

Přidružené publikace v České republice / Enhanced publications in the Czech Republic

PhDr. Petra Pejšová, Mgr. Hana Vyčítalová

petra.pejsova@techlib.cz, hana.vycitalova@techlib.cz

Národní technická knihovna v Praze, Technická 6, 160 80 Praha 6

Abstrakt

Cílem článku je představit termín „přidružené publikace“ (anglicky enhanced publications) českému publiku. Přidružené publikace jsou novým typem vědeckého publikování. Jedná se o zvláštní dodatečné materiály propojené s elektronickými publikacemi, jako jsou výzkumná data, modely, algoritmy, obrázky, metadata, post-publikovaná data (např. komentáře) a další. Součástí článku je prezentace průzkumu stavu přidružených publikací v českých výzkumných institucích.

Klíčová slova

Přidružené publikace, výzkumná data, výzkumné publikace, výzkumné instituce

Resumé v anglickém jazyce

The aim of this article is to introduce the notion of enhanced publications to the Czech audience. Enhanced publications are a new type of scholarly publishing. This concerns extra materials linked with electronic publications, such as research data, models, algorithms, figures, metadata, post-publication data (e.g. commentaries) and other items. The article includes a presentation of a survey of the state of enhanced publications in Czech research institutions.

Klíčová slova v anglickém jazyce

Enhanced publications, research data, research publications, research institutions

Pojem Enhanced publications

Přidružené publikace jsou známější i u nás pod anglickým termínem „enhanced publications“. Dle definice z projektu DRIVER-II se jedná o publikace, které jsou rozšířeny o tři kategorie informací - výzkumná data, zvláštní či dodatečné materiály a post-publikovaná data. (DRIVER, 2009) Výzkumná data jsou například záznamy z měření, výsledky pokusů apod. Pod dodatečnými materiály si můžeme představit modely, algoritmy, obrázky, metadata apod. Mezi post-publikovaná data se řadí komentáře, hodnocení a další typy informací, které vznikají až následně.

Jedná se o další stupeň vědeckého publikování, jehož cílem není pouze publikace jako text, ale zároveň by dílo mělo obsahovat i takové prvky jako podkladová data modelů, algoritmů apod. Přidání těchto základních údajů z výzkumu k publikacím usnadní ověřit, reprodukovat, a znovu použít výsledky výzkumu. Síla přidružených publikací je v popisu vztahů mezi podkladovými daty a výstupy z výzkumu smysluplným způsobem, jak v čitelné, tak ve strojově zpracovatelné formě. (SURF, 2014b) Možnosti odkazování umožňují výzkumníkům propojit libovolný typ objektu s jiným objektem. Je tak snadné propojit článek s přednáškami, recenzí či rozhovorem, které lze nalézt na internetu na různých místech. Tak mohou být popsány vztahy mezi jednotlivými typy informací přehledně na jediném místě. To je to, co nazýváme přidruženými publikacemi.

Evropská síť spolupracujících vědeckých repozitářů DRIVER byla jednou z prvních, která se začala zabývat přidruženými publikacemi. Posláním DRIVER je optimalizovat způsob elektronické infrastruktury k ukládání znalostí tak, aby byla přidána hodnota primárním výzkumným datům a informace ze sekundárního výzkumu byly efektivnější. (DRIVER 2011a) Tyto znalosti mají cenný přínos pro průmysl a pomáhají ve spolupráci mezi výzkumem a vzděláváním. Díky své síti volně přístupných digitálních repozitářů poskytuje nyní DRIVER přístup k více jak 3,5 miliónům vědeckých publikací, článků, disertací, přednášek, zpráv a dalších typů dokumentů ze všech vědních disciplín. V síti DRIVER je zapojeno 295 digitálních repozitářů z 38 zemí světa.

Rozvoj DRIVER probíhá díky projektu DRIVER II, jedním z jeho hlavních cílů je širší pokrytí pomocí přidružených publikací. V rámci projektu DRIVER II vznikla studie „Enhanced Publications: Object Models and Functionalities“, která definuje požadavky na ukládání a správu přidružených publikací v rámci infrastruktury DRIVER. Tato zpráva je volně ke stažení na http://www.driver-repository.eu/component/option,com_jdownloads/Itemid,83/task,view.download/cid,54/.

