

Výzkumná data

Barbora Drobíková, Ph.D.

Připraveno spolu s Mgr. Adélou Jarolímkovou, Ph.D.

Proč se zabývat správou výzkumných dat

- Přínos pro vědce
 - Zlepšení kvality výzkumných dat
 - Přístup ke spolehlivým „pracovním“ datům
 - Závěry mohou být kdykoliv externě potvrzeny
 - Data a informace získané z výzkumu „přežijí“ výzkumný projekt
 - Dobře spravovaná data mohou být využita při dalším výzkumu

Pokr.

- Veřejné blaho – přínos pro společnost
 - Otevřený přístup
 - Return On Investment
- Soulad s požadavky poskytovatelů finanční podpory a vydavatelů
- Rostoucí úloha dat v různých oblastech života
- Otevřený přístup nejen k vědeckým datům
- Big data
- „citizen science“

Výzkumná data - EU

- EU – Horizon 2020
- OpenAIRE
 - <https://www.openaire.eu/>

Datové politiky – grantové agentury

- EU - Horizon 2020 - <https://www.openaire.eu/ec-policies-and-mandates>
- https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf
 - Open Research Data Pilot (ORDP) – v roce 2017 rozšířen na všechny tematické okruhy
 - Výjimky – ochrana osobních dat, bezpečnost a další

ORDP - pokr.

- Otevřený přístup k digitálním výzkumným datům - opatření, která umožní koncovým uživatelům přístup k datům, jejich vytěžování, používání, reprodukci a šíření prostřednictvím úložiště výzkumných dat bez jakýchkoli poplatků.
- Výzkumná data - data v digitální podobě (zejména fakta nebo čísla), sbíraná za účelem prozkoumání, argumentace, diskuse nebo výpočet; zahrnují statistiky, výsledky experimentů, měření, pozorování vyplývající z práce v terénu, výsledky průzkumu, záznamy rozhovorů a obrázky.

ORDP - pokr.

- Pilot se vztahuje na 2 typy digitálních výzkumných dat:
 - data potřebná k ověření výsledků předložených ve vědeckých publikacích a souvisejících metadatech (tj. údaje popisující uložené výzkumné údaje)
 - další data a související metadata, jak si je určili samotní příjemci v plánu managementu dat
- Repozitář výzkumných dat – online archív výzkumných dat, může být oborový, institucionální, centralizovaný apod.

Datové politiky – grantové agentury

- National Science Foundation, division of Social and Economic Science
- <https://www.nsf.gov/sbe/ses/common/archive.jsp>
 - Povinnost připravit plán zveřejnění dat
 - Kvantitativní – vyčištěná, plně dokumentovaná data uložit do archivu do jednoho roku od ukončení grantu
 - Kvalitativní – zveřejnění je na zvážení tvůrce, vhodná forma, přiměřené náklady
 - Experimentální výzkum – upravená data z experimentů včetně detailního popisu experimentu
 - Matematické/počítačové modely – v některých případech postačuje publikace ve formě časopiseckého článku, u některých modelů musí být připravena dodatečná dokumentace

Datové politiky - časopisy

- [Journal of Applied Psychology](#)
 - Zveřejnění dat k článkům je doporučováno, ne vyžadováno, APA nabízí vlastní repozitář
- [The Quarterly Journal of Economics](#)
 - Zveřejnění dat je požadováno (po přijetí článku), možnost udělit výjimku
- [American Sociological Review](#)
 - Zveřejnění dat je vyžadováno po ukončení projektu nebo zveřejnění hlavních publikačních výstupů, výjimky zdůvodnitelné smluvními ujednáními nebo možným ohrožením účastníků výzkumu

Data

Data = opakovaně interpretovatelná vyjádření informací ve formalizované podobě vhodné pro komunikaci, interpretaci nebo zpracování; mezi příklady dat patří posloupnost bitů, tabulka s čísly, znaky na stránce, nahrávka zvuků pořízená mluvičkou nebo vzorek měsíční horniny (ČSN ISO 14751)

