

Výzkumná data na Univerzitě Karlově

Adéla JAROLÍMKOVÁ, Barbora DROBÍKOVÁ, Martin SOUČEK

Univerzita Karlova, Filozofická fakulta, Ústav informačních studií a knihovnictví
{adela.jarolimkova, barbora.drobikova, martin.soucek}@ff.cuni.cz

INFORUM 2017: 23. ročník konference o profesionálních informačních zdrojích
Praha, 30.-31.5.2017

Abstrakt

Správa a sdílení výzkumných dat jsou jedním z témat, kterým se akademické knihovny nově zabývají v rámci podpory vědy. Praxe v České republice však dosud nebyla dostatečně zmapována. Příspěvek shrnuje základní výsledky dotazníkového šetření, které proběhlo počátkem tohoto roku na Univerzitě Karlově v rámci mezinárodní srovnávací studie zaměřené na výzkumná data. Cílem výzkumu bylo zjistit, jaká data vědci při svém výzkumu využívají a vytvářejí, zda data sdílejí nebo jsou ochotni sdílet, do jaké míry jsou obeznámeni s principy managementu výzkumných dat, a to v různých vědních oborech. Dotazník byl původně vytvořen pro studii ve Velké Británii, Francii a Turecku, následně byl využit ve více než patnácti dalších zemích včetně České republiky.

V rámci Univerzity Karlovy bylo získáno celkem 1434 kompletních dotazníků. Výsledky ukazují, že většina respondentů svá data nějakým způsobem sdílí, byť ne otevřeně (zcela otevřený přístup ke svým datům poskytuje pouze 14% respondentů), nicméně téměř polovina má obavy z problémů, které mohou v souvislosti se sdílením dat vzniknout, zejména z možnosti zneužití či chybné interpretace dat. Ačkoliv jsou respondenti obeznámeni se základními koncepty z oblasti managementu dat, jako jsou metadata, DOI, personální identifikátory aj., pro své stávající projekty nemají plán managementu dat a nejsou si jisti, zda takový plán je užitečný. Značná nejistota panuje také v oblasti institucionálních doporučení týkajících se managementu dat. Celkově z výsledků i připojených komentářů vyplývá, že tato oblast je pro vědce důležitá, avšak na institucionální úrovni zatím nedořešená.

Publikace vznikla v rámci programu UK Progres č. Q15 s názvem „Životní dráhy, životní styly a kvalita života z pohledu individuální adaptace a vztahu aktérů a institucí“.

1. Úvod

Správa a sdílení výzkumných dat jsou jedním z témat, kterým se knihovny, zejména akademické, nově zajímají v rámci podpory vědy. Pracovní skupina pro e-vědu a management výzkumných dat Asociace evropských vědeckých knihoven (LIBER) vydala již v roce 2012 doporučení knihovnám týkající se služeb na podporu managementu dat (Dalsgaard, 2012).

Doporučení LIBER zahrnují

1. Nabídnout podporu managementu dat
2. Angažovat se ve vývoji metadatových a datových standardů
3. Vytvořit pozice datových knihovníků
4. Aktivně se podílet na tvorbě institucionální „datové“ politiky
5. Vyjednávat a vytvářet partnerství s vědci, výzkumnými skupinami a archivy na podporu infrastruktury pro přístup k datům a jejich sdílení
6. Podpora životního cyklu dat
7. Podpora správného citování dat
8. Poskytovat institucionální datový katalog či repozitář
9. Zabývat se oborovými specifiky managementu dat
10. Nabídnout nebo zprostředkovat bezpečné uložení dat.

Aby bylo možné efektivně zaměřit služby podpory dat, je nejprve nutné zmapovat potřeby vědců pracujících s výzkumnými daty a stávající praxi v managementu dat. Praxe v České republice nebyla v tomto směru dosud popsána, výzkumnými daty na Univerzitě Karlově se systematicky zabývá pouze E. Pavlásková ve své dizertaci zaměřené na analýzu výzkumných dat na základě fondu disertačních prací UK s ohledem na dlouhodobé uložení digitálních objektů (Pavlásková, 2016).

Následující příspěvek shrnuje základní výsledky dotazníkového šetření, které proběhlo počátkem tohoto roku na Univerzitě Karlově v rámci mezinárodní srovnávací studie zaměřené na výzkumná data. Cílem výzkumu bylo zjistit, jaká data vědci při svém výzkumu využívají a vytvářejí, zda data sdílejí nebo jsou ochotni sdílet, do jaké míry jsou obeznámeni s principy managementu výzkumných dat, a to v různých vědních oborech.