Podporovat koncepci otevřených přidružených publikací je též jedním z cílů portálu OpenAIRE¹. „Hlavním cílem projektu OpenAIREplus je vytvoření robustní, participativní služby pro vzájemné propojování recenzovaných vědeckých publikací a souvisejících datových souborů. Vzhledem k tomu, že se vědecká komunikace dotýká celé řady disciplín,

¹ <http://www.openaire.eu/en/component/content/article/326-openaireplus-press-release>

by měl projekt díky svému horizontálnímu dosahu napomoci spolupráci napříč datovými infrastrukturami, přičemž informace by měly sloužit uživatelům z okruhu vědeckých i nevědeckých zájemců a také poskytovatelům služeb s přidanou hodnotou. Součástí projektu bude vytvoření e-infrastruktury, která bude získávat, doplňovat a ukládat metadata z vědeckých souborů dat dostupných v režimu otevřeného přístupu. Pro účely podpory správy a vzájemného propojování souvisejících vědeckých dat bude využita podpůrná technická struktura inovativního charakteru.

Pro přístup k propojeným publikacím a jejich ukládání prostřednictvím portálu OpenAIRE bude zajištěna podpora formou helpdesku a bude rozšířena struktura pro podporu spolupráce v rámci sítě OpenAIRE za účelem propagace konceptu otevřených přidružených publikací v rámci uživatelských komunit. V každé z 31 evropských zemí zapojených do projektu je k dispozici styčný pracovník, který bude pomáhat s naplňováním potřeb evropských výzkumníků. V projektu budou také aktivně využity mezinárodní kontakty jako způsob jak posílit společné standardy, datovou problematiku a interoperabilitu na globální úrovni.“ (Paving the way, 2011b)

Významným iniciátorem myšlenky zpřístupňování přidružených publikací je holandské sdružení SURF², které spojuje holandské výzkumné univerzity, univerzity aplikovaných věd a výzkumné instituce. Tyto instituce spolupracují na inovativních projektech s cílem zlepšit kvalitu vysokoškolského vzdělávání a výzkumu v oblasti informačních a komunikačních technologií. V roce 2011 bylo v rámci programu SURFshare financováno šest projektů, které umožnily výzkumníkům získat zkušenosti s vytvářením přidružených publikací. Tyto projekty jsou popsány v krátkém filmu, kde výzkumníci popisují, jak propojení výzkumných výstupů poskytuje větší přehled, vyšší kvalitu a více příležitostí rozvíjet akademický výzkum. Video je dostupné na Youtube v kanále SichtingSURF pod názvem „Enhanced Publications: from experiment to practice“³. V kanále SichtingSURF je mnoho dalších videí o přidružených publikacích.

Díky sdružení SURF vznikly různé podpůrné materiály, které jsou všechny v anglickém jazyce. Jedním z nich je kniha Book Enhanced Publications: Linking Publications and Research Data in Digital Repositories⁴, která poskytuje aktuální přehled základních prvků přidružených publikací. Hlavním zjištěním je, že vydavatelé a provozovatelé digitálních repozitářů mají nástroje pro vytváření všech třech typů přidružených publikací, ale nevyužívají je. Vydavatelé a provozovatelé digitálních repozitářů by měli nabídnout služby a nástroje k přidání výzkumných dat, dodatečných materiálů a post-publikovaných dat k publikacím. Výzkumní pracovníci by měli být zodpovědní za obsah. Další zajímavou knihou je Emerging Standards for Enhanced Publications and Repository Technology⁵, která se věnuje dvěma oblastem novým technologiím a komunitám a interoperabilitě. V části věnované novým technologiím a

² <http://www.surf.nl/en/themes/research/research-data-management/enhanced-publications/index.html>

³ <http://www.youtube.com/watch?v=FHI4J94YUk&list=PLA936B4E394A29093&index=16>

⁴ <http://dare.uva.nl/document/150723>

⁵ <http://dare.uva.nl/aup/nl/record/316870>

komunitám jsou popsány sítě, dlouhodobé uchovávání dat, strategie a projekty zabývající se tématem přidružených publikací a také evropské společenství CRIS (Computer Research Information Systems). V druhé části se autoři zabývají novými standardy, formáty a vývojem digitálních úložišť, neboť to vše souvisí s interoperabilitou přidružených publikací. Posledním zde uvedeným zajímavým výstupem je vizualizér přidružených publikací, díky kterému jsou objekty v přidružených publikacích dostupné v pěkném intuitivním přátelském rozhraní.