Data = cosi, co je dáno před začátkem argumentace

Observační data

Komputační data

Experimentální data

Záznamy

Výzkumná data = zaznamenaná faktická informace uchovávaná a akceptovaná vědeckou komunitou pro potřeby validování výsledků výzkumu

Obrazová data

Textová data

Zvuková data

Audiovizuální data

Data softwarového charakteru

Numerická data

Data nedigitální povahy

Životní cyklus dat

- Tvorba příp. získání dat
 - Zpracování dat
 - Analýza dat
 - Dlouhodobé uchování dat
 - Zpřístupnění dat
 - Další využití dat
-
- (podle PAVLÁSKOVÁ, Eliška, Analýza výzkumných dat na základě fondu disertačních prací Univerzity Karlovy ..., 2016, dizertační práce)

Mapování / transformace dat

- Surová data
- Odvozená data
- Výsledná data

Vývoj pojmu datová gramotnost a její souvislosti

- Různá označení pro datovou gramotnost
 - **Data literacy**
 - **Data information literacy** (Carlson, 2011)
 - Research data literacy (Quin, 2010)
 - Data management literacy (Haendel, 2012)
 - Science data literacy (Schneider, 2013)

Vývoj pojmu datová gramotnost a její souvislosti

- obecně nejdříve se tohoto pojmu používalo v sociologii (sociologických výzkumech)
- Hunt, 2004 – datová gramotnost = statistická, kvantitativní gramotnost
- Shield, 2004 – datová gramotnost je součástí statistické gramotnosti
- Qin, 2010 – vědecká datová gramotnost = schopnost shromažďování, uchovávání, využívání a vytváření jakýchkoliv vědeckých dat

Datová gramotnost - kompetence

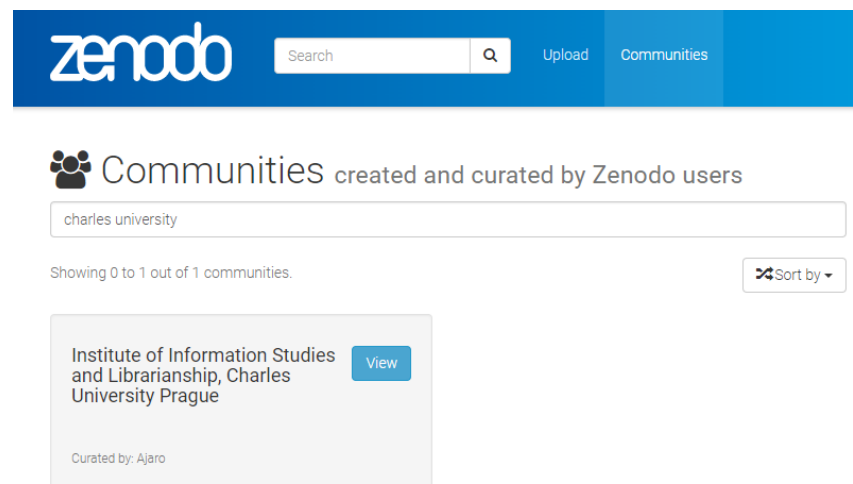


Životní cyklus dat

- Tvorba příp. získání dat
- Zpracování dat
- Analýza dat
- Dlouhodobé uchování dat
- Zpřístupnění dat
- Další využití dat

Hledání, získávání dat – datové repozitáře

- [Zenodo](#) – datový repozitář CERNu, partner OpenAIRE



- [Dryad](#)

