spíše náhodné. Přechod na nové protokoly je sice vnímán jako nákladný a pomalý, ale postupně k němu dochází. Projekt vyvíjí nástroj **LiSST** založený na poloautomatickém procesu, který využívá AI (model BERTopic) pro shlukování témat z textů otázek a proměnných, doplněný o expertní sociologickou analýzu. Bylo identifikováno 10 hlavních tematických okruhů (např. demografie, zdraví, práce, vzdělávání, příjmy), které slouží k lepší organizaci dat. Prezentace také popisuje dilemata při klasifikaci, jako je řešení komplexních kategorií nebo nepřímých měření (např. otázka na velikost bot dítěte jako indikátor chudoby). Závěrem této prezentace je zjištění, že pro úspěšnou standardizaci je nezbytné najít rovnováhu mezi potřebami producentů dat, uživatelů a specifickými tradicemi jednotlivých institucí a národních archivů.

How & When To Use A Data Steward? The Use-case Of The Czech RDM Ecosystem

Georgia Koutentaki, Matyáš Hiřman, Marek Suchánek, Jana Martínková a Robert Pergl

Prezentace dostupná zde: <https://doi.org/10.5281/zenodo.18983691>

Micro-Creds: Innovative Training for Sustainable Communities

Jenny O'Neill a Aoife Coffey

Prezentace dostupná zde: <https://doi.org/10.5281/zenodo.18983697>

Establishment of a Certification Program for Research Data Management Support Personnel at Kyushu University, Japan

Emi Ishita, Yoichi Tomiura a Toshiyuki Shimizu

Prezentace dostupná zde: <https://doi.org/10.5281/zenodo.18983703>

Tyto tři příspěvky se věnovali rozvoji a vzdělávání datových kurátorů (data stewards) v ekosystémech správy výzkumných dat (RDM) v České republice, Irsku a Japonsku. Společným tématem všech příspěvků byla profesionalizace správy dat a snaha o to, aby se RDM stalo standardní součástí vědecké práce, nikoliv jen administrativní zátěží.

V České republice je kladen důraz na budování **Národní repozitářové platformy (NRP)** a integraci do struktur **EOSC (European Open Science Cloud)**. Umělá inteligence přináší výhody v podobě automatizace manuálních úkolů, rychlejšího vyhledávání publikací a tvorby skriptů pro datové toky. Zároveň však představuje rizika, jako je snížená transparentnost, bezpečnostní hrozby při nahrávání citlivých dat a poskytování falešných citací. Pro čištění a správu dat jsou doporučovány nástroje jako **OpenRefine**, Python (Pandas), R, SQL nebo specializovaný **Data Stewardship Wizard (DSW)**.

Případová studie z Irsku nazvaná **Sonraí** (<https://datastewards.ie/>) se zaměřuje na vzdělávání data stewardů. K udržitelnému rozvoji komunity využívají tzv. mikro-certifikáty (micro-credentials), což jsou krátké, akreditované a flexibilní kurzy určené pro zvyšování kvalifikace. Obdrželi Skills4EOSC Data Stewards curriculum. Jedná se převážně o online kurz, ale je možné absolvovat i 2-in-person sessions. Zatím se jedná o pilotní kurz, po jehož skončení bude provedena evaluace a následně by chtěli rozvíjet modul pro Health Science. Zájem je veliký, původně byl kurz koncipován pro 10 studentů, ale má jich 24. Uchazeči jsou z řad

knihovníků, kteří již pracují jako data stewardi nebo výzkumníci, kteří se chtějí naučit jak lépe organizovat svá data.

V Japonsku reaguje Kyushu University na vládní plány, které od roku 2025 vyžadují plány správy dat (DMP) u všech nových projektů a okamžitý otevřený přístup k výsledkům výzkumu. Byl zaveden speciální certifikační program pro pracovníky podpory RDM, zejména pro zaměstnance knihoven a univerzitní administrátory. Výuka zahrnuje teoretické základy informačního managementu, praktickou podporu během celého životního cyklu dat (čištění dat, metadata, ukládání do repozitářů) a nácvik zjišťování potřeb vědců pomocí rozhovorů. Výuka probíhá pouze v japonštině.

Helpdesk on Research Data Management: AI-based Chatbot for FAIR Research Data in Portugal

António Lopes, Clara Boavida, Carina Cunha, João Pedro Oliveira, Pedro Príncipe a André Vieira

Prezentace dostupná zde: <https://doi.org/10.5281/zenodo.18988619>

Prezentace představuje projekt **portugalské národní sítě Re.Data** zaměřený na podporu správy výzkumných dat (RDM) a implementaci **principů FAIR**. Hlavním cílem je vytvořit koordinovaný národní přístup k otevřené vědě a propojit komunitu datových kurátorů a vědců s evropským cloudem pro otevřenou vědu (EOSC).