Vizualizér je volně ke stažení na tomto linku

<http://wiki.surf.nl/display/vp/4.3+'InContext'+Visualiser> .

Dalším významným příkladem z Nizozemí je systém DANS⁶ (Data Archiving and Networking Services), který zajišťuje digitální repozitář specializující se na dlouhodobou archivaci výzkumných dat. DANS je provozovaný v systému EASY⁷ a nabízí možnost on-line ukládání a přístupu k tisícům souborů dat z humanitních a sociálních věd a dalších disciplín. (DANS, 2014a) Jedná se především o vědecké soubory dat, e-publikace a další výzkumné informace. DANS je zdarma pro stahování i ukládání souborů, ale s nutnou registrací. DANS ukládá data trvale a udržitelným způsobem a umožňuje jejich znovuvyužití. Soubory jsou zpřístupňovány jiným výzkumníkům za podmínek definovaných vkladatelem. Uživatel se zavazuje citovat zdroj dat ve svých publikacích a k propojení publikace na data v DANS. Kromě toho je v rámci DANS zajišťováno také školení a poradenství a prováděn výzkum trvalého přístupu k informacím. DANS je provozován společně Královskou nizozemskou akademií věd a umění (KNAW) a Nizozemskou organizací pro vědecký výzkum (NWO). Jako součást svého poslání podporuje DANS otevřený přístup, i když si je vědom, že ne všechna výzkumná data mohou být volně dostupná a bez omezení. I v případě, že data nejsou volně k dispozici nebo jen v omezené míře, je nutné, aby výzkumná data byla archivována udržitelným způsobem. Aby bylo zajištěno, že archivovaná data bude možné stále nalézt, zpřístupnit a znovu využít v budoucnu, byla v rámci DANS vyvinuta pečeť Data Seal of Approval⁸. Tato pečeť zaručuje, že datové úložiště splňuje celou řadu jasně stanovených kritérií v oblasti kvality, ochrany a dostupnosti dat. DANS také financuje malé datové projekty, jako jsou studie proveditelnosti a pilotní řešení pro archivaci dat v nových vědních disciplínách. Data uložená v DANS jsou přístupná také přes informační bránu k vědeckým informacím v Nizozemí NARCIS⁹. NARCIS poskytuje informace o výzkumu, programech, projektech, výzkumných pracovnících, výzkumných institucích a jejich profilech včetně datových souborů a přidružených publikací.

Evropská unie se ve svých politikách nevěnuje přímo přidruženým publikacím, ale vztahuje se na ně politika otevřeného přístupu ve vědeckém publikování, která má zajistit rozvoj vzdělávání, podpořit inovace a kreativní znovu využití výsledků evropského výzkumu.

V rámcovém programu EU pro výzkum a inovace Horizont 2020 je definována povinnost uložení výzkumných dat potřebných pro ověření výsledků prezentovaných ve vědeckých

⁶ <http://www.dans.knaw.nl/en/content/data-archive/depositing-data>

⁷ <https://easy.dans.knaw.nl/ui/home>

⁸ <http://www.datasealofapproval.org/en/>

⁹ <http://www.narcis.nl/>

publikacích, které jsou výsledkem projektů, v režimu otevřeného přístupu viz 29.2 (a) příručky Multi-beneficiary General Model Grant Agreement: Version 1.0¹⁰. Dle bodu 29.3 je uložena povinnost ukládat výzkumná data v režimu otevřeného přístupu pouze příjemcům podpory v rámci aktivit „Research Data Pilot: Regarding the digital research data generated in the action“.

