

# ÚVT MU zprava o daj

Bulletin pro zájemce o výpočetní techniku na Masarykově univerzitě • prosinec 2009 • roč. XX • č. 2

## Open access – otevřený přístup k vědeckým informacím.

### Úvod do problematiky

*Miroslav Bartošek, ÚVT MU*

Stále větší množství akademických a vědeckovýzkumných institucí po celém světě zveřejňuje intelektuální produkci svých členů v režimu *otevřeného přístupu* (Open Access – OA), čili volně a bezplatně komukoliv na internetu. Důvody pro to jsou různé. Na jedné straně stojí pohnutky odvolávající se na sociálně zodpovědné chování institucí. Slušná akademická instituce usiluje o to, aby plody ducha jejích členů byly dostupné v co nejširší možné míře a bez zbytečných bariér každému zájemci; notabene jde-li o instituci placenou z veřejných prostředků. Dalším z hnacích motorů je snaha podpořit úsilí o nápravu současného nevyhovujícího modelu vědecké komunikace trpícího řadou neduhů. Mezi ty nejpalčivější, ne však jediné, patří dlouhé lhůty při publikování v prestižních tištěných časopisech a dramaticky se zvyšující počty a ceny vědeckých časopisů, zejména pak u velkých vydavatelů, kteří ve značné míře monopolizovali trh vědeckých publikací. Důležité jsou však i pragmatické důvody spočívající v širším zviditelnění instituce ve vědeckém světě a v lepším citačním ohlasu odborných prací členů takto „osvícené“ instituce.

Přechod do otevřeného přístupu v institucionálním měřítku není však zrovna snadná záleži-

tost. Tou relativně nejméně problémovou částí je vlastní technické řešení a technologická podpora. Podstatně složitější je řešení politických otázek, autorsko-právních aspektů, financování a organizace celého procesu tak, aby znamenal skutečný přínos jak pro mezinárodní odbornou komunitu, tak i pro jednotlivé členy příslušné instituce. Mezi klíčové „politické“ otázky patří rozhodnutí, zda vůbec, jakou formou a v jakém rozsahu se má instituce do OA zapojit, a následně jaká má být míra povinnosti či dobrovolnosti jejích členů předávat své práce do otevřeného přístupu. Pod autorsko-právní problematiku zase spadají nelehké otázky typu, v jaké formě a za jakých podmínek je autor oprávněn nabídnout do OA svůj článek publikovaný v recenzovaném (komerčním) časopise. Organizační aspekty zahrnují rozdělení rolí a zodpovědností řešitelů, stanovení postupů uvnitř instituce, podporu akademikům při zveřejňování jejich prací v OA a další. Je zřejmé, že celkově jde o natolik významnou změnu v zavedených výzkumných a publikačních zvyklostech dané instituce, že se nedá provést bez silné podpory od vedení instituce i od významné části její akademické obce.

Akademické instituce v ČR se zatím do tohoto procesu výrazněji nezapojily. Jedním z důvodů může být i nižší míra povědomí o iniciativě otevřeného přístupu – jejím smyslu, cílech, postupech a přínosech. Článek si proto klade za cíl přispět k osvětě v této oblasti a napomoci dis-

kusi o možnostech širšího zapojení české akademické komunity. Mimo jiné i z toho důvodu, že zpřístupnění výsledků vědeckého výzkumu se stále častěji stává závaznou podmínkou poskytovatelů výzkumných grantů jak na národní tak i celoevropské úrovni.

## 1 Východiska: BBB-iniciativy

Cílem hnutí otevřeného přístupu je využít možností globální elektronické komunikace (internetu) k prosazení nové, efektivní, rychlé a bezplatné výměny vědeckých poznatků, a tím vytvořit alternativu k tradičnímu modelu vědeckého publikování. Přestože principy OA jsou rozvíjeny již delší dobu, formální vymezení a východiska poskytly především tři události z počátku tohoto století označované někdy souhrnně jako tzv. *BBB-iniciativy*: Budapešťská iniciativa, Prohlášení z Bethesdy a Berlínská deklarace.

*Budapešťská iniciativa* (Budapest Open Access Initiative - BOAI) [1] vznikla na setkání členů Open Society Institute (podporovaném Sorosovou nadací) v únoru 2002 v Budapešti. Vymezuje základní principy, zdůvodnění, strategie a postupy uplatňované při prosazování otevřeného přístupu k vědeckým informacím a vyzývá vědeckou komunitu k širokému zapojení.

*Prohlášení z Bethesdy* (Bethesda Statement on Open Access Publishing) [2] vzniklo v dubnu 2003 v rámci biomedicínské výzkumné komunity. Jeho cílem bylo podnítit diskusi, jak postupovat při urychleném prosazování zásad otevřeného přístupu k primární vědecké literatuře; současně vymezilo konkrétní kroky nezbytné pro přechod k OA-publikování pro jednotlivé subjekty ve sféře vědeckého publikování (výzkumné instituce, grantové agentury, knihovny, vydavatele vědeckých časopisů, vědce a učené společnosti).

*Berlínská deklarace* (Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities) [3] vzešla z konference pořádané společností Maxe Plancka v Berlíně v říjnu 2003. Přihlásila se k předchozím dvěma iniciativám a vyzvala vědce a výzkumné instituce ke změně jejich publikačních návyků ve prospěch publikování v režimu OA. K dnešnímu dni podepsalo tuto deklaraci a zavázalo se tak podporovat principy OA

na 300 významných výzkumných institucí z celého světa. Z České republiky tak učinily dosud jen dvě instituce - Akademie věd ČR a Grantová agentura ČR; obě podepsaly deklaraci v roce 2008.

## 2 Co je to Open Access

Existuje vícero definicí pojmu otevřený přístup. Základní vymezení poskytla Budapešťská iniciativa [1]:

*„Literatura, která by měla být volně dostupná online, je ta, kterou vědci poskytují světu, aniž by za ni očekávali platbu. Primárně tato kategorie zahrnuje recenzované časopisecké články; patří sem ale i nerecenzované preprinty, které vědci mohou chtít nabídnout online pro připomínkování nebo jako upozornění kolegům na důležité výzkumné poznatky. Existuje mnoho stupňů a druhů širšího a snazšího přístupu k takové literatuře. Pojmem „otevřený přístup“ k této literatuře myslíme její volnou dostupnost na veřejném internetu umožňující libovolnému uživateli číst, stahovat, kopírovat, distribuovat, tisknout, prohledávat nebo vytvářet odkazy na plné texty těchto článků, sklízet je pro potřeby indexace, předávat je jako data pro software, nebo používat je k jakýmkoliv jiným legálním účelům bez finančních, právních nebo technických omezení s výjimkou těch, která jsou neoddělitelnou součástí získání přístupu k internetu samotnému. Jediným omezením na reprodukci a distribuci a jediným uplatněním autorskoprávní ochrany (copyrightu) v této oblasti by mělo být poskytnout autorům kontrolu nad integritou jejich prací a právo na řádné uznání a uvedení autorství.“*

Z výše uvedeného je zřejmé, že hnutí otevřeného přístupu se zaměřuje především na kvalitní vědecké publikace - recenzované články, preprinty, ale i další materiály - a že v něm nejde pouze o bezplatný přístup, ale současně i o možnost neomezeného následného využití (*free availability and unrestricted use*). V praxi je samozřejmě tento ideál ne vždy dosažitelný, takže je často nutné vycházet z konkrétních aktuálních možností a podmínek (např. licenčních podmínek vydavatelů).

Pojem otevřeného přístupu se neustále vyvíjí. Základní vymezení Budapešťské iniciativy je v prohlášení jiných iniciativ dále rozvíjeno a doplňováno. Například Prohlášení z Bethesdy [2] připojuje následující požadavek:

„Úplná verze práce a všech doplňkových materiálů, včetně souhlasu k využití, jsou ihned po prvotním publikování uloženy ve vhodném standardizovaném elektronickém formátu alespoň v jednom on-line repozitáři, který je podporován akademickou institucí, učenou společností, vládní agenturou nebo jinou dobře zavedenou organizací, která usiluje o poskytnutí otevřeného přístupu, neomezenou distribuci, interoperabilitu a dlouhodobou archivaci (pro oblast biomedicíny je takovým repozitářem PubMed Central).“

Je vhodné zdůraznit, že koncepce otevřeného přístupu neusiluje o potírání, rušení či porušování dosud zavedených institutů jako je autorské právo (copyright), recenzní řízení, výdělečné publikování vědecké literatury aj. Je s nimi kompatibilní, hledá však nové efektivnější cesty vědecké komunikace. Základní rozdíl je ten, že účty nehradí čtenář - a proto také finance neslouží jako překážka v přístupu [4].

