

POSTOJE AKADEMICKÝCH PRACOVNÍKŮ UNIVERZITY KARLOVY KE SDÍLENÍ DAT

Adéla Jarolímková

Adela.jarolimkova@ff.cuni.cz

Univerzita Karlova, Filozofická fakulta, Ústav informačních studií a knihovnictví, Česká republika

Barbora Drobíková

barbora.drobikova@ff.cuni.cz

Univerzita Karlova, Filozofická fakulta, Ústav informačních studií a knihovnictví, Česká republika

Martin Souček

martin.soucek@ff.cuni.cz

Univerzita Karlova, Filozofická fakulta, Ústav informačních studií a knihovnictví, Česká republika

This paper is licensed under the Creative Commons licence: CC-BY-SA-4.0 (<http://creativecommons.org/licenses/by-sa/4.0/>).

Abstrakt

Management a sdílení dat jsou v současnosti nedílnou součástí výzkumné činnosti. V rámci Univerzity Karlovy jsme provedli průzkum, který se týkal vybraných aspektů práce s výzkumnými daty a postojů vědců k managementu a sdílení dat. Bylo vyhodnoceno celkem 596 dotazníků od akademických pracovníků, rozdělených do čtyř základních oborů - humanitní vědy, lékařství, přírodní vědy a sociální vědy. Důležitým zjištěním výzkumu je, že akademici a vědci UK jsou ochotni sdílet svá výzkumná data, vnímají však různá rizika, která jsou se sdílením a zejména s otevřeným přístupem k datům spojená. Pro podporu otevřeného přístupu k datům na UK bude tedy potřeba vytvořit bezpečnou infrastrukturu pro sdílení dat, která by vyhovovala specifikům jednotlivých oborů, ale i zabezpečit podporu managementu dat např. v akademických knihovnách, aby vědci nebyli zatíženi dalšími administrativními povinnostmi.

Klíčová slova

Management výzkumných dat, sdílení výzkumných dat, otevřený přístup, vědci, akademičtí pracovníci

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“.

Úvod

Práce s daty je nedílnou součástí současné vědy. Ke spravování, uchovávání a sdílení dat jsou však třeba kompetence a znalosti, které nejsou doposud běžnou součástí vysokoškolského vzdělávání. Proto se stále více výzkumů zabývá úrovní tzv. datové gramotnosti, tj. souborem kompetencí, které umožňují vyhledávat, interpretovat, kriticky hodnotit, spravovat a eticky využívat data (Calzada Prado 2013), potřebami, které vědci v této oblasti mají, a návrhy vzdělávacích programů (např. Carlson 2011, Haendel 2012, Sapp Nelson 2017 aj.). V našem prostředí se výzkumnými daty zabývala doposud pouze Pavlásková ve své dizertaci (Pavlásková 2016).

Proto jsme využili možnost zúčastnit se mezinárodní srovnávací studie datové gramotnosti a managementu výzkumných dat¹, která nejprve proběhla ve Francii, Turecku a Velké Británii a jejíž první výsledky byly prezentovány na konferenci ECIL (Chowdhury 2016).

Za data je v rámci tohoto výzkumu považována jakákoli informace uložená v nejen digitální formě, a to včetně textu, čísel, obrázků, videa nebo filmu, zvuku, softwaru, algoritmů, rovnic, animací, modelů, simulací atd.

¹ Data ze všech zúčastněných zemí budou zveřejněna na společné webové stránce, data z české části výzkumu jsou dostupná jako DROBIKOVA, Barbora, JAROLIMKOVA, Adela, a Martin SOUCEK, 2017. Data literacy and research data management survey [Data set]. Zenodo. <http://doi.org/10.5281/zenodo.997844>

Metody

V průzkumu byla využita česká verze dotazníku, jeho popis a vybrané výsledky byly již publikovány (Jarolímková et al. 2017, Drobíková et al. 2017). Dotazník byl rozeslán e-mailem všem akademickým pracovníkům a doktorandům UK.

V tomto článku se zaměřujeme na část výsledků, která se týká postojů akademických pracovníků ke sdílení výzkumných dat a jejich stávajícího přístupu ke sdílení a zahrnuje celkem šest otázek (viz. Tab. 1).