Průzkum přidružených publikací v ČR

V České republice nebyl zatím stav přidružených publikací zmapován. Proto Národní technická knihovna provedla v letošním roce průzkum. Průzkum měl za cíl zjistit, jaká data z výzkumu výzkumné instituce produkují a archivují. Dále nás zajímal i způsob ukládání a archivace dat, další využívání dat a to nejdůležitější – zda jsou data propojená s publikacemi z výzkumu a můžeme tedy hovořit o přidružených publikacích v ČR.

Cílovou skupinou průzkumu byly veřejné výzkumné instituce, soukromé výzkumné instituce a další instituce, které se výzkumem zabývají. Osloveni byli zástupci těchto institucí, u kterých se předpokládalo, že budou mít kompletní přehled o výzkumné činnosti dané instituce, vybráni byli lidé na pozici vědeckých tajemníků, vědeckých sekretářů, náměstků pro výzkum nebo vědu. Pokud nebylo možné zjistit, kdo podobné pozice v instituci zastává, případně takové funkce nebyly v instituci zřízeny, byli vybráni zástupci z vedení instituce.

Průzkum byl proveden pomocí online dotazníku, složeného z devatenácti otázek. Jednalo se o strukturovaný dotazník, většina otázek byla uzavřená s možností jedné nebo více odpovědí. Některé otázky navazovaly na předchozí v závislosti na výběru odpovědi na předchozí otázku. K šíření dotazníku byla vybrána jedna z online aplikací na webu.

Vybraným respondentům byl zaslán informativní email s prosbou o vyplnění dotazníku. Email obsahoval informace o termínu přidružené publikace, cíli průzkumu i o využití získaných dat. Připojena byla také prosba, aby v případě, že adresát nemůže průzkum vyplnit, předal informaci a odkaz na dotazník někomu ze svých kolegů.

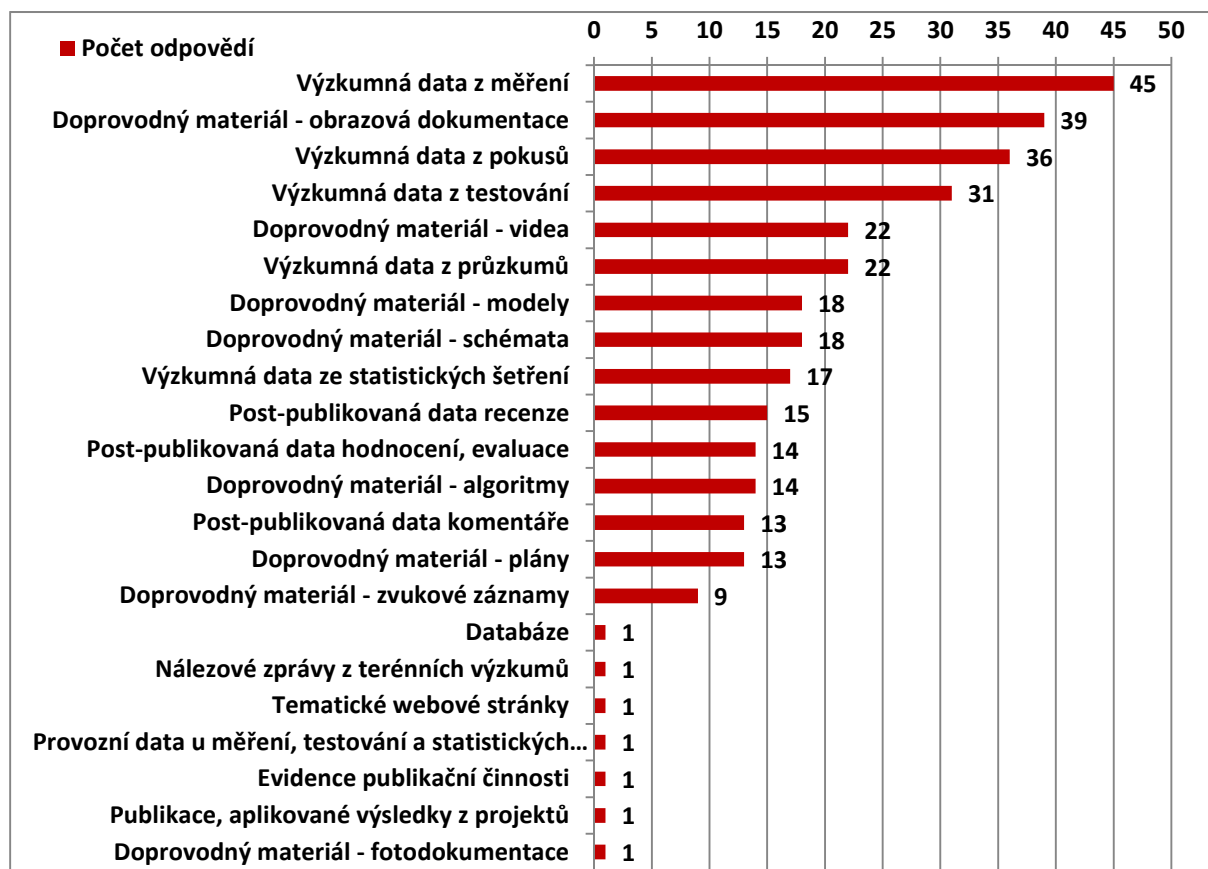
Průzkum probíhal od 11. září do 7. října 2013. Ze začátku jsme obdrželi pouze malý počet vyplněných dotazníků, proto jsme respondenty žádali znovu o vyplnění telefonicky. Celkem jsme oslovili 113 institucí, potažmo jejich zástupců, získali jsme 66 vyplněných dotazníků. Čtyři instituce odmítli nebo nemohli dotazník vyplnit. Odmítnutí bylo z důvodu utajování informací z výzkumu, zástupce dané instituce nechtěl poskytnout žádné informace. Dalším uvedeným důvodem bylo, že instituce nemají výzkumná data, souhrn poznatků z výzkumu rovnou formulují do publikovaných výstupů. Jedna instituce uvedla, že se nezabývá přímo výzkumnou činností, spíše přípravou podkladů pro výzkum. Detailní souhrn oslovených institucí a získaných odpovědí je v následující tabulce a grafu.