### 3 Zlatá a zelená cesta k OA

Existují dvě hlavní cesty k otevřenému publikování, které vytýčila již Budapešťská iniciativa. Je to:

1. publikování v *otevřených časopisech* (tzv. zlatá cesta - gold OA; otevřený přístup k publikacím poskytují vydavatelé);
2. autoarchivace v *otevřených repozitářích* (tzv. zelená cesta - green OA; otevřený přístup k publikovaným pracím poskytují autoři).

Prvotní rozdíl spočívá v tom, že zatímco OA časopisy provádí recenzní řízení, tak OA repozitáře nikoliv. Rozdílů je však více.

#### 3.1 Otevřené časopisy

Otevřené časopisy vychází z tradičního modelu vědeckého publikování prostřednictvím recenzovaných vědeckých časopisů. Umožňují však všem čtenářům okamžitý bezplatný přístup k plným

textům článků. Copyright obvykle zůstává autorům, kteří poskytují vydavatelům časopisů (neexkluzivní) licenci k publikaci svého článku, mohou tedy s článkem dál volně nakládat, například ho vložit do některého z otevřených repozitářů. Z hlediska krytí nákladů na publikování a tvorby zisku lze rozdělit OA-časopisy do dvou krajních skupin (mezi nimi existuje velká škála nuancí<sup>1</sup>):

**Nekomerční časopisy:** bezplatné pro čtenáře i autory; náklady na publikování, recenzní řízení, zpřístupnění, archivaci a další výdaje hradí třetí strana (např. vědecká instituce).

**Komerční časopisy:** bezplatné pro čtenáře; náklady hradí autoři (nebo jejich sponzoři - zaměstnavatelé, grantové agentury aj.), a to platbami za publikování článku. Cena se dnes pohybuje v rozpětí 700 - 3 500 USD za článek podle výše nákladů a prestiže daného časopisu. Některé z komerčních časopisů jsou neziskové (např. PLoS - Public Library of Science), jiné pracují v tradičním výdělečném režimu (např. BioMed Central).

Uvádí se, že v současnosti tvoří otevřené časopisy zhruba 15% ze všech vydávaných recenzovaných vědeckých časopisů; zhruba jde o 4 400 z celkového počtu 25 000 časopisů. Z nich pak méně než polovina využívá obchodní model „platí autor“ (a často i ten je dále „změkčován“ různými úlevami).

Aktuální přehled kvalitních plně otevřených vědeckých časopisů poskytuje *Adresář otevřených časopisů* <http://www.doaj.org/> (Directory of Open Access Journals - DOAJ). Časopisy jsou členěny podle oborů, a zhruba pro třetinu z nich nabízí adresář i možnost prohledávání na úrovni článků. V době psaní článku obsahoval adresář 4 400 časopisů, z toho 23 časopisů od českých vydavatelů.

---

<sup>1</sup>Z mnoha variant, které se vyskytují mezi těmito dvěma póly, zmiňme alespoň *hybridní OA-časopisy* (obecně jsou dostupné na bázi předplatného, ale některé články z těchto časopisů jsou zařazeny - na základě platby autorů resp. jejich zástupců - do otevřeného přístupu) a *OA-časopisy s pozdrženým přístupem* (umožňují otevřený přístup k článkům až po uplynutí stanovené doby od publikování).

### 3.2 Otevřené repozitáře

Otevřené repozitáře jsou založeny na tom, že autoři samotní poskytují otevřený přístup ke svým pracím tím, že vloží elektronické verze prací do digitálního repozitáře (a to včetně těch prací, které byly publikovány mimo oblast otevřeného přístupu a k nimž mají autoři oprávnění ke zveřejnění). Repozitář tyto práce zpřístupňuje světu a stará se o jejich dlouhodobé uchování. V principu existují tři různá místa, kam mohou autoři své práce vkládat: institucionální repozitáře, předmětové repozitáře, osobní webové stránky.

**Institucionální repozitáře:** řada univerzit a výzkumných institucí provozuje repozitáře, jejichž prostřednictvím shromažďuje, uchovává a distribuuje svou intelektuální produkci (preprinty či postprinty publikovaných recenzovaných článků, technické zprávy, studentské a vědecké kvalifikační práce atd.). Jako příklad uveďme repozitář Kalifornské univerzity *eScholarship Repository* <http://escholarship.org/>.

**Předmětové repozitáře:** jsou zaměřeny na určitou konkrétní vědní oblast. Patrně nejznámějším zástupcem je *arXiv.org* <http://arxiv.org>, pokrývající původně oblast fyziky a matematiky, později i computer science a některých dalších přírodních věd. Vznikl již v roce 1991 s cílem shromažďovat a zpřístupňovat elektronické verze prací (e-prints), zejména preprintů časopiseckých a sborníkových článků. Kromě toho, že nabízí volný přístup k obrovskému množství kvalitních dat (v současnosti více než 570 tisíc e-printů), dramaticky zvyšuje rychlost šíření vědeckých poznatků: okamžitě poté, kdy vědec vloží svůj e-print do repozitáře, je dostupný odborné komunitě na celém světě. Z dalších známých a velkých předmětových repozitářů zmiňme alespoň *PubMed Central* <http://pubmedcentral.nih.gov/> pro oblast biomedicíny a *RePEc - Research Papers in Economics* <http://repec.org/> pro oblast ekonomie.

**Osobní stránky:** někteří vědci vystavují své práce na vlastních osobních webových stránkách. Problémem ale bývá větší pracnost, menší viditelnost, nízká interoperabilita a

krátkodobá životnost takových stránek. Lepším řešením může být vytvoření osobního odborného profilu v rámci některé z trvale udržovaných vědeckých sítí, jako je například ResearchGATE <https://www.researchgate.net/>. Další možností je automatické generování osobních stránek s otevřenými publikacemi jako nadstavba nad institucionálním repozitářem; příkladem je služba *SelectedWorks* v rámci repozitářového systému Bepress (<http://works.bepress.com/>).

V některých zemích existují ještě další možnosti pro autoarchivaci. Například ve Velké Británii existuje volně dostupná služba *the Depot* <http://depot.edina.ac.uk/>, což je vlastně sběrný repozitář pro ty výzkumníky, kteří chtějí zveřejnit své práce v OA, ale jejichž mateřská instituce žádný institucionální OA-repozitář (zatím) neprovozuje. Jakmile instituce svůj repozitář vytvoří, všechny práce jejích členů jsou do něj z Depotu přesunuty.

Podobně jako u zlaté cesty k otevřenému přístupu (publikování v OA-časopisech) tak i u „zelených“ (autoarchivace v OA-repozitářích) jde o zajištění přístup hlavně ke kvalitní vědecké literatuře. Primárně jde tedy o články, které prošly nějakou formou kontroly vědecké kvality (nejčastěji v podobě recenzního řízení v časopisech<sup>2</sup>). Existují ale i specializované repozitáře, které se zaměřují na jiný druh vědeckých informací; jako příklad mohou posloužit datové OA-repozitáře, které shromažďují a zpřístupňují vědecká data a výsledky experimentů.

Aktuální přehled o dostupných OA-repozitářích lze získat na stránkách ROAR - *Registry of Open Access Repositories* <http://roar.eprints.org>

---

<sup>2</sup>Je vhodné zdůraznit, že různé vědecké obory používají různé formy vědecké komunikace. Zatímco ve většině přírodních věd převažují časopisecké články, v oblasti computer science jsou to spíše články ve sbornících z konferencí a v humanitních vědách zase jiné formy, například kapitoly v monografiích. Různé obory mají také různou tradici a zkušenosti s využíváním e-printů (elektronických verzí vědeckých prací). Dobře navržený repozitář by měl zohledňovat a podporovat zavedené formy komunikace v té-které disciplíně.

nebo OpenDOAR – *Directory of Open Access Repositories* <http://www.opendoar.org>. Užitečným nástrojem pro ty, kdo se zabývají nastavením vnitřních pravidel upravujících v dané instituci postupy a povinnosti pro vkládání prací do OA-repozitáře, může být služba ROARMAP – Registry of Open Access Repository Material Archiving Policies <http://www.eprints.org/openaccess/policysignup/>.