2. Metody

V průzkumu byla využita česká verze dotazníku, který byl původně vytvořen pro studii ve Velké Británii, Francii a Turecku, a jehož částečné výsledky byly prezentovány na konferenci ECIL 2016 společně s výzvou ke spolupráci na mezinárodním šetření (Chowdhury, 2016). Dotazník byl následně přeložen a využit ve více než patnácti dalších zemích včetně České republiky, což umožní kromě popisu situace u nás také v budoucnu srovnání se zahraniční praxí. Dotazník je zaměřen na vybrané otázky datové gramotnosti a managementu výzkumných dat, které lze shrnout do následujících tří oblastí:

- Vytváření a využívání dat (typy dat, jejich objem atd.)
- Postoje ke sdílení dat a chování při sdílení dat
- Povědomí o různých aspektech sdílení dat (metadata, identifikátory, verzování) a praxe sdílení dat

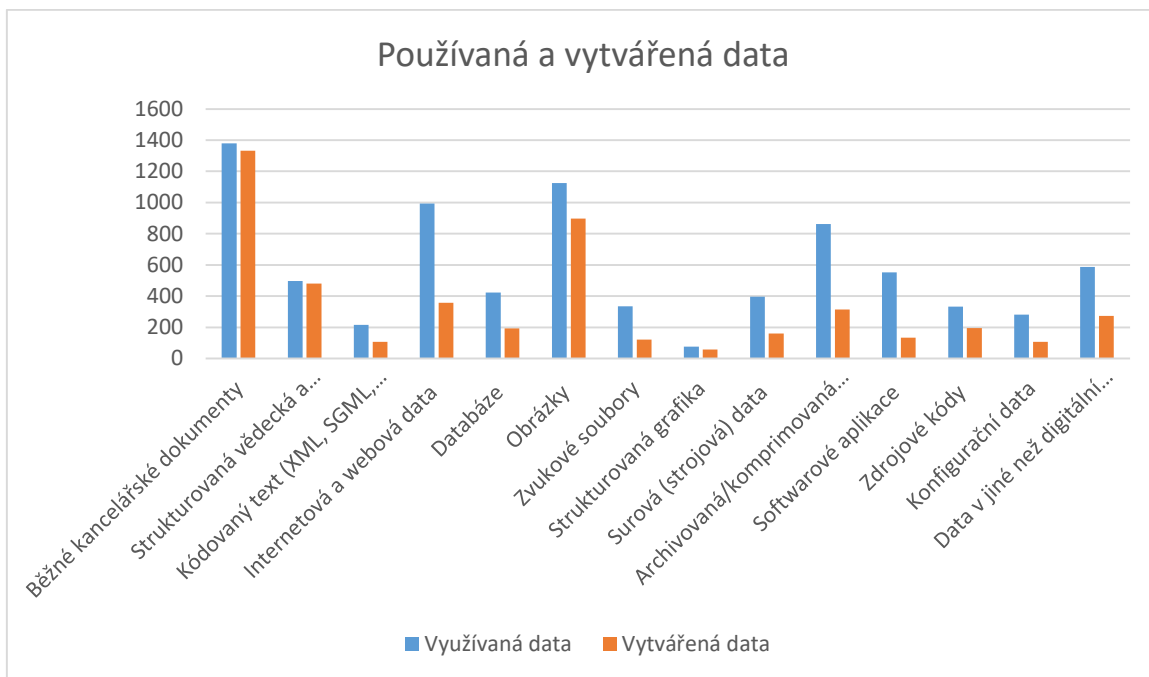
3. Výsledky

Dotazníkem byli osloveni akademičtí a vědečtí pracovníci a doktorandi Univerzity Karlovy, což představuje více než 10 000 oslovených. Na výzvu zareagovalo celkem 2381 respondentů, což naznačuje, že se vědci UK zajímají o problematiku výzkumných dat, avšak část dotazníků zůstala nedokončená, patrně proto, že otázky byly pro některé respondent příliš složité. Nicméně se podařilo získat 1434 kompletních odpovědí, z toho 603 (42%) od akademiků a 831 (58%) od doktorandů. Nejvíce zastoupená byla věková skupina 26-35 let (52,2%), následovaná skupinou 36-45 let (21,5%). Co se týče oborů, nejvíce respondentů bylo zaměřeno na biologii (220, resp. 15,3%), klinickou medicínu (138, resp. 9,62%) fyziku a vědy týkající se jazyků a literatury. Zastoupeny byly ale i další přírodní, společenské a humanitní vědy.

3.1 Vytváření a využívání dat

Data byla v rámci výzkumu chápána velmi široce, jako jakákoliv informace uložená nejen v digitální formě, a to včetně textu, čísel, obrázků, videa nebo filmu, zvuku, softwaru, algoritmů, rovnic, animací, modelů, simulací, atd. Tento přístup umožňuje podchytit i unikátní data ze společenských a humanitních věd, která jsou často pouze v analogové podobě (Pavlásková, 2016).