¹⁰ http://ec.europa.eu/research/participants/data/ref/h2020/mga/gga/h2020-mga-gga-multi_en.pdf

Typ instituce	Ústavy AV	VVI	Soukromé výzkumné instituce	Jiné výzkumné instituce	Celkem
Počet oslovení	54	21	34	4	113
Počet responsí	33	14	16	2	65
Negativní odpověď	2	0	1	1	4

Tab. č. 1 Počet odpovědí

Nejprve bylo třeba získat přehled, jaká data z výzkumu ve výzkumných institucích vznikají. V dotazníku byly vyjmenovány jednotlivé typy dat, které obvykle v rámci přidružených publikací mohou být připojeny k publikacím z výzkumu. Navíc mohli respondenti uvést další možnosti. Obvykle respondenti vybrali více možností. 48% výsledků tvořily různé typy výzkumných dat (z měření, testování, pokusů...), 42% tvořil doprovodný materiál (obrazová dokumentace, videa, modely, schémata...) a 10% tvořila post-publikovaná data (recenze, hodnocení).

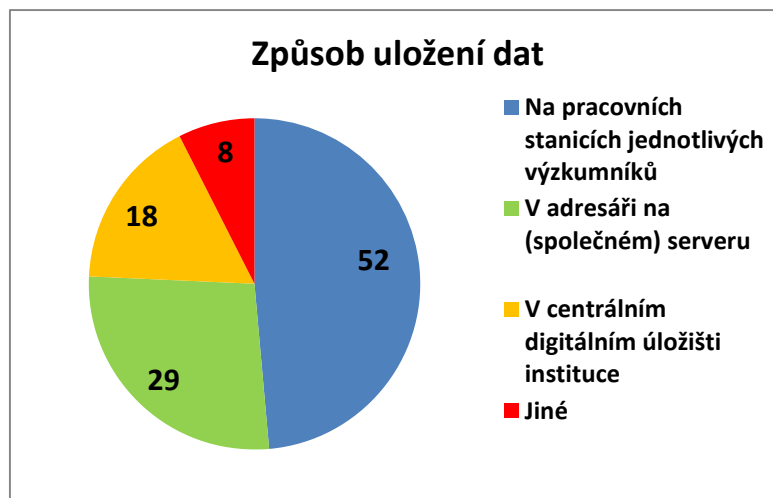


Obr. č. 1 Typy výzkumných dat

Nejčastěji jsou data ukládaná a archivovaná ve formátu PDF, DOC a XLS: Další formáty dat nebyly uváděny s takovou četností. Ke známějším formátům patřil také formát XML, CSV nebo obrázkové formáty (JPEG, TIFF aj.). Méně známé byly formáty SAV, CIF a různé formáty pro oblast kartografie.