Souběžně s tím, jak vznikaly jednotlivé OA-repozitáře, objevila se i volně dostupná programová řešení pro implementaci institucionálních repozitářů (uvedme alespoň systémy EPrints a DSpace). Následně pak vznikly standardy OAI (Open Archive Initiative), které umožňují sklízet metadata z různých repozitářů a budovat nad nimi globální vyhledávací služby. Díky tomu je možné propojit jednotlivé distribuované OA-repozitáře do jednoho masivního virtuálního digitálního archivu; v něm může čtenář požadovanou práci vyhledat, aniž by věděl, ve kterém konkrétním repozitáři je vlastně fyzicky uložena.

#### 4 Autorská a vydavatelská práva

Klíčovou otázkou pro vkládání publikovaných vědeckých prací do otevřených repozitářů je ošetření autorských a vlastnických práv (copyrightu). Se samotnými právy autorů problém vcelku není – koneckonců jsou to vědci samotní, kteří chtějí dát své práce bezplatně k dispozici a mají zájem na jejich co nejširším zpřístupnění. Míru ochrany svých prací si mohou zvolit volbou vhodné licence *Creative Commons* resp. *Science Commons* (podrobnosti viz článek Lukáše Grubera v tomto čísle Zpravodaje ÚVT MU). Problém je ovšem s právy vydavatelů.

Komerční vydavatelé vědecké literatury tradičně vyžadovali od autorů exkluzivní právo na vydání/zpřístupnění jejich prací. Následně pak nikdo jiný (ani autor samotný) nemohl práci bez souhlasu vydavatele zveřejnit. Pod tlakem vědecké komunity a hnutí OA začali vydavatelé v posledních letech tuto svou praxi měnit. Pozitivní zprávou je, že dnes již většina vydavatelů vědeckých časopisů (včetně takových gigantů jako je Elsevier či Springer) uplatňuje licenční politiky, které jsou vůči OA-repozitářům

mnohem vstřícnější. A to nejen co se týká možnosti vkládat do repozitářů s otevřeným přístupem *preprinty* prací (preprintem se rozumí ta verze práce, kterou autor nabídl vydavateli časopisu; tedy verze, která ještě neprošla recenzním řízením a z něj vyplývajícími úpravami)<sup>3</sup>.

Podle [5] umožňuje v současnosti 64% vydavatelů, aby autoři zveřejnili v OA-repozitářích i *postprinty* (verze práce po recenzním řízení). Pokud se toto číslo převede na počet časopiseckých titulů, pak to představuje již více než 90% všech recenzovaných vědeckých časopisů!

Je třeba však mít na paměti, že pojem „postprint“ není definován u všech vydavatelů stejně. Jedni tím rozumí verzi práce se všemi obsahovými změnami po recenzním řízení, ale ještě před finálními jazykovými a grafickými úpravami provedenými redakcí časopisu. Jiní rozumí pod tímto pojmem hotovou elektronickou verzi práce vysázenou v designu a stylu daného časopisu. Zatímco v některých případech tak může autor vložit do OA-repozitáře přímo pdf-verzi článku poskytnutou mu vydavatelem, v jiných případech musí vyrobit z finální verze článků svůj vlastní pdf-soubor, který pak teprve může použít pro OA. Jak uvádí [6], v současnosti existuje dokonce více různých licenčních politik, než kolik je vydavatelů (je to dáno tím, že někteří vydavatelé uplatňují vícero politik pro různé jimi vydávané časopisy).

Obrovským pomocníkem, který umožňuje zorientovat se v džungli licenčních politik a zjistit aktuální podmínky, za kterých umožňuje ten který vydavatel vkládat publikované práce do OA-repozitářů, je služba SHERPA/RoMEO <http://www.sherpa.ac.uk/romeo/>. Rozděluje všechny vydavatele do různobarevných skupin podle volnosti poskytnuté autorům: zelená, modrá, žlutá, bílá. Zatímco zelení vydavatelé povolují zveřejnit v OA všechny verze práce – od preprintu až po finální vydavatelskou pdf-verzi článku, tak bílí vydavatelé na opačném

<sup>3</sup>Právo volně (tj. bez souhlasu vydavatele) nakládat s preprintem ostatně měl autor vždy. Jen se někteří vydavatelé v minulosti zdráhali přijmout k publikování práce, které již byly ve formě preprintu zveřejněny.

pólu vůbec OA-přístup nepodporují. Aspoň nějakou formu OA podporuje 61 % vydavatelů uvedených v databázi (těch bylo v době psaní článku 646). U každého vydavatele je uveden přesný popis jeho licenčních podmínek. Podrobnější popis autorsko-právní problematiky a služby SHERPA/RoMEO podává článek Pavly Rygelové v tomto čísle Zpravodaje ÚVT MU.

## 5 Podpora OA ze strany grantových agentur a politických orgánů

Stále větší část poskytovatelů dotací na výzkum (grantové agentury, vládní instituce, ale i soukromé nadace) požaduje, aby výsledky jimi podporovaného výzkumu byly *povinně* zpřístupněny v režimu otevřeného přístupu. A jsou dokonce ochotni náklady na otevřené publikování finančně podpořit. Mezi prvními velkými grantovými agenturami, které začaly důsledně vyžadovat otevřený přístup k výsledkům podporovaného výzkumu, byl americký National Institutes of Health (od roku 2005 vyžadoval OA nepovinně, od roku 2007 již povinně). Od té doby počet poskytovatelů výzkumných dotací vyžadujících OA-publikování rychle roste. Jejich přehled (včetně podrobného popisu pravidel jednotlivých agentur) je průběžně aktualizován v rámci služby SHERPA/JULIET <http://www.sherpa.ac.uk/juliet/>.

Významnou událostí v evropském měřítku je pilotní iniciativa Evropské komise „Open Access Pilot in FP7“ [9] vyhlášená v srpnu 2008. Ta požaduje, aby řešitelé zajistili otevřený přístup během 6-12 měsíců od publikování ke všem recenzovaným článkům vzniklým jako výsledek projektů Sedmého rámcového programu EU FP7 2007-2013 v oblastech: Energy, Environment, Health, ICT, e-Infrastructures (6měsíční lhůta) a Science in Society, Socio-economic Sciences and Humanities (12měsíční lhůta). Projekty ve vyjmenovaných oblastech tvoří asi 20% celého rozpočtu na vědu ve FP7. Odhaduje se, že by mohly vygenerovat kolem 100 000 článků [6], které musí být uloženy do institucionálních nebo předmětových OA-repozitářů (pilotní iniciativa podporuje „zelenou cestu“ k OA; vedle toho ale poskytuje EC také možnost refundovat náklady na publikování článků v OA-časopisech). EC rovněž

vypsala tendr na vytvoření celoevropského repositoře, kam by mohli vkládat své články autoři, kteří nemají k dispozici odpovídající institucionální nebo předmětový OA-repozitář [10]. Při publikování v časopisech, které neumožňují otevřený přístup, jsou autoři zavázáni vyjednat s vydavatelem takové licenční smlouvy, které jim umožní naplnit požadavky pilotní iniciativy Evropské komise. Řešitelé těchto projektů, které pod pilotní iniciativu nespádají, jsou vyzváni, aby také zvážili možnost publikovat výsledky v režimu otevřeného přístupu. Na základě výsledků pilotní iniciativy FP7 plánuje Evropská komise stanovit politiku otevřeného přístupu pro následující rámcový program FP8.

Iniciativa Evropské komise není ani zdaleka jediná. Podobnou politiku nastavila již dříve Evropská výzkumná rada [11]. Také EUROHORCS - European Heads of Research Councils, který sdružuje většinu hlavních veřejných agentur z 23 evropských zemí, vydal pro své členy doporučení v podobě „minimálního standardu“ pro podporu OA [12]. V něm doporučuje členským organizacím, aby se připojily k Berlínské deklaraci a přijaly doporučení k otevřenému přístupu vydané Evropským poradním výborem pro výzkum EURAB [13].

Podpora otevřenému přístupu v podobě různých iniciativ, deklarací, aliancí, programů, projektů a aktivit po celém světě je obrovská. Není naším cílem podat jejich vyčerpávající přehled - ten lze nalézt jinde, např. v [8]. Stejně tak roste politická podpora, a to i z těch nejvyšších míst. Zmíňme, že již v roce 2004 podepsaly členské země OECD (včetně České republiky) *Deklaraci o přístupu k výzkumným datům z veřejně podporovaného výzkumu*, která přebírá řadu myšlenek OA, a vzaly tak na sebe politický závazek postupovat v souladu s přijatými doporučeními [14].