Která z následujících možností se vztahuje na Vaše data? (Moje data jsou otevřeně přístupná komukoliv/Moje data jsou otevřeně přístupná pouze mému výzkumnému týmu/Moje data jsou přístupná na vyžádání/Přístup k mým datům je omezen (např. je přístupná pouze určitá část dat)/Moje data nejsou přístupná nikomu)

Máte nějaké obavy týkající se sdílení dat s jinými vědci? (Žádné obavy nemám/Mám obavy ze ztráty náskoku ve výzkumu/Mám obavy z právních a etických aspektů sdílení dat/Mám obavy ze zneužití dat/Mám obavy z chybné interpretace dat/Mám obavy z nedostatku zdrojů (technických, finančních, personálních atd.) pro management výzkumných dat/Mám obavy z neexistence odpovídajících směrnic a opatření chránících má práva v oblasti sdílení dat/Jiné)

Spolupracujete s ostatními vědci a sdílíte data? Ne, nesdílím data/Ano, sdílím data s vědci z mého týmu/Ano, sdílím data s vědci z mé univerzity/Ano, sdílím data s vědci z jiných institucí/Jiné)

Jsem obeznámen/a s požadavky otevřeného přístupu (Silně souhlasím/souhlasím/nemám názor/nesouhlasím/silně nesouhlasím)

Jsem ochotný/á sdílet data s ostatními (Silně souhlasím/souhlasím/nemám názor/nesouhlasím/silně nesouhlasím)

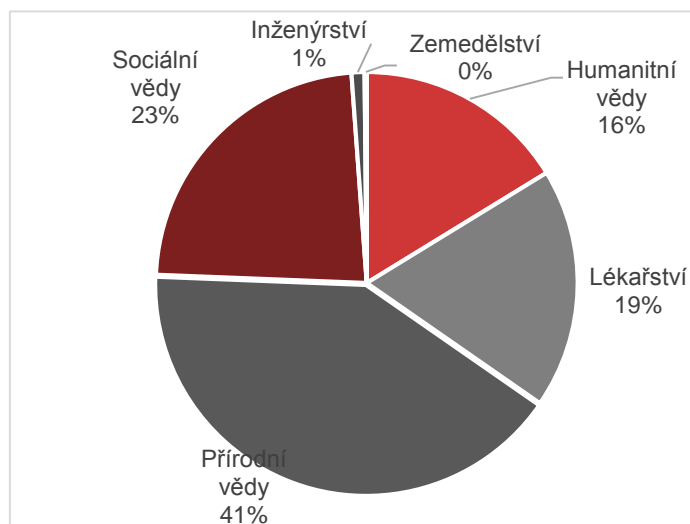
Jestliže budou data sdílena, mohou vyvstat problémy v oblasti etiky (Silně souhlasím/souhlasím/nemám názor/nesouhlasím/silně nesouhlasím)

Tab. 1: Otázky týkající se sdílení výzkumných dat

Při interpretaci dat byla využita popisná statistika a Pearsonův chí-kvadrát test pro hledání souvislostí mezi otázkami na úrovni spolehlivosti $\alpha = 0,05$. V otázkách, kde byla použita 5 stupňová Likertova škála, byly odpovědi pro větší přehlednost sloučeny následovně: silně souhlasím a souhlasím jako ano, nesouhlasím a silně nesouhlasím jako ne.