Zajímal nás také celkový objem dat, žádali jsme respondenty o alespoň hrubý odhad, odpověď nebyla povinná. I tak se jednalo zřejmě o obtížně zodpověditelnou otázku, kterou zodpověděl jen malý počet respondentů. Celkový objem výzkumných dat patrně souvisí i s oborovým zaměřením instituce. Obvyklá odpověď byla v řádu desítek nebo stovek gigabytů, několik odpovědí sahalo do řádu terrabytů.

Dalším podstatným zjištěním mělo být, kde a jak jsou výzkumná data ukládána a archivována. Dotazník respondentům opět nabídl několik možností, mohli zvolit jednu, více nebo doplnit svoji. Potvrdila se nám hypotéza, že nejčastěji nejsou data ukládána centrálně na společném úložišti, ale zůstávají na pracovních stanicích (počítačích) jednotlivých pracovníků, tato možnost byla zvolena 52x. Dalšími častými způsoby uložení byly společný adresář na disku (29 odpovědí) a centrální úložiště dat instituce, tuto odpověď vybralo 18 respondentů. Někteří respondenti uvedli ale např. dvě nebo tři tyto vyjmenované možnosti, koncepce ukládání dat není tedy evidentně ve spoustě institucí dořešená a jednotná. Většina institucí ale nehodlá zatím nic měnit, na to se ptala následující otázka.

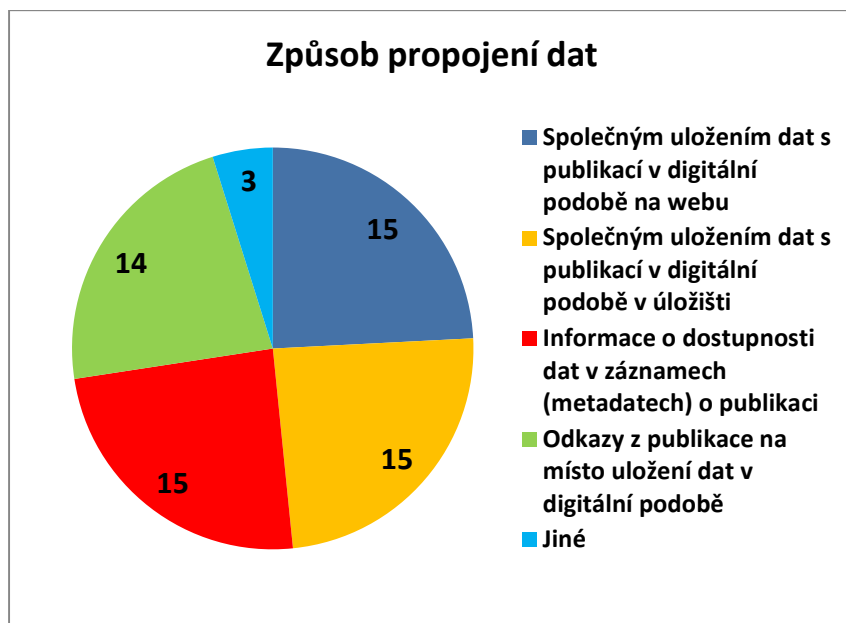


Obr. č. 2 Způsob uložení dat

Zajímavé zjištění poskytla také otázka, kterou jsme se ptali, zda instituce data archivují více jak deset let. 54% respondentů uvedlo, že ano, 31 % nevědělo, ale 15% (10 respondentů) odpovědělo, že se data tak dlouhou dobu nearchivují, to je poměrně velké množství. Přitom ale také většina respondentů uvedla, že data jsou znovu využívána při dalších výzkumech.