Z akademického světa připomeňme již jen doporučení Asociace evropských univerzit EUA přijaté v březnu 2008 [15], které doporučuje vedením univerzit iniciovat vytváření institucionálních OA-repozitářů a ustanovení univerzitních postupů, které by vyžadovaly od vědců vkládat své odborné publikace do těchto repositořů.

## 6 Závěr

Správně zvolený otevřený přístup může přinést výhody všem hráčům v oblasti vědeckého publikování – autorům, čtenářům, knihovnám, univerzitám, grantovým agenturám i běžným občanům.

Přestože má otevřený přístup řadu výhod a stále rostoucí podporu na všech úrovních, neznamená to, že se prosadí sám od sebe. Je nutné, aby všechny subjekty podílející se na procesu tvorby vědeckého poznání a jeho šíření změnilы zavedené způsoby chování. Od způsobů publikování a zpřístupňování informací, přes efektivnější procesy pro posuzování vědecké kvality, až po řízení vědy a výzkumu v jednotlivých institucích – včetně nových metod měření výkonnosti a hodnocení jednotlivých vědeckých pracovníků. Současně s tím je však nutná i aktivní podpora jednotlivců. V České republice převládá zatím spíše rezervovaný postoj – jak od akademických pracovníků, tak i vedení institucí. První vlaštovkou se zdá být Akademie věd ČR, která se k hnutí OA oficiálně přihlásila a začala podnikat i první praktické kroky. Přidají se i univerzity?

## Literatura

- [1] BOAI – Budapest Open Access Initiative, 2002. <http://www.soros.org/openaccess/read.shtml>
- [2] Bethesda Statement on Open Access Publishing, 11.4.2003. <http://www.earlham.edu/~peters/fos/bethesda.htm>
- [3] Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, 2003. <http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>
- [4] Peter Suber. *Open access Overview*. <http://www.earlham.edu/peters/fos/overview.htm>
- [5] DRIVER. Digital Repository Vision for Europe. Open Access Frequently Asked Questions. <http://www.driver-support.eu/faq/oafaq.html>
- [6] Leo Waaijers. *A short overview in international trends in Open Access*. Sborník konference CASLIN 2009 “Institutional Online Repositories and Open Access”, ZČU Plzeň, 2009. Online dostupný na <http://www.knihovna.zcu.cz/Caslin/Caslin09.pdf>
- [7] Sherpa. Authors and Open Access. <http://www.sherpa.ac.uk/guidance/authors.html#whatoa>
- [8] Open Access Directory. Compendium of factual lists about open access. [http://oad.simmons.edu/oadwiki/Main\\_Page](http://oad.simmons.edu/oadwiki/Main_Page)
- [9] European Commission. *Open Access Pilot in FP7*. <http://ec.europa.eu/research/science-society/index.cfm?fuseaction=public.topic&id=1680>
- [10] Repository supporting FP7. [ftp://ftp.cordis.europa.eu/pub/fp7/ict/docs/e-infrastructure/20081216-infoday-morais\\_en.pdf](ftp://ftp.cordis.europa.eu/pub/fp7/ict/docs/e-infrastructure/20081216-infoday-morais_en.pdf)
- [11] European Research Council. *ERC Scientific Council Guidelines for Open Access*. 17.12.2007. [http://erc.europa.eu/pdf/ScC\\_Guidelines\\_Open\\_Access\\_revised\\_Dec07\\_FINAL.pdf](http://erc.europa.eu/pdf/ScC_Guidelines_Open_Access_revised_Dec07_FINAL.pdf)
- [12] European Heads of Research Councils. *EUROHORCs' Recommendations on Open Access (OA)*. [http://www.eurohorcs.org/SiteCollectionDocuments/EUROHORCs\\_Recommendations\\_OpenAccess\\_200805.pdf](http://www.eurohorcs.org/SiteCollectionDocuments/EUROHORCs_Recommendations_OpenAccess_200805.pdf)
- [13] European Research Advisory Board. *Scientific Publication: Policy on Open Access*. December 2006. [http://ec.europa.eu/research/eurab/pdf/eurab\\_sci/pub\\_report\\_recomm\\_dec06\\_en.pdf](http://ec.europa.eu/research/eurab/pdf/eurab_sci/pub_report_recomm_dec06_en.pdf)
- [14] OECD Principles and Guidelines for Access to Research Data from Public Funding. 2007. <http://www.oecd.org/dataoecd/9/61/38500813.pdf>
- [15] Recommendations from the EUA Working Group on Open Access. Dostupné z WWW: [http://www.eua.be/fileadmin/user\\_upload/files/Policy\\_Positions/Recommendations\\_Open\\_Access\\_adopted\\_by\\_the\\_EUA\\_Council\\_on\\_26th\\_of\\_March\\_2008\\_final.pdf](http://www.eua.be/fileadmin/user_upload/files/Policy_Positions/Recommendations_Open_Access_adopted_by_the_EUA_Council_on_26th_of_March_2008_final.pdf) □

# Licence Creative Commons a jejich využití v OA

Lukáš Gruber, Národní knihovna ČR

## 1 Úvod

Open Access (OA) jako filosofie přístupu k vědeckým informacím přináší dva základní požadavky na proces publikování informací. Jedná se o požadavky kladené na autorská práva vztahující se k publikovanému dokumentu, a na požadavky týkající technické specifikace metadatavého popisu a uložení textu dokumentu v otevřených repozitářích či v otevřených časopisech. Zatímco oblasti „fyzického“ uložení OA publikací a návazných služeb jejich registrace a vyhledávání byl věnovaný prostor v úvodníku, tento článek pojednává o problematice vypořádání autorských práv. Jako zprostředkující nástroj jsou zde popsány veřejné licence Creative Commons.

Z dokumentů BBB iniciativy se podrobněji autorskými právy k dokumentům zabývá Prohlášení z Bethesdy [1] a Berlínská deklarace [2], a to formou výčtu možností nakládání s dokumentem. Obě specifikace jsou vzájemně slučitelné. Každý dokument, publikovaný v souladu s OA lze „kopírovat, využívat, distribuovat, přenášet a veřejně zobrazovat a vytvářet a distribuovat dokumenty od něho odvozené, a to na jakémkoliv digitálním médiu pro kterýkoliv rozumný účel za předpokladu uvedení údajů o autorství.“<sup>1</sup>

Aby bylo možné tato práva zajistit, je potřeba ošetřit vztah „autor/nakladatel – čtenář“ nad rámec autorského zákona (AutZ). Ochrana majetkových a osobnostních autorských práv, která je vtělena v AutZ, sama o sobě nakládání s autorskými díly (článek, preprint apod.) s požadavky OA neumožňuje. AutZ však umožňuje držitelům autorských práv část svých majetkových práv

<sup>1</sup>Přeloženo z textu Prohlášení z Bethesdy: „The author(s) and copyright holder(s) grant(s) to all users a free, irrevocable, worldwide, perpetual right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship, as well as the right to make small numbers of printed copies for their personal use.“

převést, a de facto umožnit druhé straně liberálnější nakládání s autorským dílem. K převodu práv slouží licenční smlouvy neboli licence.

Licence Creative Commons (CC) jsou licenční schémata používaná pro účely legálního šíření a sdílení autorských děl převážně na internetové síti. Počet licencovaných děl se na konci roku 2009 pohyboval okolo 250 mil. [3] a licence jsou již zavedeny v 52 státech světa [4]. Díky své srozumitelnosti, jednoduchosti aplikace a mezinárodní podpoře jsou CC licence využívány v oblastech, kde je potřeba v souladu se autorským právem zastřešit šíření, sdílení a upravování autorských děl. Takovými oblastmi jsou mezi jinými problematika využívání informací veřejného sektoru (PSI), budování kolekcí e-learningových materiálů (OER), nebo právě OA.

Mezi největší světové OA nakladatele používající CC licence patří BiomedCentral <http://www.biomedcentral.com>, Public Library of Science <http://www.plos.org> nebo Hindawi Publishing Corporation <http://www.hindawi.com>. Za Českou republiku uvádím časopis Central European Journal of Public Policy <http://www.cejpp.eu>, který vydává Fakulta sociálních věd Univerzity Karlovy.

## 2 Charakteristika licencí Creative Commons

CC licence byly poprvé představeny světu organizací Creative Commons na sklonku roku 2002. Jedná se o sadu předem připravených licenčních ujednání, které vlastník autorských práv může přiřadit ke svému dílu. Podnětem pro vznik CC licencí byla snaha vytvořit nástroj v souladu s autorským právem na podporu kreativity v souvislosti s téměř neomezenými možnostmi šíření a vzájemného sdílení autorských děl v prostředí internetu.