Datové časopisy

- Open Data Journals

- <https://www.fosteropenscience.eu/foster-taxonomy/open-data-journals>

- Příklad: Giga Science

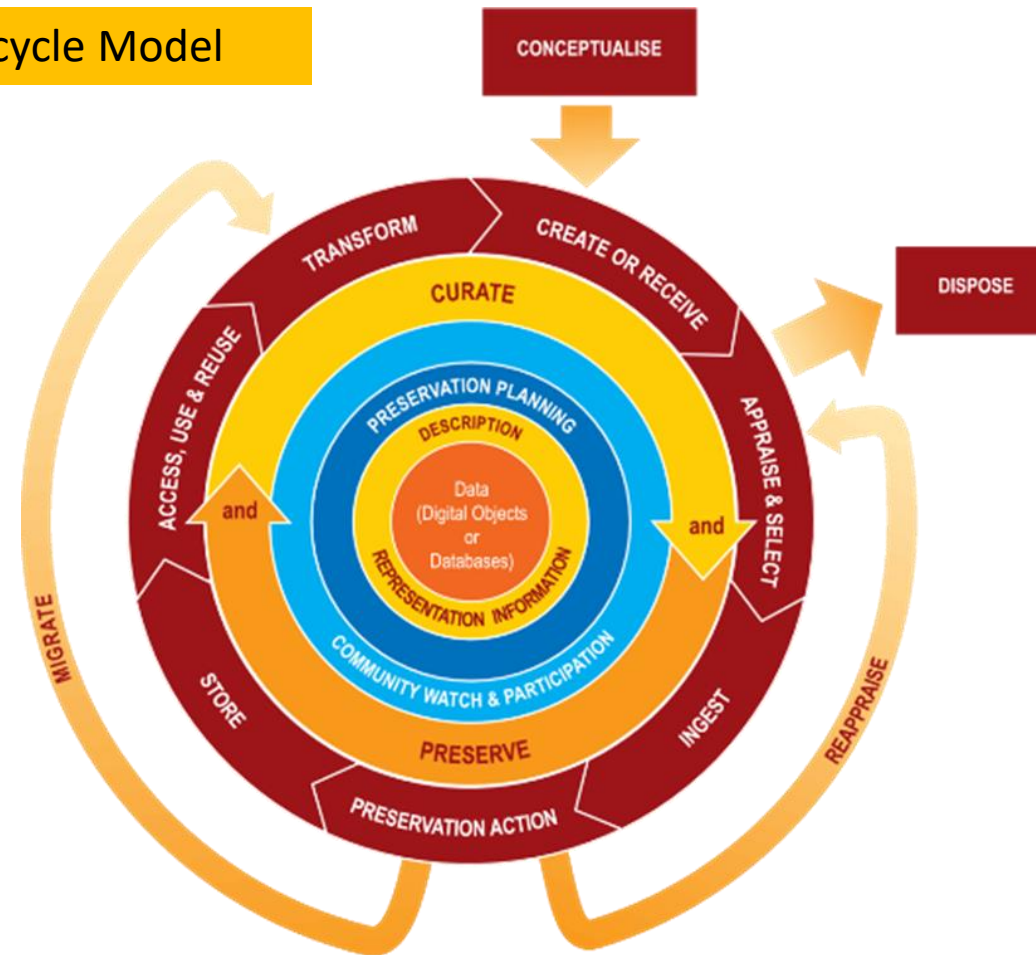
- <https://academic.oup.com/gigascience>

Vytváření dat

- Znalost výzkumných metod
- Výběr a aplikace vhodné výzkumné metody
- Oborové rozdíly ve zdrojích dat
 - Humanitní vědy – záznamy
 - Společenské vědy – záznamy, pozorování
 - Přírodní vědy – experimenty
- Ukládání dat ve vhodném formátu
- Zaznamenání důležitých informací týkajících se okolností vzniku dat - metadata

Management dat

DCC Curation Lifecycle Model



Management dat - metadata

- Popisná – zajišťují identifikaci a nalezení
- Technická – popis technické infrastruktury, která byla použita k vytvoření nebo přístupu k datům
- Administrativní
- Metadata pro užívání
- Metadata pro uchování