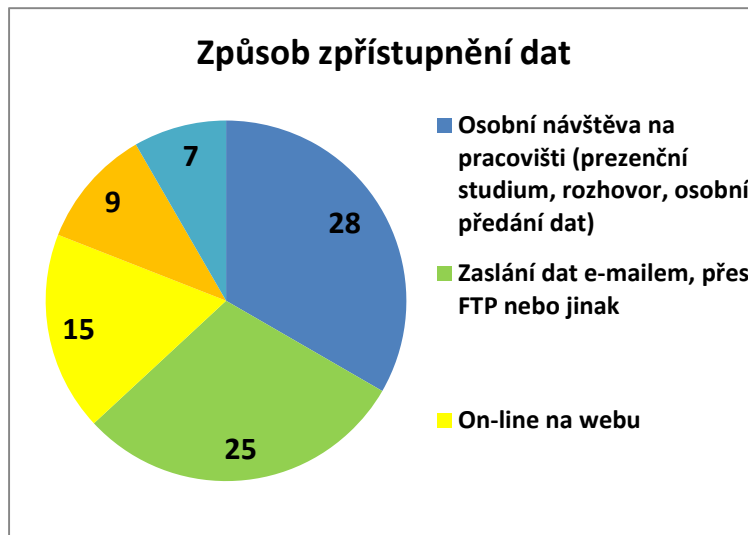
O data se zpravidla starají sami výzkumníci, uvedlo to tak 53 respondentů. I u této otázky mohli ale respondenti vybrat více možností. Dále se o data často starají pověřené osoby v knihovně, archivu, IT oddělení nebo vědeckí sekretáři nebo tajemníci, případně vedoucí.

Jádro průzkumu spočívalo ve zjištění, zda v českém prostředí lze hovořit o přidružených publikacích, čili zda jsou výzkumné publikace propojovány s daty z výzkumu. V dotazníku bylo nastíněno několik možností, jak takové propojení může v praxi fungovat. Kromě ideálního způsobu propojení – společné uložení publikace v elektronické podobě spolu s příslušnými výzkumnými daty v digitálním úložišti - bylo navrženo několik dalších alternativních možností. Poměrně velkým překvapením bylo, že 58% respondentů uvedlo, že nějakým způsobem výzkumné publikace s daty propojují. 5 zástupců institucí, ve kterých zatím k propojování publikací a dat nedochází, uvedlo, že by situaci chtěli změnit.



Obr. č. 3 Způsob propojení dat s publikacemi

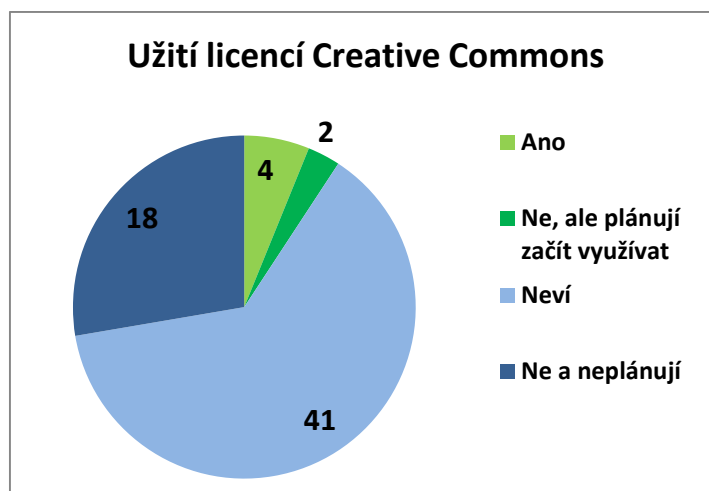
Dalším faktem, který nás zajímal, bylo, zda a jakým způsobem jsou instituce ochotné poskytnout data z výzkumu kolegům z jiných institucí. Čtyřicet čtyři respondentů uvedlo, že nějakým způsobem je možné jejich data poskytnout. Obvykle musí ale zájemci navštívit pracoviště a osobně si data převzít nebo vyžádat. Takovým způsobem je ochotno data poskytnout 68% respondentů (44 institucí), z toho 37 veřejných výzkumných institucí a 7 soukromých výzkumných institucí. Data není ochotno poskytnout 12 veřejných výzkumných institucí a 9 soukromých výzkumných institucí. Poskytnout data ideálním způsobem (on-line na webu nebo on-line v digitálním úložišti) může 24 institucí. Někdy musí instituce splnit určité podmínky, např. získat souhlas autora nebo majitele výzkumných dat.