CC licence pochází z rodiny tzv. *veřejných licencí*, stejně jako známé GNU licence<sup>2</sup>. Na rozdíl od GNU je možno CC licence aplikovat na širokou škálu autorských děl tak, jak je v § 2, odst. 1

<sup>2</sup>Nejznámější zástupci GNU licencí jsou GPL (General Public License) používaná pro šíření a sdílení zdrojových kódů open source softwaru a FDL (Federal Public License), která primárně slouží pro šíření dokumentace k softwaru a hardwaru.



definuje AutZ, a to jak v hmotné, tak v nehmotné (elektronické) podobě. Veřejné licence se od pojetí tradičních licenčních smluv liší tím, že jsou upraveny speciálně pro síťové prostředí. Mezi jejich základní vlastnosti patří [8]:

- jsou poskytovány neurčitému okruhu osob (resp. každému, kdo má přístup k internetu);
- jsou poskytovány na celou dobu ochrany díla (70 let po smrti autora);
- získá je automaticky každý, kdo díla v souladu s licencí využívá;
- nabyvatelé licence jsou oprávněni dílo šířit;
- licence je nabízena bezúplatně.

## 2.1 Typy Creative Commons licencí

CC licence lze chápat jako jednotlivé kombinace poskytnutých a vyhrazených práv autora k dílu, které tvoří konkrétní typ CC licence. Zjednodušeně lze tato práva zobrazit jako tzv. licenční prvky. Ke každému licenčnímu prvku je přiřazen konkrétní, mezinárodně srozumitelný grafický symbol. Mezi licenční prvky, které označují *poskytnutá práva* autora, viz tabulka 1.

Právo dílo šířit znamená souhlas autora k vytváření kopií jeho díla a tyto kopie dále v nezměněné podobě šířit. Zároveň je možné licencované dílo šířit jako součást souborného díla (např. článek v časopisu, příspěvek ve sborníku). Právo dílo upravovat opravňuje uživatele dílo pozměnit, nebo-li zasáhnout do jeho obsahové integrity. Příkladem takového užití může být např. dramaturgie, překlad literárního díla, úprava digitálních fotografií či remixování hudebních skladeb.

*Vyhrazená práva* autora jsou podmínky, za jakých je možné poskytnutá práva k dílu vykonávat. Pro tato práva existují licenční prvky dle tabulky 2.

Prvek s názvem „uved'te autora“ obsahují všechny typy CC licencí. Údaje o autorovi a dílu je třeba uvést způsobem, který autor díla preferuje. Pokud tento způsob není znám, CC licence stanovují minimální povinný rozsah uvádění autorských údajů: jméno autora (nebo pseudonym), název díla a v případě úpravy díla také způsob, jakým bylo dílo upraveno. Povinnost zachovat licenci vychází z účelu, aby veškerá, následně

vzniklá odvozená díla byla šířena pod stejnou licenci, jaká se vztahuje na původní dílo. Podmínka nekomerčního využití pramení z práva autora zakázat uživateli jakékoliv finanční obohacení na jeho úkor. Prvek „nezasahujte do díla“ je aplikací osobnostního práva autora na nedotknutelnost jeho díla.

Srozumitelnou kombinací licenčních prvků vzniká šest typů CC licencí. V tabulce 3 jsou seřazeny z hlediska možností nakládat s dílem od nejvíce liberální k nejvíce restriktivní. V oblasti OA se nejčastěji používá licence BY.



Přestože se jednotlivé licence od sebe liší, několik vlastností mají společných [9]:

- při nakládání s dílem musí být uvedeny údaje o autorství;
- nabyvatel licence nesmí technickými prostředky omezovat přístupnost díla;
- každá vzniknuvší kopie licencovaného díla musí obsahovat odkaz na původní licenci;
- CC licence jsou neodvolatelné (autor sice může licenci stáhnout zpět, ale nemá právo zasahovat do práv nabyvatele původní licence při nakládání s dílem);
- CC licence zanikají v případě porušení licenčních podmínek.





CC licence nikterak neomezují ochranu autorských práv. Naopak jsou její extenzí a přizpůsobením zejména v internetovém prostředí, kde nakládání s autorskými díly vykazuje jiné parametry než je tomu v „analogovém“ světě. A kde se často vyplatí deklarovat práva spjatá se zpřístupněním informací, a tím posílit svou pozici v případě porušení autorských práv druhou stranou.

## 2.2 Vrstvy CC licencí

CC licence, na rozdíl od ostatních typů veřejných licencí, používají pro zobrazení licenčních podmínek dvě vrstvy určené různým cílovým skupinám: zkrácený text (Commons Deed) pro běžného uživatele a text v plném znění (Legal Code) vytvořený v právní terminologii. Zároveň se pro účely indexace používá speciální syntaxe metadata (Digital Code), na základě které je možné licencovaná díla v prostředí internetu vyhledat.

Symbol	Název	Práva k dílu
	Právo dílo šířit	Dílo je možné šířit, tzn. kopírovat, distribuovat a sdělovat veřejnosti.
	Právo dílo upravovat	Dílo je možné samostatně pozměňovat či upravovat nebo jej jako celé či jako část zpracovat s jiným dílem.

Tabulka 1: Poskytnutá práva autora

Symbol	Název	Podmínky užívání díla
	Uveďte autora	Kdykoliv je dílo šířeno ve své změněné či nezměněné podobě, je nutno uvést údaje o původním autorovi a dílu.
	Zachovejte licenci	Pokud uživatel dílo upraví, má povinnost výsledek své práce šířit pod stejnou licenci, jaká se vztahuje k podkladovému dílu.
	Neužívejte dílo komerčně	Uživatelovi nesmí z užívání díla plynout zisk.
	Nezasahujte do díla	Uživatel nesmí dílo upravovat.

Tabulka 2: Vyhrazená práva autora

Text určený pro běžného uživatele obsahuje výčet poskytnutých a vyhrazených práv, která autor ke svému dílu přisoudil. Smyslem textu je seznámit uživatele se základními podmínkami při nakládání s dílem, aniž by byl nucen studovat kompletní text licence. Hlavní důraz je kladen na srozumitelné vysvětlení licenčních podmínek i za pomoci výše uvedených grafických symbolů. Součástí zkráceného znění je odkaz na plné znění licence.

Plné znění licence je legislativně závazný dokument, který je vytvořen v souladu se stávajícím autorským právem. Kompletní text licence zahrnuje mimo jiné definici základních pojmů, podmínky pro poskytnutí licence, veškerá omezení licence, podmínky ukončení licence apod. Stejně jako v případě zkráceného znění, i text plného znění je vytvořen zvlášť pro každý typ licence.

Pro účely vystavení díla pod CC licenci se používají metadata ve vlastní formátové specifikaci s názvem Creative Commons Rights Expression Language (ccREL). Jedná se o úpravu popisného modelu RDF s využitím Dublin Core. Metadata je možno pro své dílo manuálně vytvořit nebo pomocí dostupných nástrojů vygenerovat.

### 2.3 Aplikace Creative Commons licencí

Aby bylo dílo pod CC licencí náležitě vystaveno, je nutné k dílu připojit text CC licence nebo URL odkaz na adresu, kde se text nachází. Tímto úkolem je zajištěno, aby se uživatel díla seznámil s licenčními podmínkami a na jejich základě s dílem nakládal, a tím de facto uzavřel s autorem licenční smlouvu. Aby byla skutečnost, že je dílo vystaveno pod CC licencí zdůrazněna, doporučuje se taktéž u URL odkazu uvádět grafické symboly obsažených licenčních prvků nebo samotných CC licencí<sup>3</sup> a libovolně formulované sdělení, že dané dílo podléhá CC licenci.

Pro aplikaci CC licencí na webové stránky lze využít webový generátor (<http://creativecommons.org/choose/?lang=cs>). Generátor se ovládá pomocí webového formuláře, kde na základě zodpovězení dvou otázek je vygenerován HTML řetězec, který obsahuje odkaz na text a logo vybrané CC licence. Po vložení řetězce do zdrojového kódu stránek se informace objeví na stránkách.

<sup>3</sup>Tato loga jsou pro účely licencování autorských děl volně k dispozici na serveru organizace Creative Commons: <http://creativecommons.org/about/downloads/>.