V rámci pracovní skupiny WP5 byl vyvinut **AI Chatbot pro Helpdesk založený na umělé inteligenci**, který slouží jako vstupní bod pro výzkumníky, knihovníky a datové kurátory. Systém využívá metodu **Retrieval Augmented Generation (RAG)**. Ta kombinuje schopnosti velkých jazykových modelů (LLM) s ověřenou bází znalostí, která obsahuje více než 280 párů otázek a odpovědí v portugalské a angličtině. Chatbot čerpá z dokumentace Re.Data a pokrývá témata jako FAIR principy, data steward, právní otázky, ochranu údajů a licencování. Služba je spravována na národní úrovni a financována agenturou FCT/FCCN. Model je navržen jako škálovatelný a flexibilní, což umožňuje ostatním institucím jej replikovat a obohatit o vlastní lokální pravidla.

Chatbot je veřejně dostupný na webových stránkách projektu **redata.pt**.

Introducing FJORD: Orchestrating Research Data Management through Enhanced FAIRness

Federico Grasso Toro

Prezentace dostupná zde: <https://doi.org/10.5281/zenodo.18892563>

V rámci této přednášky byl představen **FJORD (FAIRly Jointed Open Research Data)**, inovativní rámec (framework) navržený pro řešení výzev spojených se správou výzkumných dat (RDM), který má za cíl zvýšit

úroveň naplňování principů FAIR. FIORD není navržen jako náhrada stávajících systémů, ale funguje jako vrstva metadat, která sedí mezi nimi. Působí jako „překladač“, který propojuje informace a mapuje stávající data do výzkumných informací. Jedná se o jakýsi pre-DMP před podáním žádosti, díky němuž se např. univerzita v dostatečném předstihu dozví, jaké finanční prostředky na projekt vyčlenit. Vědci získávají sjednocený dashboard pro správu a automatizovaný reporting, což snižuje jejich administrativní zátěž. Datoví kurátoři mohou sledovat a hodnotit úroveň FAIR (tzv. FAIRness) výstupů v reálném čase pomocí poloautomatizovaných procesů. Vedení institucí a poskytovatelé financí získávají ověřený a jasný přehled o výzkumných aktivitách pro strategické plánování a kontrolu souladu s pravidly. Jedná se o dotazník s 16ti „multiple“ otázkami, který lze následně vyexportovat do formátu JSON a je možné jej opatřit komentáři.

Enhancing the Benefits of Machine-Actionable DMPs with Generative AI

Becky Grady

Prezentace dostupná zde: <https://doi.org/10.5281/zenodo.18964620>

GenAI For Research Support: A Case To Dmps Pre-Filling

Davide Nardi, Liz Guzman Ramirez a Erwin Hoogerwoord

Prezentace dostupná zde: <https://doi.org/10.5281/zenodo.18964622>

Exploring the Use of LLMs to Support Research Data Documentation in the Social

Auriane Marmier

Combining Machine Learning tools and Terminology Services to support FAIR Data Management in ESS

Claudia Martens a Alexander Wolodkin

Umělá inteligence byla tématem celé konference. Pracovníci v informačních službách a knihovnách sdíleli své představy o uplatnění AI v oboru repozitářů. Blok A spolu s blokem M daly prostor pro příspěvky, které vyzvedly výzvy i příležitosti při práci s AI, například jako užití AI pro sémantické vyhledávání v repozitáři. Velké téma bylo využití AI pro první kontrolu záznamů v repozitářích AI kurátorem, zejména metadata doplněná do záznamu. Ten poskytne prvotní zpětnou vazbu, navrhne oblasti, které potřebují dopracovat nebo vůbec nesplňují markery pro dobrý záznam. Umělá inteligence zakládá své výstupy na kolekci dat od lidských kurátorů a po kontrole AI je záznam stále podstoupený lidským kurátorům.

Další zajímavý příspěvek narazil na úskalí AI, kdy AI naopak přetěžuje infrastrukturu repozitářů kvůli opakovaným přístupům a vytěžování informací. Knihovna při Univerzitě v Oxfordu plánuje sepsat politiku specificky pro případy, kdy AI boti nebo jiné strany přistupují do repozitářů a sbírají data pro trénování jazykových modelů.

AI se vrací i v oblasti podpory výzkumné činnosti, kde může dle některých příspěvků vypomáhat při tvorbě Data Management Plans nebo doplňující dokumentaci datových sad. Opět na principu kontroly benchmark metrik nebo předvyplnění údajů na základě jiných existujících dokumentů.