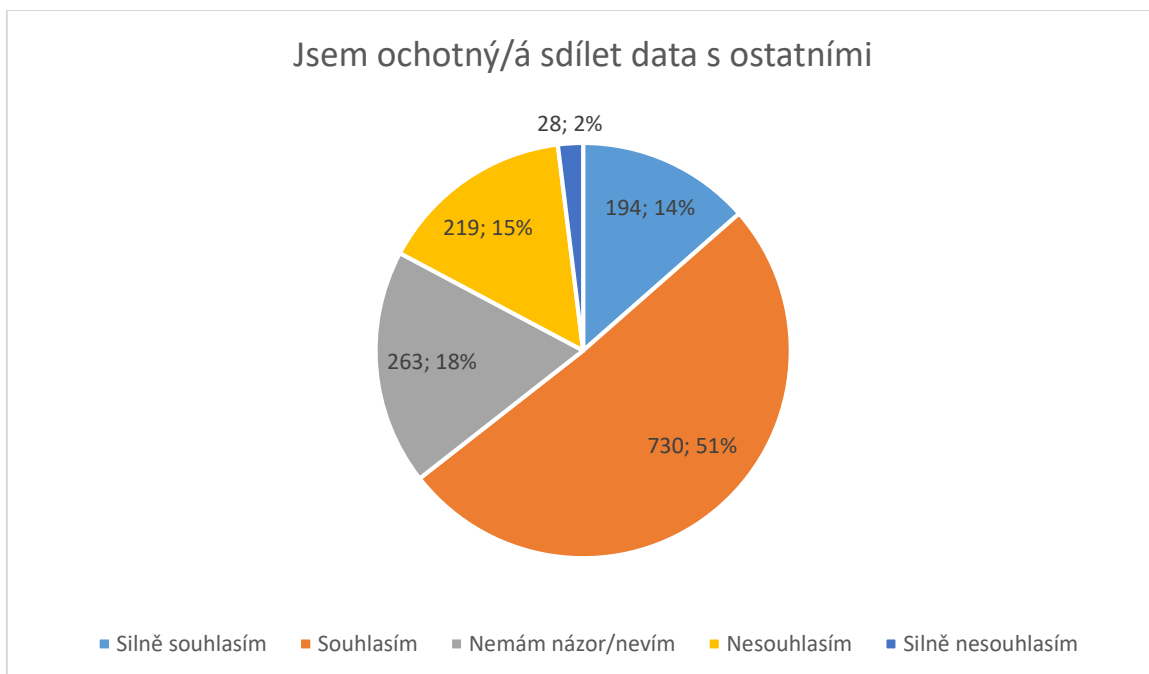
Na obr. 1 je vidět, v jaké míře jsou různé typy dat v rámci výzkumu využívány, a také vytvářeny. Objem používaných i vytvářených dat se nejčastěji pohybuje v megabytech a gigabytech.



Obr. 1. Typy používaných a vytvářených dat

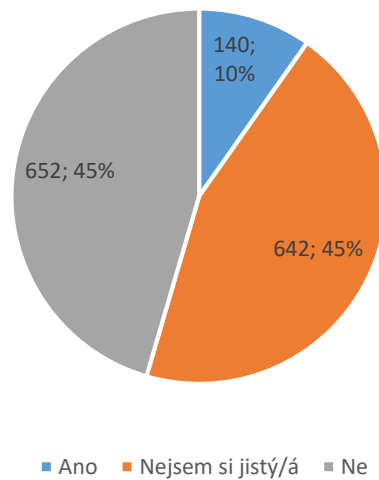
3.2 Sdílení dat

Jeden z okruhů otázek byl zaměřen na postoje vědců ke sdílení dat a současnou praxi v tomto směru. Výsledky ukazují, že většina vědců je ochotna data sdílet (obr. 2) a také je nějakým způsobem sdílí, pouze 17% nesdílí data vůbec. Zcela otevřený přístup ke svým datům však poskytuje pouze 14% respondentů, a pouze 9,8% je přesvědčeno, že je univerzita aktivně vede ke sdílení dat v režimu otevřeného přístupu (obr. 3).



Obr. 2. Ochota sdílet data

Vede Vás Vaše univerzita aktivně ke sdílení dat v režimu otevřeného přístupu?