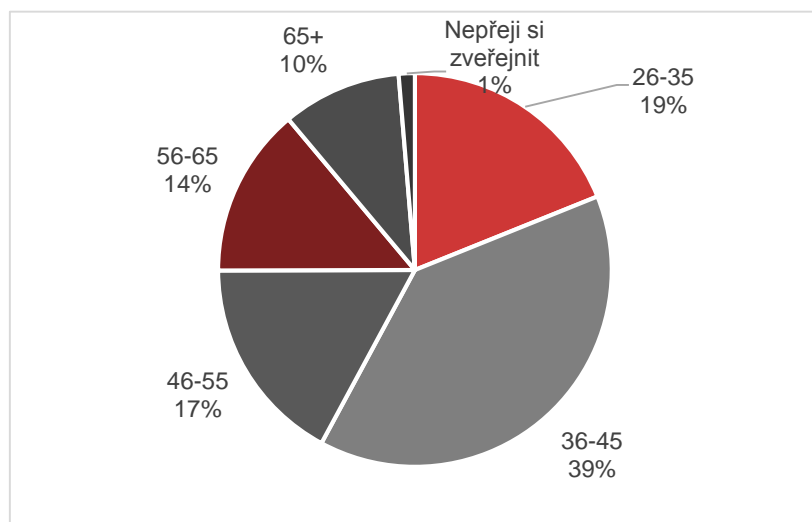
Výsledky

Celkem bylo získáno 2381 odpovědí, z toho však pouze 1434 dotazníků bylo vyplněno kompletně. Vzhledem k tomu, že u nekompletních dotazníků byl vyplněn pouze oddíl demografických charakteristik, byly tyto dotazníky z analýzy vypuštěny. Od akademických pracovníků UK bylo získáno celkem 603 kompletních dotazníků. Pro přehlednost byly dotazníky rozděleny do čtyř základních oborů – humanitní vědy, lékařství, přírodní vědy a sociální vědy (viz. Obr. 1), v odpovědích bylo zastoupeno také inženýrství a zemědělství celkem 7 dotazníky, které však byly z další analýzy vyřazeny, vzhledem k malému zastoupení těchto oborů.



Obr. 1: Respondenti podle oborů

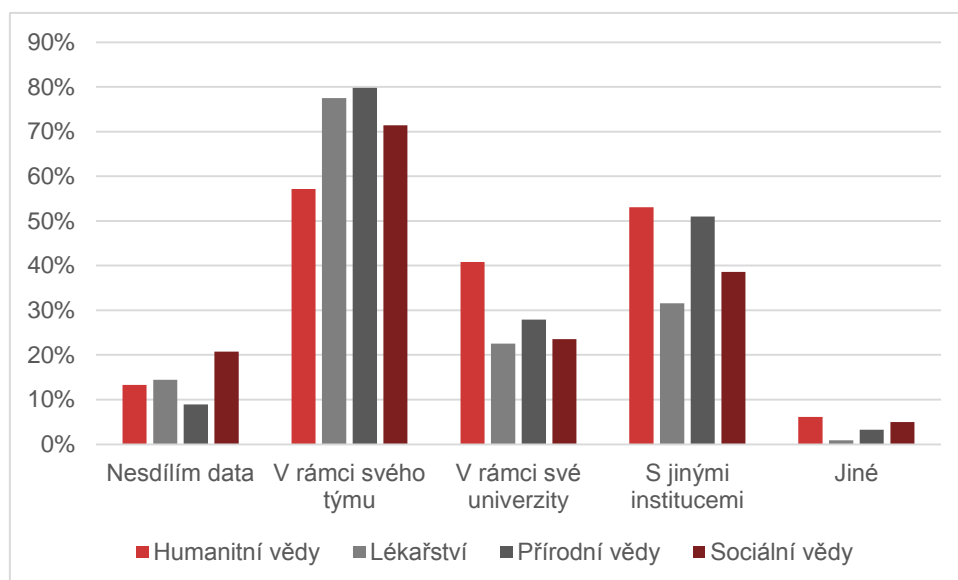
Co se týče věkové struktury respondentů, nejvíce respondentů bylo z kategorie 36-45 let, nejméně byla zastoupena kategorie 65 a více let (obr. 2).