Obr. č. 4 způsob zpřístupnění dat

Pokud instituce nemohou nebo nechtějí data zpřístupnit jiným zájemcům, byly požádány o zdůvodnění. Jasně je stanovisko v otázce komerčních výzkumu – výzkumů na zakázku. Pak jsou data majetkem zadavatele, který výzkum zaplatil. Zadavatel by tedy musel souhlasit se zpřístupněním. Poměrně často uváděným důvodem bylo riziko zneužití dat. V některých institucích jsou vlastní data považována za duševní vlastnictví jejich autora (výzkumníka) a ten by se zpřístupněním musel souhlasit. Jako veřejná jsou také často považována jen ta data, která jsou součástí konečného výstupu z výzkumu, tzn., která byla publikována v odborných publikacích nebo na konferencích. Vyskytla se i otázka patentové ochrany.

Poslední oblastí, kterou průzkum prověřoval, byla znalost a použití veřejných licencí (konkrétně Creative Commons). Díky použití veřejných licencí může autor poskytnout své dílo nebo data veřejnosti za určitých podmínek. Za stejných podmínek, jaké určí autor pomocí zvolené licence, může další zájemce dílo užívat nebo šířit. Pouze šest procent (4 respondenti) uvedlo, že pomocí licencí označují a tím pádem zpřístupňují data z výzkumu. Další dva respondenti potvrdili, že by chtěla začít veřejné licence používat.



Závěr

Ze získaných informací lze usoudit, že v českých výzkumných institucích není častější existence přidružených publikací vyloučena. Zatím jsou opravdové přidružené publikace, tzn., výzkumné publikace propojené s výzkumnými daty dostupné nejlépe on-line, teprve v začátcích. Bylo by třeba vylepšit způsob ukládání a archivace výzkumných dat na českých výzkumných institucích. Vhodnější by byl ujasněný koncept s centrálním úložištěm pro celou instituci a také větší technická podpora výzkumníkům.

Lepší přístup k technické otázce ukládání dat by usnadnil také jejich postupné zpřístupňování online. Samozřejmě zde vždy bude otázka autorskoprávní a jednotlivé instituce budou muset vždy zvážit, v jakém režimu a za jakých podmínek by mohla být jejich data dostupná. Byla by také lepší širší osvěta v oblasti autorskoprávní problematiky a otázek veřejných licencí, aby jednotlivci i instituce neměly obavu ze zneužití dat.

Použitá literatura

DANS. 2014a. About DANS. *DANS: Data Archiving and Networked Services* [online]. DANS [cit. 2014-01-02]. Dostupné z: <http://www.dans.knaw.nl/en/content/about-dans>.

DRIVER. 2011a. DRIVER: Networking European Scientific Repositories. *DRIVER – Digital Repository Infrastructure Vision for European Research* [online]. 02 June 2011 [cit. 2014-01-02]. Dostupné z: <http://www.driver-repository.eu/>.

DRIVER. 2009. Enhanced Publications. *DRIVER – Digital Repository Infrastructure Vision for European Research* [online]. 18 February 2009 [cit. 2014-01-02]. Dostupné z: <http://www.driver-repository.eu/Enhanced-Publications.html>.

SURF. 2014b. Enhanced publications. *SURF: Collaborative organisation for ICT in Dutch higher education and research* [online]. SURF, 26 Nov 2013 [cit. 2014-01-02]. Dostupné z: <http://www.surf.nl/en/themes/research/research-data-management/enhanced-publications/index.html>.

Paving the way to an open scientific information space: OpenAIREplus – linking peer-reviewed literature to associated data. 2011b. In: *OpenAIRE: Open Access Infrastructure for Research in Europe* [online]. 15 December 2011 [cit. 2014-01-02]. Dostupné z: <http://www.openaire.eu/en/component/content/article/326-openaireplus-press-release>.