Metadata pro výzkumná data

- Oborové rozdíly, vlastní specifikace
 - <http://www.dcc.ac.uk/resources/metadata-standards>
- Obecné struktury – např. pro harvestování do OpenAIRE
 - <https://guidelines.openaire.eu/en/latest/data/index.html>
- DataCite – v OpenAire - příklad
 - https://guidelines.openaire.eu/en/latest/data/use_of_datacite.html

Management dat

- Udržování vztahů mezi datovými soubory, verzování
- Dokumentace
- Znalost typů a datových formátů
- Dlouhodobé uchování dat
 - Výběr dat pro dlouhodobé uchování
- Management rizik
- Plán managementu dat (data management plan, DMP)
 - <https://www.openaire.eu/rdm-handbook>

Plán managementu dat podle ORDP (Horizon 2020)

- PMD - musí podporovat životní cyklus správy pro všechna data, která budou shromážděna, zpracována nebo vygenerována v daném projektu
- FAIR use – viz dále
- metodologie a standardy
- Sdílení dat – otevřený přístup, pokud ne – proč?
- Správa a uchování dat
- PMD by měl být aktualizován v průběhu projektu upřesňován a aktualizován. Nové verze by měly být vytvořeny vždy, když dojde k důležitým změnám projektu (např. nové soubory údajů, změny v zásadách konsorcia atd.), alespoň v polovině období (jako součást přezkumu, pokud existuje) a na konci projektu.

Plán managementu dat (DMP)

- plán, jak budou data vytvářena/získávána, jak s nimi bude zacházeno v průběhu projektu i po jeho skončení
- obsah (podle [ICPSR, Data Management&Curation](#))
 - popis dat – jaká data budou shromažďována
 - existující data – jsou k projektu relevantní už existující data?
 - formát
 - metadata – použité standardy
 - uložení, zálohování – způsob uložení a zálohování dat

Plán managementu dat (DMP) – pokr.

- bezpečnost – technické a procedurální informace k zajištění bezpečnosti dat
- odpovědnost
- autorská práva, duševní vlastnictví
- přístup a sdílení – jak budou data sdílena, embargo, přístupová práva (otevřená data x omezený přístup)
- potenciální uživatelé
- výběr a doba uchování
- dlouhodobá archivace
- etika, ochrana soukromí
- rozpočet potřebný na management dat
- organizace dat – práce s daty v průběhu projektu, verzování, standardní jména souborů
- zajištění kvality
- požadavky grantové agentury nebo státní politiky

Pomůcka - DMPonline

- <https://dmponline.dcc.ac.uk/> - autorizace
- Volné šablony: https://dmponline.dcc.ac.uk/public_templates

Užívání dat

- Analýza dat
- Vizualizace dat
- Využití dat při argumentaci
- Prezentace kvantitativních dat
- Etické a právní aspekty
- Citování dat

Sdílení dat

- Spolupráce v týmech
- Otevřená data
- Etické a právní aspekty
- Ochrana osobních dat

FAIR principy sdílení dat

- Findable
 - Identifikátory
 - Metadata
 - Indexace nebo registrace v prohledatelném zdroji
- Accessible
 - (meta)data jsou přístupná prostřednictvím standardizovaného komunikačního protokolu

FAIR principy sdílení dat

- Interoperable
 - (meta)data užívají formální, dostupný, sdílený a široce aplikovatelný jazyk reprezentace znalostí
 - (meta)data užívají slovníky, které odpovídají FAIR principům
 - (meta)data obsahují reference na jiná (meta)data
- Re-usable
 - Kvalitní metadata umožňující propojení, užívání a citování dat

FAIRsharing.org

- <https://fairsharing.org/>

Jak jsme na tom s datovou gramotností na UK

- Data literacy and research data management survey – mezinárodní srovnávací studie datové gramotnosti a managementu výzkumných dat
 - Na UK proběhla v lednu/únoru 2017
 - 30 dalších zemí
 - 1434 kompletních dotazníků od akademických pracovníků a doktorandů UK
- Dále navázal kvalitativní výzkum v sociálních vědách