Obr. 2: Respondenti podle věku

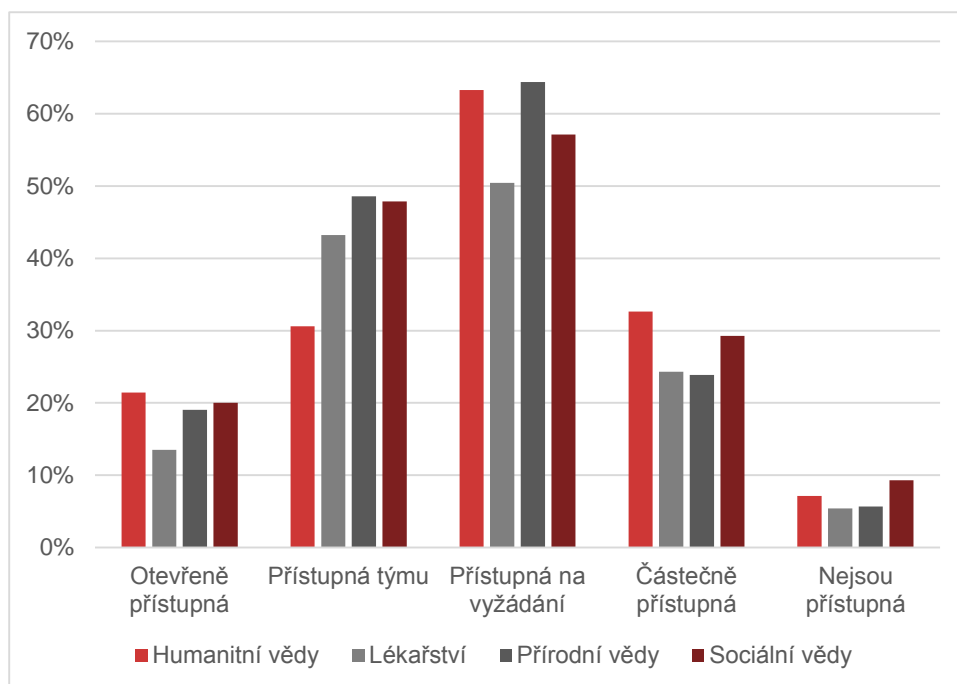
S požadavky otevřeného přístupu je obeznámeno 48% (n=596) respondentů z řad akademiků, což je více než u doktorandů, v jejichž skupině deklarovalo obeznámenost s otevřeným přístupem 36% (n=826).

Většina akademiků v současnosti již svá data sdílí (viz. Obr. 3). Pouze 13% (n=596) uvedlo, že data nesdílí žádným způsobem (obr.). Nejvíce těch, kteří data nesdílí, je ze sociálních věd (21%, n=140), naopak v přírodních vědách je podíl nesdílejících pouhých 9% (n=247). Nejčastěji jsou data sdílena v rámci týmu, s vědci z jiných institucí sdílí svá data 45% respondentů (n=596). V porovnání s výsledky z výsledků publikovaných v Drobíková aj., 2017 akademici sdílejí svá data častěji než doktorandi ($p < 0,001$) a rozdíly byly nalezeny také mezi věkovými kategoriemi, v nejnižší (26-35 let) nesdílí data 8%, zatímco v nejvyšší (65+) celkem 25% respondentů.