Označení licence	Licenční prvky		Název licence
	Práva	Povinnosti	
BY			Uveďte autora
BY-SA			Uveďte autora - Zachovejte licenci
BY-NC			Uveďte autora - Neužívejte dílo komerčně
BY-ND			Uveďte autora - Nezasahujte do díla
BY-NC-SA			Uveďte autora - Neužívejte dílo komerčně - Zachovejte licenci
BY-NC-ND			Uveďte autora - Neužívejte dílo komerčně - Nezasahujte do díla

Tabulka 3: Typy CC licencí

### 3 Česká verze licencí Creative Commons

Přestože byly licence CC původně vytvořeny pro americké prostředí, jejich myšlenka se rychle přenesla do celého světa. Licence postupně začaly využívat i autoři jiné než americké provenience. Postupem času vznikla mezinárodní iniciativa s cílem vytváření národních verzí CC licencí. Organizace Creative Commons zřídila pro účely koordinace prací při vytváření národních verzí licencí kancelář se sídlem v Berlíně. Národní verze vycházejí z obecné verze licenčního ujednání, jsou vytvořeny v národním jazyce a přizpůsobeny příslušnému právnímu řádu dané země. V současné době je k dispozici 52 různých národních verzí CC licencí, z nichž největší procento zastupují země evropského kontinentu.

V České republice byla česká verze CC licencí uvedena v rámci tiskové konference 16. dubna 2009. Vytvořilo ji sdružení Iuridicum Remedium ve spolupráci s Národní knihovnou České republiky a Ústavem práva autorského, práv průmyslových a práva soutěžního při Právnické fakultě Univerzity Karlovy v Praze. Při vytváření textu licence autoři přihlíželi zejména k aktuálnímu znění českého AutZ, aby byl text licence v souladu se zákonem z věcného a terminologického hlediska. Součástí procesu bylo i dvoukolové veřejné připomínkové řízení, kdy byl návrh textu vystaven na webu a kdokoliv se k němu mohl vyjádřit. Pozdější fáze textu byla na základě anglického překladu konzultována s berlínskou pobočkou organizace Creative Commons.

Všechny národní verze CC licencí jsou hostovány na serveru organizace Creative Commons, viz

<http://www.creativecommons.cz/zakladni-informace-o-cc/ceske-cc-licence/>.

### 4 Science Commons

Science Commons (SC) je vedlejší projekt organizace Creative Commons, který je oborově zaměřen na oblast vědy a výzkumu v přírodních vědách, s důrazem na oblast medicíny. Program navazuje na využití CC licencí v oblasti OA a zároveň hledá cesty, jak vytvořit optimální podmínky pro sdílení a využívání vědeckých dat. Jeho hlavním smyslem je najít lék na nešvary dnešní vědy, jakými jsou duplicita výzkumných programů, prodleva mezi objevem a jeho publikací, překážky pro sdílení výsledků výzkumných činností mezi jednotlivými výzkumnými institucemi či roztržitost relevantních informačních zdrojů. Činnost SC je v současné době rozdělena na tři hlavní cíle: vytvoření nástrojů na registraci a licencování výzkumných dat pro účely vzájemného sdílení, vytvoření platformy, která umožní jednoduchý („one-click“) přístup k výzkumným materiálům a zvýšení efektivity vyhledávání informačních zdrojů formou vývoje a implementace vlastní metadatové syntaxe [6].

Přestože byl program SC spuštěn v roce 2005, již nyní vykazuje aplikovatelné výsledky. Asi nejzajímavější v tomto ohledu je vytvoření nástroje s označením *Material Transfer Agreement* (<http://mta.sciencecommons.org/chooser>), který zastřešuje výměnu výzkumných dat mezi jednotlivými institucemi. Jedná se o nástroj, který funguje jednak jako licence, upravující podmínky využívání výzkumných dat, ale také

jako nástroj jejich registrace. SC se v současné době soustředí na dokončení pilotního projektu s NeuroCommons [7], jehož výstupem má být otevřená databáze vědeckých materiálů z oblasti věd o nervovém systému (neuroscience). Tato databáze má být realizací tří hlavních výše uvedených cílů SC.

## 5 Závěr

CC licence, zejména licence CC-BY, jsou v kontextu OA často používány pro účely deklarace práv vztahující se k publikovanému dílu. Více informací o CC licencích naleznete na stránkách organizace Creative Commons <http://creativecommons.org> nebo na stránkách věnovaných české verzi CC licence <http://www.creativecommons.cz>. Pokud zde požadované informace nenaleznete, můžete se obrátit na autora tohoto článku, potažmo na celý tým pro vytvoření české verze CC licence, a to zasláním dotazu na adresu [info@creativecommons.cz](mailto:info@creativecommons.cz).

## Literatura

- [1] Bethesda Statement on Open Access Publishing [online]. 2003-, [cit. 2009-11-15]. <http://www.earlham.edu/~peters/fos/bethesda.htm>
- [2] Open Access to Knowledge in the Sciences and Humanities [online]. 2003-, [cit. 2009-11-15]. <http://oa.mpg.de/openaccess-berlin/berlindeclaration.html>
- [3] Creative Commons : metrics [online]. 2005-, [cit. 2009-11-15]. <http://wiki.creativecommons.org/Metrics>
- [4] Creative Commons : international [online]. 2005-, [cit. 2009-11-15]. <http://creativecommons.org/international>
- [5] Creative Commons Česká republika [online]. 2008-, [cit. 2009-11-15]. <http://www.creativecommons.cz>
- [6] Science Commons : about Science Commons [online]. 2005-, [cit. 2009-11-15]. <http://sciencecommons.org/about>
- [7] Science Commons : The Neurocommons [online]. 2005-, [cit. 2009-11-15]. <http://sciencecommons.org/projects/data>
- [8] JANSÁ, Petr. Právní aspekty implementace projektu „Creative Commons“ v České republice. [cit. 2009-11-15] Praha, 2008. 99 s., 6. s. příl. Diplomová práce (Mgr.). Univerzita Karlova, Právnická fakulta. [http://www.creativecommons.cz/wp-content/uploads/dp\\_petr\\_jansa\\_komplet\\_xmp.pdf](http://www.creativecommons.cz/wp-content/uploads/dp_petr_jansa_komplet_xmp.pdf)
- [9] FITZGERALD, Brian. Open Content Licensing (OCL) for Open Educational Resources [cit. 2009-11-15] In: Proceedings OECD Expert Meeting on Open Educational Resources. Malmö, 2005. <http://eprints.qut.edu.au/archive/00003621> □

## Služba SHERPA/RoMEO – další život vědeckého článku

*Pavla Rygelová, VŠB-TU Ostrava*

### 1 Úvod

Poslední desetiletí dochází ke vzniku obrovského množství iniciativ na podporu otevřeného přístupu k vědeckým informacím. Internet a další informační technologie umožňují zásadní změny ve způsobu vědecké komunikace. Vědecká komunita se začala zajímat o bezplatné a neomezené šíření a sdílení vědeckých poznatků a do popředí se dostala problematika šíření elektronických preprintů a postprintů, neboli tzv. *zelená cesta otevřeného přístupu*. Téma úzce souvisí s otázkou autorsko-právních vztahů mezi autorem článku a vydavatelem vědeckého časopisu. Vlastníky práv k šíření díla se zpravidla stávají vydavatelé vědeckých časopisů. S postupnou adaptací otevřeného přístupu jako standardního postupu šíření vědeckých informací se však tato situace mění a autoři vědeckých článků získávají větší kontrolu nad tím, jak mohou se svým článkem nakládat. Jaké možnosti autoři získali, ukazuje služba SHERPA/RoMEO [1], která se zabývá vztahem autora, vydavatele a autorského práva k autoarchivaci elektronických preprintů a postprintů v digitálních repozitářích.

Vzhledem k tomu, že autoarchivaci e-printů (souhrnný název pro elektronické preprinty a postprinty) lze stále považovat za poměrně nový

model distribuce vědeckých poznatků a je především předmětem zájmu vědeckých pracovníků, ze strany vydavatelů nedošlo zatím k jejímu jednoznačnému akceptování. Existují různé kombinace toho, co vydavatel svému autorovi povoluje a co zakazuje. Autoři mohou být v této situaci bezradní, a proto vznikla služba SHERPA/RoMEO, která autorům, knihovnám a ostatním institucím pomáhá orientovat se v nepřehledné síti různých přístupů u různých vydavatelů.