Jak jsme na tom s datovou gramotností na UK

61,3% je obeznámeno s
pojmem metadata

72,3% používá
standardizovaný styl
citování metadat

4,5% použilo plán
managementu dat

41,6% je obeznámeno s
požadavky otevřeného
přístupu

52% je obeznámeno s
konceptem DOI

64,5% je ochotno sdílet
data

46,3% nikdy nepoužívá
metadata

36,5% nikdy nepoužívá
konvenci nebo standard
pro pojmenování souborů

51% nikdy nepoužívá
techniky pro verzování
souborů

Faktory ovlivňující sdílení dat

- Původce dat
- Příjemce dat
- Charakteristika dat
- Výzkumná komunita
- Organizace výzkumu v daném oboru
- Infrastruktura pro sdílení dat

Původce dat (data donor/provider)

- Individuální charakteristiky – věk, postavení, výzkumná a publikační aktivita, počítačové dovednosti, postoj k otevřené vědě
- Vnitřní motivace – altruismus, zájem o rozvoj oboru
- Vnější motivace – přínos pro kariéru, zvýšení citovanosti
- Bariéry – obava z ohrožení kariéry, nedostatek času

Příjemce/uživatel dat (data recipient/re-user)

- Kdo je uživatel? – kolega z oboru/vědec z jiného oboru/veřejnost?
- Jak a k čemu data použije?
 - Zneužití dat – krádež dat, plagiátorství
 - Špatná interpretace dat – neznalost metod, nedostatečná dokumentace

Charakteristika dat

- Data v soc. vědách mají mnoho různých forem, ne všechny jsou v digitální podobě
- Kvantitativní x kvalitativní data
- Etika
 - Poškození účastníků výzkumu
 - Informovaný souhlas
 - Osobní údaje, citlivé informace, GDPR
 - Anonymizace – časově náročná, hrozí ztráta kontextu

Výzkumná komunita

- Převládající „kultura sdílení“ – velmi důležitá
- Důvěra
- Neformální sdílení na základě vyžádání nebo na základě osobních kontaktů je poměrně časté

Organizace výzkumu v daném oboru

- Politika grantových agentur
- Odborné společnosti
- Požadavky oborových časopisů
- Existující oborové standardy – metadata, citování

Infrastruktura pro sdílení dat

- Dostupnost repozitáře
- Uživatelská vlídnost
- Stupeň kontroly tvůrce nad svými daty
 - Možnost stanovit podmínky použití
 - Ochrana před zneužitím
- Udržitelnost repozitáře

Literatura

TENOPIR, Carol, Elisabeth D. DALTON, Suzie ALLARD, Mike FRAME, Ivanka PJESIVAC, Ben BIRCH, Danielle POLLOCK a Kristina DORSETT. Changes in data sharing and data reuse practices and perceptions among scientists worldwide. PLoS One [online]. 2015, 10(8) [cit. 2016-09-29]. ISSN 19326203. Dostupné z: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0134826>

CALZADA PRADO, Javier a Miguel Ángel MARZAL. Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. Libri [online]. 2013, 63(2), - [cit. 2017-01-30]. DOI: 10.1515/libri-2013-0010. ISSN 18658423. Dostupné z: <http://www.degruyter.com/view/j/libr.2013.63.issue-2/libri-2013-0010/libri-2013-0010.xml>

CHRISTENSEN-DALSGAARD, Birte. Ten recommendations for libraries to get started with research data management [online]. b.r. [cit. 2017-10-26]. Dostupné z: <http://libereurope.eu/wp-content/uploads/The%20research%20data%20group%202012%20v7%20final.pdf>

KOLTAY, Tibor. Data literacy for researchers and data librarians. Journal of Librarianship and Information Science [online]. 2017, 49(1), 3-14 [cit. 2017-04-09]. DOI: 10.1177/0961000615616450. ISSN 09610006. Dostupné z: <http://journals.sagepub.com/doi/10.1177/0961000615616450>

CARLSON, Jacob, Michael FOSMIRE, C.C. MILLER a Megan SAPP NELSON. Determining Data Information Literacy Needs: A Study of Students and Research Faculty. Libraries and the Academy. 2011, 11(2), 629-657.