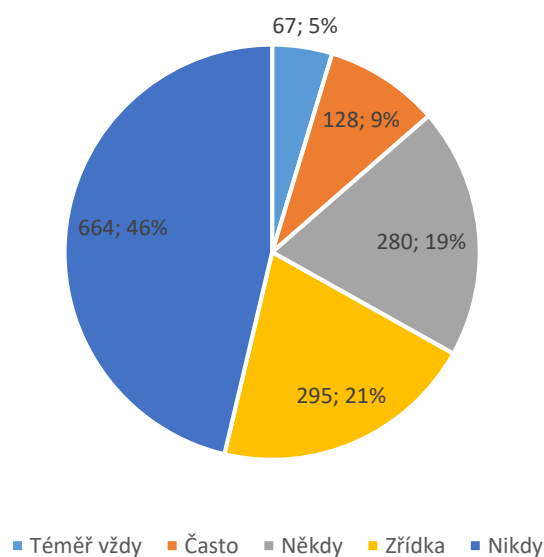
Obr. 3. Vnímání podpory univerzity v otevřeném sdílení dat

I přes ochotu data sdílet jsou si vědci vědomi možných rizik, která s sebou sdílení nese. Obavy se týkají především možné chybné interpretace dat (35,6%), jejich zneužití (30,5%) a právních a etických aspektů sdílení dat (30,4%). Naopak žádné obavy nemá 35% respondentů.

3.3 Povědomí o různých aspektech sdílení dat

Aby bylo možné data sdílet, vyhledávat a využívat, je především nutné, aby byly datové soubory opatřeny odpovídajícími metadaty. S pojmem metadata je sice obeznámeno 61% respondentů, pro popis svých vlastních dat je však užívá pouze menšina (obr. 4).

Jak často obvykle používáte metadata k popisu svých dat?



Obr. 4. Využívání metadat

Naprosté minimum respondentů také použilo někdy plán managementu dat (DMP) pro své projekty (4,5% v minulosti, 5,3% má DMP pro aktuální projekt), o užitečnosti DMP je přesvědčeno pouze 16% vědců, ostatní si nejsou jisti (73%) nebo DMP za přínosný nepovažují (11%). V komentářích k dotazníku se objevily v této souvislosti zejména obavy z nárůstu administrativy.

Co se týče ukládání dat, v současnosti uchovává většina vědců svá data na svých zařízeních (94%), někteří využívají kromě toho také cloudová (31%) či centrální úložiště (23%) v rámci univerzity (centrální univerzitní repozitář sice zatím není k dispozici, ale v rámci fakult či ústavů mohou dílčí řešení existovat. Třetina respondentů využívá také úložiště externí. U otázky, kde by data měla být uchovávána, zvolilo 76,6% respondentů odpověď „Na Vaší univerzitě“.

4. Závěr

Ačkoliv jsme si vědomi toho, že výsledky výzkumu mohou být zkresleny tím, že na dotazník reagovali ti, kterých se problematika výzkumných dat bezprostředně dotýká, a nelze tedy výsledky zobecňovat, získali jsme alespoň základní vhled do problematiky výzkumných dat na Univerzitě Karlově, který může sloužit jako podklad jednak pro další výzkum, jednak pro utváření institucionálních pravidel týkajících se výzkumných dat. Z výsledků je zřejmé, že otázky managementu výzkumných dat zajímají celou řadu vědců Univerzity Karlovy a bude nutné je na úrovni univerzity v nejbližší době řešit,

Vědci jsou ochotni svá data sdílet, část z nich tak už v současnosti činí, i když si jsou vědomi možných problémů souvisejících s otevřeným přístupem k datům. Jsou obeznámeni s pojmem metadata, ale pro popis svých datových souborů používá metadata jen malá část z nich, stejně tak nemají pro své projekty plán managementu dat.

Otvírá se zde velký prostor pro akademické knihovny, které mohou vědcům pomoci s různými aspekty managementu dat v duchu zmíněných doporučení LIBER, ale také připravit podpůrné a školící materiály.

Použitá literatura

CHOWDHURY, G., G. WALTON, S. KURBANOĞLU, Y. ÜNALA J. BOUSTANY. Information Practices for Sustainability: Information, Data and Environmental Literacy. In Spiranec, S., Kurbanoğlu, S., Landová, H. et al (Eds), *The Fourth European Conference on Information Literacy (ECIL)*. Prague: Association of Libraries of Czech Universities, 2016.

CHRISTENSEN-DALSGAARD, Birgitte et al. *Ten recommendations for libraries to get started with research data management* [online]. [cit. 2017-05-03]. Dostupné z: <http://libereurope.eu/wp-content/uploads/The%20research%20data%20group%202012%20v7%20final.pdf>

PAVLÁSKOVÁ, Eliška. Analýza výzkumných dat na základě fondu dizertačních prací Univerzity Karlovy v Praze s ohledem na dlouhodobé uložení digitálních objektů [online]. Praha, 2016 [cit. 2017-05-06]. Univerzita Karlova, Filozofická fakulta. Dizertační práce. Dostupné z: <https://is.cuni.cz/webapps/zzp/detail/103851/>