## 2 Vymezení některých pojmů

Než rozvineme praktické aspekty autoarchivace e-rintů, je potřeba si nejprve ujasnit některé pojmy, které jsou v textu používány. Jejich přesným definováním se vysvětlí některé nejasnosti a usnadní se pochopení celé problematiky.

Nejprve několik slov k samotnému pojmu autoarchivace (*self-archiving*). *Autoarchivací* je myšleno, že *autor sám vloží* svůj digitální dokument (v tomto případě preprint nebo postprint svého článku) do digitálního repozitáře nebo jej zveřejní například na své webové stránce. Z hlediska autorsko-právního je v podstatě jedno, jestli akt uložení provede opravdu sám autor, a také v mnohých případech se praktikuje spíše zprostředkovaný způsob vložení dokumentů do repozitáře (*mediated deposit*) jinou pověřenou osobou. Proto dále tento pojem budeme psát závorkami u předpony „auto“, tedy (auto)archivace, čímž chceme vyjádřit, že vložení dokumentu do repozitáře samotným autorem není bezpodmínečně nutné. Autor samozřejmě musí být seznámen a měl by souhlasit se skutečností, že jím vytvořené práce jsou jeho jménem dále zveřejňovány.

Je nutné vysvětlit také pojmy preprint, postprint a e-print. Databáze SHERPA/RoMEO používá termín *preprint* pro podobu článku před recenzním řízením. Tato definice je blízká autorům, pro které jakékoli modifikace článku znamenají samotný recenzní proces. Jinak chápou termín preprint někteří vydavatelé, kteří za preprint považují obsahově konečnou podobu článku s dokončeným recenzním řízením, ale bez konečného formátování pro tisk. To, co vydavatelé chápou jako preprint, je v případě databáze

SHERPA/RoMEO bráno jako postprint. *Postprint* označuje podobu článku, který prošel recenzním řízením. Pro zjednodušení budeme v textu dále pro preprinty a postprinty užívat slovo *e-print* [2] jako souhrnné označení obou pojmů.

## 3 Smlouva s vydavatelem

Článek vychází z modelové situace, kdy se autorovi vědeckého článku podařilo dopsat odborný text a usiluje o jeho zveřejnění ve vědeckém, nejlépe prestižním časopise. Autor uzavírá s vydavatelem smlouvu o vydání článku (*copyright transfer agreement*), která zpravidla určuje, že autor přenechává svá autorská práva vydavateli, jak vyplývá už ze samotného označení této smlouvy. Zároveň vydavatel může autorovi garantovat některá práva, např. k šíření e-rintu vědeckého článku, pořizování a šíření kopie článku, zahrnutí článku do publikace napsané nebo sestavované autorem, využití článku k výukovým účelům apod.

Na příkladu smlouvy vydavatele Wiley [3] si lze názorně ukázat, jak často vypadá základní nastavení autorsko-právních vztahů mezi vydavatelem a autorem vědeckého článku. Ve smlouvě se říká, že vydavatel získává autorská práva článek nebo jeho části například publikovat, znovu publikovat, přenášet, prodávat, distribuovat a dále využívat. Na druhou stranu vydavatel Wiley uděluje autorovi jmenovitě *právo sdílet s kolegy* tištěný nebo elektronický preprint ještě nevydaného článku v podobě, v jaké byl přijat k vydání v časopise. A dále jsou vyjmenovány podmínky, jak může být sdílení provozováno. Preprint může být vystaven k osobnímu nebo pracovnímu užití na autorových vlastních webových stránkách, nebo v interní síti jeho univerzity nebo společnosti, anebo na zabezpečených webových stránkách autorovy instituce.

Ze smlouvy vyplývá, že Wiley nechává autorům poměrně dost volného prostoru, jak mohou se svým článkem naložit, bohužel ovšem s výjimkou kýženého uložení preprintu do otevřeného digitálního repozitáře.

Další dva z významných vydavatelů, Springer a Elsevier, jsou ve vztahu k (auto)archivaci e-rintů

progresivnější. Oba vydavatelé vycházejí autorem vstřícně v tom, že jim umožňují ukládání e-printů také do digitálních repozitářů.

#### 4 Autorská práva ve vztahu k (auto)archivaci

Databáze SHERPA/RoMEO vznikla z jednoletého projektu RoMEO (*Rights Metadata for Open Archiving*) [4], jehož realizátoři si položili několik základních otázek vztahujících se k autorskoprávní stránce (auto)archivace. Především je zajímavé, kdo je skutečně vlastníkem autorských práv, jak vydavatelé zdůvodňují svůj požadavek, aby se autoři vzdali svých autorských práv, kdy dochází k podepsání smlouvy s vydavatelem, jak vypadá širší workflow (auto)archivace, jaké možnosti má autor při hledání vhodného vydavatele a zda může dojednat s vydavatelem příznivější podmínky. V rámci projektu vzniklo několik průzkumů hledajících odpovědi na tyto otázky a vzniknuvší databáze SHERPA/RoMEO představuje velice užitečné a využívané pokračování původního projektu.

Nejdůležitější je zřejmě otázka, proč vůbec vydavatelé vyžadují podepsání smlouvy *Copyright Transfer Agreement* (dále jen CTA), v níž si vyhrazují exkluzivní autorská práva. Předně, podle jedné z prezentací [5] projektu RoMEO z let 2002-2003, žádalo tehdy 90 % ze zkoumaných vydavatelů podepsání CTA, dalších 6 % vydavatelů požadovalo exkluzivní licenci a jen 4 % žádalo neexkluzivní licenci k publikování článku, kdy copyright zůstává v rukou autorů (dnes již je situace z pohledu autorů o něco příznivější). Vydavatelé vycházejí z dualistické podstaty autorského práva, které vymezuje osobnostní práva autora a vlastnická práva tvůrce původního díla. Vlastnická práva zaručují exkluzivní právo dílo užít, tj. dílo reprodukovat, šířit, distribuovat jeho kopie, pronajímat dílo, sdělovat dílo veřejnosti atd., viz např. přesné znění *Zákona 106/2006 Sb., O právu autorském ... , oddíl 3, paragraf 12* [6]. Jelikož se jedná o exkluzivní právo, pokud chce někdo využít jen některé z těchto práv, musí k tomu získat souhlas od vlastníka copyrightu.

Další z průzkumů projektu RoMEO se zajímal o to, jak a jestli vydavatelé dostatečně zdůvodňují podepsání smlouvy CTA. Pokud vůbec vydavatelé poskytují nějaké vysvětlení, většinou se omezí jen na krátké vyjádření, že CTA jim umožňuje ochranu před porušováním autorských práv. Takto to například zdůvodňuje vydavatel Taylor & Francis: *“Copyright assignment has the advantage of elimination of ambiguity, for example in negotiating subsidiary licences (for instance with database aggregators) or in dealing on an author’s behalf with plagiarism, abuse of moral rights, protecting intellectual property or copyright infringement.”* [7]

Trvání na podepsání smlouvy CTA pro autory spíše nevýhodné, ukazuje na exkluzivní postavení vydavatelů, jaké zaujímají ve vydavatelském průmyslu. Vydavatelé v podstatě autorům diktují své podmínky a přitom zapomínají na to, že by jejich obchody nemohly fungovat bez autorů a jejich vůle sdílet svá díla. Vedle vysokých cen předplatného je nevýhodnost smlouvy CTA dalším důvodem, proč podporovat myšlenky otevřeného přístupu, který se snaží narušit monopolní pozici vydavatelů vědeckých časopisů a vrátit autorům práva šířit své dílo způsobem adekvátním dnešním technologickým možnostem vědecké komunikace.

Důležitá je také otázka chvíle, kdy autor musí smlouvu CTA s vydavatelem podepsat. Zpravidla k tomu dochází před recenzním řízením a přijetím článku k vydání. Před podepsáním smlouvy je autor výhradním držitelem copyrightu a může s článkem nakládat jakkoli, neboli také může preprint článku uložit do repozitáře. Má to však jeden zásadní háček, neboť velká část vydavatelů vyžaduje po autorovi záruku, že článek nebyl dříve publikován. Toto prepublikační embargo, někdy označované jako tzv. „Ingelfingerovo pravidlo“ podle Franze Ingelfingera, editora časopisu *New England Journal of Medicine*, má zaručit časopisu jeho prvenství ve zveřejnění článku. Z tohoto pohledu se sice nejedná o problém právní, ale spíše jde o politiku časopisu, která implicitně vyjadřuje: *„We will not consider for publication any preprint that has been previously self-archived“* [8]

Problém však zase souvisí s přesnou interpretací obratu „dříve publikován“. Akademici souhlasí s výkladem, že publikovat znamená vydat článek v recenzovaném časopise. Vydavatelé se přiklánějí k obecnější definici, kdy publikovat znamená učinit informaci veřejně dostupnou [9]. Někteří vydavatelé sice výslovně uvádějí, že (auto)archivaci nepovažují za akt publikování, přesto autoři musí brát v potaz další nejednoznačnou překážku bránící co možná největší volné distribuci výsledků jejich práce.