On Curating HTR Training Datasets for Romanian Language with use of Transcribathon Tool

Sergiu Gordea, George Cristian Cotea, Frank Drauschke, Jasmin Salesevic, Marthe Lamote a Tom Vanallemeersch

Dealing with Unprecedented Scale and Complexity: Lessons from Archiving HS2 Digital Archaeological Data

Marco Brunello, Valeria Carrillo a Olivia Foster

Auditing the Human BioMolecular Atlas Program (HuBMAP) Human Reference Atlas (HRA): An Evaluation of Core Digital Objects

Devan Ray Donaldson, Mary Nelson a Katy Börner

První příspěvek mluvil o zapojení občanské vědy (Citizen Science) při získání vstupních dat pro trénování transkripce. Další příspěvky shrnuly způsoby, jak se vypořádat s velkým objemem dat, které potřebují dohled kurátora nebo jak postupovat při auditu repozitářů s velkým množstvím dat.

Moving Data Citation Forward: The Multilayered Approach of a Social Sciences Data Repository

Marieke Heers a Christina Bornatici

Towards Sustainable Curation: Evaluation of Cost and Accuracy of AI Tools in Scaling Annotation Tasks in Curation of Biomedical Literature

Latrice Landry, Aniete Andy, Ebelechwu Nwafor a Mary Lucas

Na závěr se prezentovaly způsoby, jak podpořit další užívání dat – ve Švýcarsku se snaží poskytnout co nejlepší zázemí pro znovupoužití vytvořených dat, například integrováním citace v repozitáři a zahrnováním datové sady do seznamu zdrojů a literatury.

Poster session

EOSC začíná pilotní program **FIDELIS** pro digitální repozitáře a pro zapojení lidí a repozitářů z méně reprezentovaných a zastoupených zemí. Nabízí výměnu informací a podporu pro méně reprezentované země, plánuje mapovat jedinečné požadavky z různých koutů Evropy a peer-to-peer mentoring program. Sledování a zapojení se do aktivit může být jedinečnou příležitostí pro NTK jak šířit praxi z oddělení repozitářů.

Závěr a naplnění cílů služební cesty

Závěrem lze konstatovat, že cíle služební cesty byly naplněny ve všech klíčových oblastech.

Účast na konferenci umožnila zmapovat aktuální trendy v oboru datového kurátorství, tentokrát se zaměřením na stále více se rozvíjející problematiku umělé inteligence (AI). Velkým přínosem pro NTK z příspěvků této konference je pracovní postup při trénování AI, nastavení škál důležitosti jednotlivých markerů nebo zasazení do workflow repozitářů. Jelikož NTK provozuje dva repozitáře pro ukládání dokumentů, je AI výpomoc oblastí, kterou se můžeme zabývat. Pokud bude NTK chtít zároveň vymezit užívání AI jiných stran, může na to pamatovat v podmínkách užívání či jiných dokumentech, kde se může inspirovat například Univerzitou v Oxfordu.

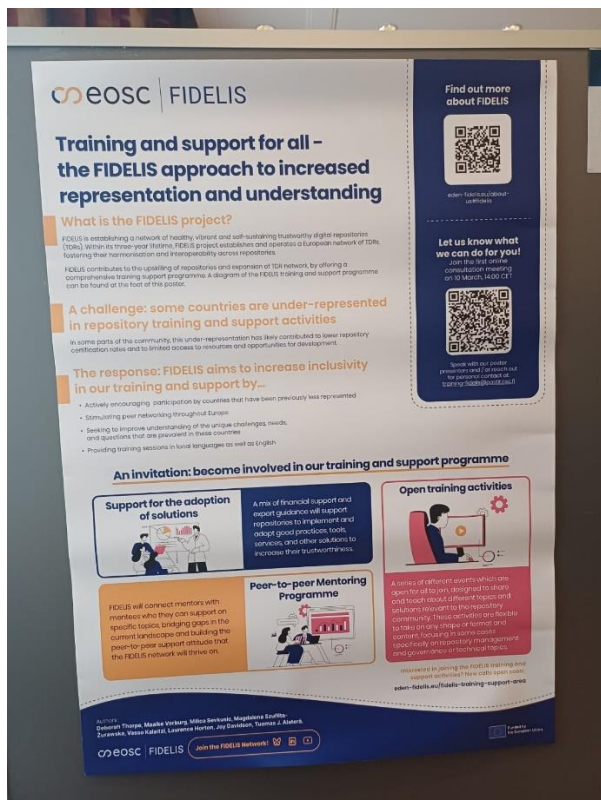
Zdroje:

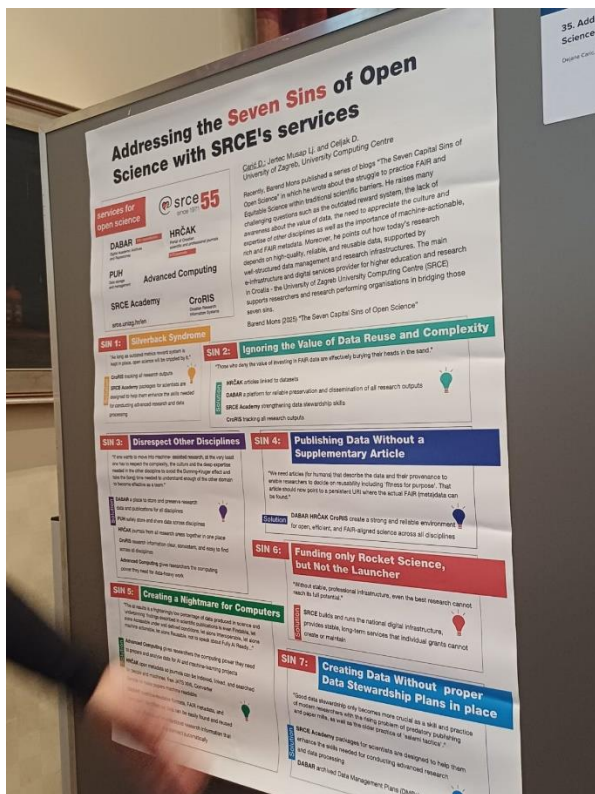
Informace o konferenci: <https://dcc.ac.uk/events/idcc26>

Program konference: <https://dcc.ac.uk/events/idcc26/programme>

Při přípravě tohoto materiálu byl použit nástroj Notebook LM, a to při práci s osobními poznámkami a strukturování dokumentu.

Fotodokumentace





AI complementing human intelligence in DMP assessment

Artificial vs. Human Intelligence: A Comparative Assessment of Feedback to Data Management Plans

Jari Pitkanen (Tampere University) ORCID: 0000-0001-6223-4169
Salla Huusanen (Tampere University) ORCID: 0000-0002-2316-283X
Maria Kaakonen (Tampere University) ORCID: 0009-0003-1304-5371
Arianna Lahti-Kauranne (CSG) ORCID: 0009-0000-2403-4107
Jukka Rantasaari (University of Turku) ORCID: 0000-0001-5927-3791
Juuso Repo (University of Turku) ORCID: 0000-0001-8756-6936

1 Introduction

This study investigated the reliability and usefulness of AI-assisted assessment of Data Management Plans (DMPs). First, we used an AI-powered DMP Analyzer to assess DMPs (n=10) against established evaluation criteria. Next, we compared AI outputs against assessments conducted by data support professionals.

The DMP Analyzer applies a modular analysis pipeline to produce structured and traceable AI outputs and exploit audit trail from each judgment to sentence-level evidence in the original DMP. This design enables experts to inspect, validate and contest AI outputs.

2 Multi-level assessment criteria

3 Traceable, staged evaluation pipeline for Data Management Plans

All steps operate over the same underlying DMP document, represented at sentence level.

4 Insights from testing

- The AI-based DMP Analyzer follows predefined criteria recursively, resulting in predictable and transparent output.
- Where performance was weak, limitations were mainly due to underspecified or context-dependent criteria rather than model errors.
- In contrast, experts were able to interpret the same criteria flexibly and take the research context into account.
- The Analyzer struggled to distinguish whether missing information was irrelevant or whether it should have been included. For example, in the legislation section, it may fail to recognize the use of laboratory animals and the specific regulations that apply to them.

5 Future expectations

- The DMP Analyzer has the potential to provide consistent and scalable support for DMP assessment.
- While currently designed for expert use, it could be adapted for researchers to improve DMP quality prior to possible expert review.
- Integration into existing DMP platforms could streamline workflows by reducing the need for multiple separate tools.
- Further evaluation in real-world research support settings is needed.
- The tool is intended to complement human expertise, improving efficiency and consistency rather than replacing expert judgement.

DMP Analyzer features

- Verifiable AI Outputs for human-in-the-loop
- Enterprise-level Data Security & Privacy
- Automatic University Statistics
- Accumulating Knowledge Base for Consistency
- Actionable Feedback for Researchers
- Customizable for Any Assessment Framework

SEE HOW IT WORKS / © 2024 by A.O.