Obr. 3: Současná praxe ve sdílení dat

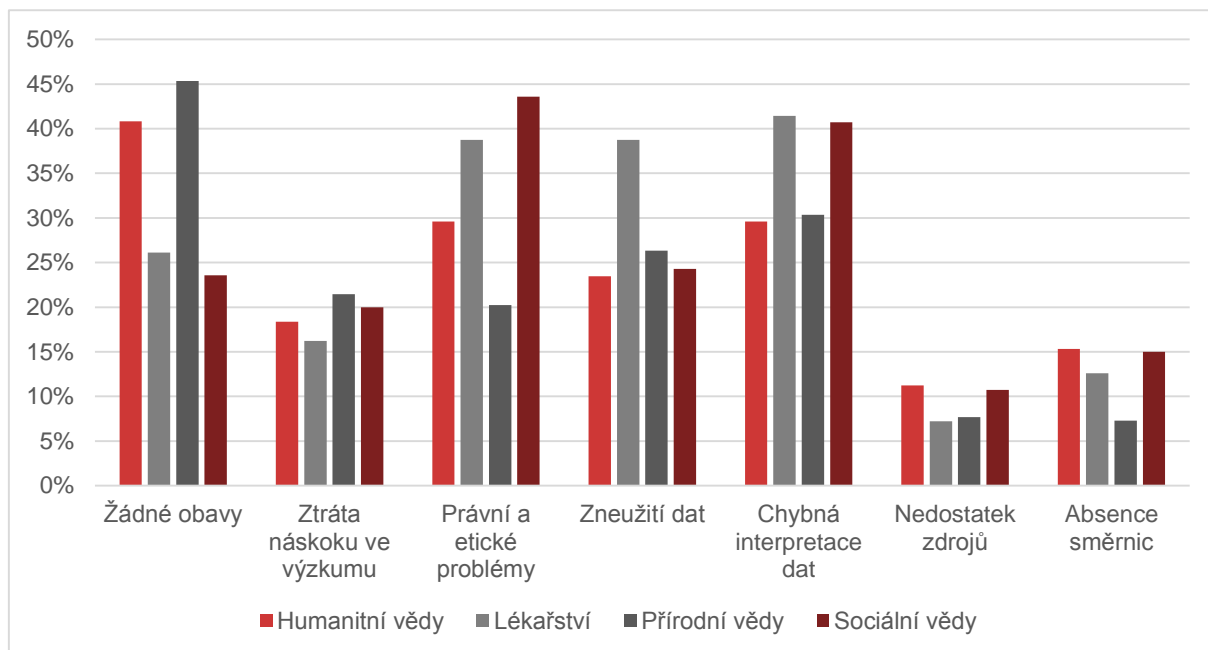
Jak bylo zjištěno už při analýze výsledků skupiny doktorandů, sdílení však neznamená automaticky otevřený přístup k datům (Drobíková et al. 2017). Naopak, otevřený přístup je nejméně častým způsobem, jak akademici svá data sdílí, přičemž zde nejsou podstatné rozdíly mezi obory (viz. Obr. 4). Otevřený přístup ke svým datům častěji poskytují ti akademici, kteří jsou obeznámeni s principy otevřeného přístupu, než ti, kteří s nimi obeznámeni nejsou ($p = 0,002$). Akademici také poskytují svá data otevřeně ve větší míře než doktorandi ($p < 0,001$) (opět ve srovnání s Drobíková et al., 2017). Nejčastěji jsou data přístupná na vyžádání.



Obr. 4: Způsob zpřístupňování dat

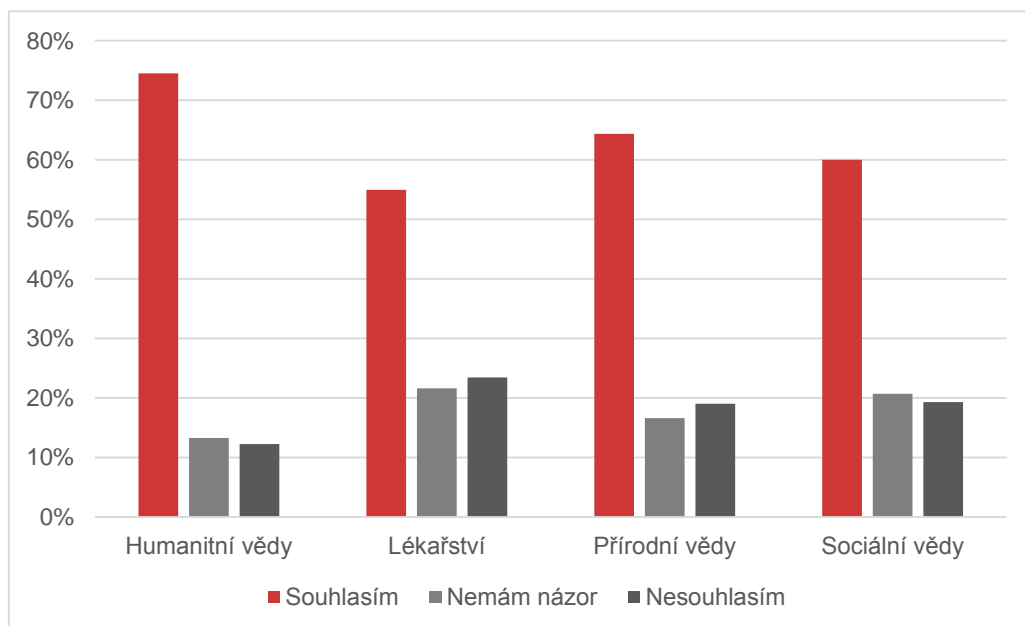
Zajímavé výsledky přinesly odpovědi na otázku týkající se obav souvisejících se sdílením dat (Obr. 5). K této otázce měli respondenti možnost zvolit více odpovědí. Více než třetina respondentů - akademiků (36%, n=596) nemá žádné obavy ze sdílení dat, přičemž je výrazný rozdíl mezi respondenty z humanitních a přírodovědných oborů, kde žádné obavy nemá 41% (n=98), respektive 45% (n=247), na straně jedné a medicínou s 26% (n=111) a sociálními vědami (24%, n=140) na straně druhé. Nejvíce akademiků se obává chybné interpretace dat (35%, n=596), rozdíly mezi obory zde však nejsou statisticky významné. 31% respondentů se obává právních a etických problémů, opět se zde projevuje rozdíl mezi medicínou a sociálními vědami, kde mají respondenti častěji tento druh obav, a ostatními obory. Respondenti z medicíny se také více než ostatní obávají zneužití dat (p=0,004). Pouze malý podíl akademiků se obává nedostatku zdrojů či absence směrnic v oblasti managementu výzkumných dat, z komentářů naopak vyplynulo, že se respondenti spíše obávají nadbytku směrnic a nařízení v této oblasti.