QIN, Jian a John D'IGNAZIO. Lessons learned from a two-year experience in science data literacy education. In: International Association of Scientific and Technological University Libraries, 31st Annual Conference. [online]. 2010 [cit. 2017-04-09].

HAENDEL, Melissa A., Nicole A. VASILEVSKY a Jacqueline A. WURZ. Dealing with Data: A Case Study on Information and Data Management Literacy. Plos Biology [online]. 2012, 10(5) [cit. 2017-04-09]. DOI: 10.1371/journal.pbio.1001339.

SCHNEIDER, René. Research data literacy. In: KURBANOGLU, Serap, Esther GRASSIAN, Diane MIZRACHI, Ralph CATTS a Sonja ŠPIRANEC. Worldwide Commonalities and Challenges in Information Literacy Research and Practice: European Conference, ECIL 2013, Istanbul, Turkey, October 22-25, 2013 [online]. Springer, 2013 [cit. 2017-10-26]. ISBN 978-3-319-03918-3. ISSN 1865-0937.

HUNT, Karen. The Challenges of Integrating Data Literacy into the Curriculum in an Undergraduate Institution. IASSIST Quarterly. 2004, (), 12-16.

SHIELD, Milo. Information Literacy, Statistical Literacy and Data Literacy. IASSIST Quarterly. 2004, (), 6-11.

FRANK, E. P. a N. PHARO. Academic Librarians in Data Information Literacy Instruction: A Case Study in Meteorology. College & Research Libraries [online]. 2016, 77(4), 536-552 [cit. 2017-04-09]. DOI: 10.5860/crl.77.4.536. ISSN 00100870. Dostupné z: <http://crl.acrl.org/cgi/doi/10.5860/crl.77.4.536>

SAPP NELSON, Megan R. A Pilot Competency Matrix for Data Management Skills: A Step toward the Development of Systematic Data Information Literacy Programs. Journal of eScience Librarianship [online]. 2017, 6(1), - [cit. 2017-04-09]. Dostupné z: <https://doi.org/10.7191/jeslib.2017.1096>

PAVLASKOVA, Eliska. Analýza výzkumných dat na základě fondu disertačních prací Univerzity Karlovy v Praze s ohledem na dlouhodobé uložení digitálních objektů [online]. Praha, 2016 [cit. 2017-04-12]. Dostupné z: <https://is.cuni.cz/webapps/zzp/detail/103851/>. Dissertation thesis. Univerzita Karlova. Filozofická fakulta. Vedoucí práce RNDr. Pavel Krbec, Ph.D.

DROBÍKOVÁ, Barbora, Adéla JAROLÍMKOVÁ a Martin SOUČEK. Data literacy of Charles University PhD students : are they prepared for their research careers?. In: ŠPIRANEC, Sonja, Serap KURBANOGLU, Joumana BOUSTANY, Esther GRASSIAN, Diane MIZRACHI, Lorie ROY a Denis KOS. The Fifth European Conference on Information Literacy (ECIL) : abstracts. Saint-Malo: Information Literacy Association, 2017, s. 41. ISBN 978-2-9561952-0-7.

JAROLÍMKOVÁ, Adéla, DROBÍKOVÁ, Barbora. Data Sharing in Social Sciences. In: ŠPIRANEC, Sonja et al. (ed.). The Sixth European Conference on Information Literacy (ECIL): abstracts. Oulu: University of Oulu, 2018, s. 46. ISBN 978-952-62-2013-0.