Obavy ovlivňují také ochotu sdílet data. Ti, kteří nemají žádné obavy, jsou současně častěji ochotni svá data sdílet (p<0,001). Negativní vliv na sdílení dat mají obavy z jejich zneužití (p<0,001) a z právních a etických problémů (p=0,003).



Obr. 5: Obavy spojené se sdílením dat

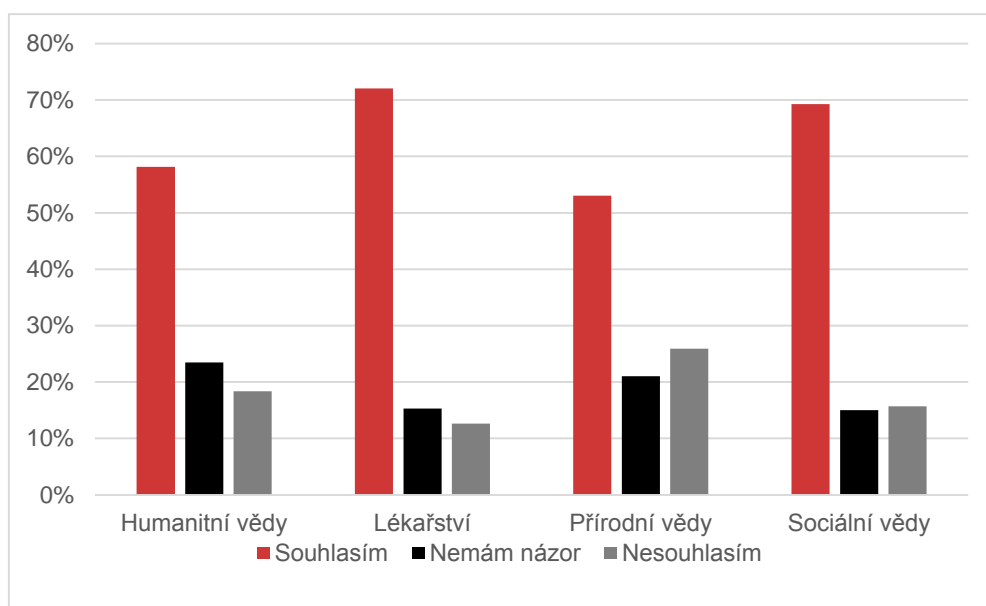
Komplementární otázkou k výše uvedené byl další dotaz, zda jsou vědci ochotni svá data sdílet (Obr. 6). Vědcům byla umožněna pouze odpověď – „silně souhlasím, souhlasím, nemám názor, nesouhlasím, silně nesouhlasím“. Odpovědi „silně souhlasím a souhlasím“ jsme v grafu pro přehlednost spojili a stejně tak odpovědi „nesouhlasím a silně nesouhlasím“. V odpovědích ve všech vědních oblastech převažoval kladný postoj k této otázce. Vědci jsou ochotni svá data sdílet. Nejsilněji je postoj souhlasím zastoupen u vědců z humanitních věd (74% souhlasím ku 12% nesouhlasím, n=98). Naopak nejméně u vědců z oboru lékařství (55% souhlasím ku 23% nesouhlasím, n=111)). Postoje vědců jsou v této otázce i s přihlédnutím na předchozí otázku konzistentní. Nicméně vypočtená hodnota chí-kvadrát testu $p=0,13$ nepotvrzuje závislost příslušnosti k oboru na ochotu svá data sdílet.



Obr. 6. Ochota sdílet data

V dotazníku byl položen dotaz přímo na to, zda se vědci domnívají, že mohou při sdílení dat vyvstat nějaké etické problémy (Obr. 7). S tím, že se nějaké etické problémy objevit mohou, souhlasila vždy větší část akademiků ze všech čtyř zastoupených vědních oblastí. Na rozdíl od jiných vědních oblastí se akademici z přírodních věd ve více než pětinovém zastoupení (26%, n=247) domnívají, že etické problémy vyvstat nemohou, což je výrazný rozdíl od názorů akademiků z jiných vědních oblastí. Naopak v oblasti lékařství (72%, n=111) a v sociálních vědách (69%, n=140) se výrazná většina akademiků domnívá, že se problémy objevit mohou.

Vypočtená hodnota chí-kvadrát testu $p=0,004$ potvrzuje závislost názorů v oblasti etických problémů na vědním oboru.



Obr. 7. Etické problémy při sdílení dat

Diskuze

Závěry výzkumu přinesly první vhled do problematiky výzkumných data na UK a lze říci, že se příliš neliší od podobných výzkumu v zahraničí (např. Tenopir 2015, Enwald 2017). Akademičtí pracovníci UK jsou ochotni svá data sdílet a většina jich tak v současnosti už činí. Nejedná se však o otevřený přístup, což může být ovlivněno několika faktory. V první řadě, méně než polovina akademiků je obeznána s principy otevřeného přístupu. Dále neexistuje jednoduché řešení pro ukládání dat v podobě otevřeného univerzitního repozitáře. Některé obory mají svůj vlastní oborový repozitář (LINDAT/CLARIN, Český sociálně vědní datový archiv), ostatní jsou odkázáni na mezinárodní služby typu Zenodo či Dryad. Vzhledem k tomu, že většina respondentů neabsolvovala žádné školení v managementu dat, je pro ně pochopitelně obtížně v této oblasti se orientovat, vybrat správný repozitář a pracovat s ním. V komentářích k výzkumu také zaznělo, že ne ve všech oborech má otevřený přístup, či vůbec sdílení dat smysl.

Ochota sdílet data otevřeně je také ovlivněna obavami z možných problémů, zejména právních a etických, ale i obava ze zneužití či chybné interpretace dat. Zde se projevují výraznější mezioborové rozdíly, vyplývající patrně z rozdílné povahy dat v jednotlivých oborech a ukazující na nutnost přizpůsobovat jakákoliv řešení pro správu dat, školení a jiné

aktivity v této oblasti na míru jednotlivým vědním oblastem. Data ukazují rozdíly v přístupu k výzkumným datům mezi humanitními obory a přírodními obory na jedné straně a lékařskými a sociálními vědami na straně druhé. Z komentářů k dotazníkům vyplývá, že se respondenti také obávají nárůstu administrativy spojené s managementem dat, což byla možnost, která nebyla v dotazníku nabídnuta. Někteří zmiňují technokratizaci bádání, obstrukci duševní činnosti či dokonce nebezpečnost myšlenky otevřeného sdílení dat, v komentářích se objevily i negativní zkušenosti se zneužitím či přímo krádeží dat. V otázce sdílení je pro respondenty důležité, zda se jedná o sdílení dat před publikováním výsledků standardní cestou, či po něm, což nebylo v dotazníku rozlišeno. Některé komentáře také poukazují na obtížnost vytvoření centralizovaného řešení vzhledem k mezioborovým rozdílům, čímž potvrzují výsledky dotazníku, nicméně objevily se i komentáře volající po centrálním repozitáři.

Závěry

Důležitým zjištěním výzkumu je, že akademici a vědci UK jsou ochotni sdílet svá výzkumná data, vnímají však různá rizika, která jsou se sdílením a zejména s otevřeným přístupem k datům spojená, a s managementem dat spojují i další navýšení administrativní zátěže. Pro podporu otevřeného přístupu k datům na UK bude tedy potřeba nejen vytvořit bezpečnou infrastrukturu pro sdílení dat, která by vyhovovala specifikům jednotlivých oborů, ale i zabezpečit podporu managementu dat např. v akademických knihovnách, aby vědci nebyli zatíženi dalšími administrativními povinnostmi.

Dále je zřejmé, že pro hlubší pochopení některých aspektů sdílení dat, bude nezbytné pokračovat ve výzkumu s použitím dalších, kvantitativních i kvalitativních metod, a zaměřit se zejména na specifika jednotlivých vědních oborů.

Použité zdroje

CALZADA PRADO, Javier a Miguel Ángel MARZAL, 2013. Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. *Libri* [online]. **63**(2), 123-134 [cit. 2017-04-09]. 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>

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

DROBÍKOVÁ, Barbora, Adéla JAROLÍMKOVÁ a Martin SOUČEK, 2017. Data literacy of Charles University PhD students : are they prepared for their research careers? In: ŠPIRANEC, Sonja, Serap KURBANOGU, Joumana BOUSTANY, Esther GRASSIAN, Diane MIZRACHI, Lorie ROY a Denis KOS. *The Fifth European Conference on Information Literacy (ECIL): September 18th - 21st, 2017, Saint - Malo, France: abstracts* [online]. Saint - Malo: Information Literacy Association, 2017, s. 42 [cit. 2017-11-28]. ISBN 978-2-9561952-0-7. Dostupné z: <http://ecil2017.ilconf.org/wp-content/uploads/sites/6/2017/09/ECIL-2017-Book-of-abstracts.pdf>

ENWALD, Heike, Terttu KORTELEINEN a Maija-Leena HUOTARI. Research data management: experiences of scholars in Finland. In: ŠPIRANEC, Sonja, Serap KURBANOGLU, Joumana BOUSTANY, Esther GRASSIAN, Diane MIZRACHI, Lorie ROY a Denis KOS. *The Fifth European Conference on Information Literacy (ECIL): September 18th - 21st, 2017, Saint - Malo, France: abstracts* [online]. Saint - Malo: Information Literacy Association, 2017, p. 46 [cit. 2017-11-28]. ISBN 978-2-9561952-0-7. Available from: <http://ecil2017.ilconf.org/wp-content/uploads/sites/6/2017/09/ECIL-2017-Book-of-abstracts.pdf>

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

CHOWDHURY, G., G. WALTON, S. KURBANOGLU, Y. UNAL a J. BOUSTANY, 2016. Information Practices for Sustainability: Information, Data and Environmental Literacy. In: SPIRANEC, S., S. KURBANOGLU a H. LANDOVA. *The Fourth European Conference on Information Literacy (ECIL)*. Prague: Association of Libraries of Czech Universities, s. 22. ISBN 978-80-270-0530-7.

JAROLÍMKOVÁ, Adéla, Barbora DROBÍKOVÁ a Martin SOUČEK, 2017. Výzkumná data na Univerzitě Karlově. In: *INFORUM 2017: 23. ročník konference o profesionálních informačních zdrojích, Praha 30.-31. května 2017* [online]. Praha: Albertina icome Praha [cit. 2017-09-23]. ISSN 1801–2213. Dostupné z: <http://www.inforum.cz/pdf/2017/jarolimkova-adela.pdf>

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

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

TENOPIR, Carol, Elisabeth D. DALTON, Suzie ALLARD a Mike FRAME, 2015. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *PLoS One* [online]. [cit. 2017-02-03]. DOI: 10.1371/journal.pone.0134826. Dostupné z: <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0134826>