Vydavatelé nebránící (auto)archivaci často vyžadují, aby k archivovanému e-printu bylo připojeno oznámení o tom, že e-print byl nebo bude vydán v příslušném časopise, s hypertextový odkazem na web vydavatele nebo na stránku s oficiálně vydaným článkem.

Jiné úskalí je skryto v tom, že vydavatelé někdy nezmiňují ve smlouvě CTA, jak budou práva vyrovnána v případě, že článek nebude přijat k vydání. Podle výzkumu projektu RoMEO 15 % zkoumaných vydavatelů tuto možnost ve smlouvě neuvádí. Znamená to, že odmítnutý článek nemůže být v podstatě už nikdy vydán u žádného jiného vydavatele. Autoři by si proto měli být vědomi, jakých práv se při podpisu smlouvy CTA vzdávají, a měli by bedlivě zvážit, zda skutečně chtějí o některá práva přijít. Zkrátka měli by vědět, že pro vydání článku není potřeba vzdát se veškerých autorských práv ve prospěch vydavatele a měli by si hledat takového vydavatele, který jim zaručuje co nejlepší podmínky – a to nejen ve vztahu k (auto)archivaci e-printů.

S řadou vydavatelů je možné vyjednávat o výhodnějších podmínkách publikování. Organizace SPARC ve spolupráci s iniciativou *Science Commons* nabízí pro takovou situaci online generátor [10] pro dodatek ke smlouvě s vydavatelem (tzv. addendum), který si každý autor může po vyplnění vytisknout a zaslat vydavateli.

## 5 Databáze SHERPA/RoMEO

Jak už bylo v úvodu zmíněno, situace kolem (auto)archivace elektronických preprintů a postprintů je velice komplikovaná. Přesto, že velká část vydavatelů se staví k (auto)archivaci spíše pozitivně (nyní je to již kolem 60 % vydavatelů

registrovaných v databázi SHERPA/RoMEO), je v podstatě nemožné najít jednotný postup, jak (auto)archivaci realizovat, aniž by v nějakém případě nedošlo k porušení konkrétní smlouvy s vydavatelem. Databáze SHERPA/RoMEO přesto pomáhá v orientaci alespoň v základních otázkách.

Databáze SHERPA/RoMEO je pokračováním projektu RoMEO podpořeného JISC z let 2002/2003. Organizace SHERPA se dohodla na přeměně výsledků projektu RoMEO na plně funkční databázi SHERPA/RoMEO, která je nyní financována organizací JISC a spravována Universitou v Nottinghamu. Dalšími partnery projektu jsou Wellcome Trust, British Library a organizace MIMAS.

V kostce se dá říct, že databáze SHERPA/RoMEO:

1. poskytuje informace o přístupu vydavatelů vědeckých časopisů k otázce (auto)archivace vědeckých článků, resp. jejich e-printů v digitálních repozitářích;
2. rozlišuje pomocí barevného označení, jak vydavatelé přistupují k (auto)archivaci e-printů. Barva zelená, modrá, žlutá a bílá indikují, jestli vydavatel (auto)archivaci povoluje, nebo ji povoluje s určitým omezením nebo ji neumožňuje vůbec;
3. umožňuje vyhledávání podle názvu časopisu, jeho ISSN nebo podle názvu jeho vydavatele;
4. sestavuje zvláštní seznam vydavatelů, kteří autorům dovolují využít přímo elektronickou verzi vydaného článku, zpravidla souboru PDF;
5. sestavuje zvláštní seznam vydavatelů, kteří za určitý poplatek nabízejí autorům zveřejnění jejich článku v režimu otevřeného přístupu (tzv. placený otevřený přístup).

Databáze nyní zahrnuje 650 vydavatelů především anglo-americké provenience. Přesná statistika vydavatelů podle zemí není k dispozici, ale čeští, stejně jako další evropští nebo mimoevropští vydavatelé, nejsou zastoupeni vůbec nebo jen v omezené míře. Paralelně vznikly k databázi SHERPA/RoMEO i další databáze. Pro španělské prostředí je budována obdobná služba DULCINEA [11], jejíž provozovatelé se zjevně inspirovali názvem projektu RoMEO a pojmenovali službu jménem postavy ze španělské literatury. V Austrálii je provozována databáze OAKList [12]. Obě databáze spolupracují

s SHERPA/RoMEO, která je stále nejobsáhlejším zdrojem.

Vydavatelé jsou rozděleni do čtyř barevných skupin podle toho, jak přistupují k (auto)archivaci e-printů. Zelení vydavatelé povolují využít preprint i postprint, modří vydavatelé pouze postprint, žlutí preprint a bílí (auto)archivaci nepodporují.

K vyhledání konkrétního vydavatele slouží webový formulář, který umožňuje vyhledávání podle několika hledisek. Vyhledávat je možné podle názvu vydavatele nebo názvu časopisu. Nedávno bylo významně rozšířeno pokrytí časopisů díky využití adresáře DOAJ, seznamu ENTREZ a služby Zetoc a vyhledávat časopisy je možné podle jejich ISSN nebo zkratky jejich názvu.

Větší pokrytí časopisů se promítlo i do podrobnějšího zobrazení výsledků vyhledávání. Vyhledání časopisu zobrazí výsledky v přehledné tabulce, z níž je okamžitě patrné, kdo časopis vydává a jestli má časopis, resp. jeho vydavatel vlastní záznam v databázi.

Informace o přístupu vydavatele k (auto)archivaci jsou v databázi udržovány na úrovni vydavatele. V záznamu se vyskytují tyto údaje:

1. název a odkaz na webové stránky vydavatele;
2. informace o archivaci e-printů;
3. informace o možnostech využití vydavatelské verze článku (PDF);
4. obecné podmínky využívání e-printů;
5. informace o povinném otevřeném přístupu ze strany poskytovatelů finančních zdrojů;
6. odkaz na dokument, odkud byly informace čerpány (často přímo text smlouvy CTA, nebo odkaz na webové stránky vydavatele, kde je popsán přístup vydavatele k otevřenému přístupu a otázkám (auto)archivace;
7. barevné označení vydavatele.

K bodu c): Služba SHERPA/RoMEO udržuje zvláštní seznam všech vydavatelů kteří k (auto)archivaci povolují využít vydavatelské verze článku ve formátu PDF. Nyní tento seznam obsahuje asi 120 vydavatelů, bohužel mezi nimi nefiguruje žádný z triumvirátu velkých vydavatelů Springer, Wiley nebo Elsevier. Celkem 84

vydavatelů (např. IEEE nebo vydavatelé otevřených časopisů PLoS a Hindawi) neklade autorům žádné překážky, ostatní podmiňují využití vydavatelské verze článku časovým embargem, které může být od 3 měsíců až po několik let. Někteří vydavatelé dovolí využít vydavatelské PDF až po udělení zvláštního povolení a několik vydavatelů zase za to vybírá určitý poplatek.

V neposlední řadě je třeba o databázi SHERPA/RoMEO říct, že všechny informace uvedené v záznamech vydavatelů jsou ověřovány přímo u vydavatelů, vychází se převážně z informací na webových stránkách a z dostupného textu samotné smlouvy s vydavatelem. Tyto informace lze považovat za korektní, mohou autorům sloužit jako vstupní zdroj při rozhodování, ve kterém časopise budou chtít publikovat; nemohou však být použity jako právní podklad pro dodatečné vyjednávání s vydavatelem o podmínkách už dříve podepsané smlouvy.

## 6 Závěr

Když se vrátíme k názvu tohoto článku, další život vědeckých článků mají v rukou jejich uvědomělí autoři, kteří si nenechají vnutit od vydavatelů nevýhodné podmínky pro distribuci svého díla. Předmětem článku nebylo vyzdvižení pozitivní přínosů aktu (auto)archivace, ale spíše naznačení okruhu autorsko-právních aspektů s (auto)archivací souvisejících. Situace není z tohoto pohledu zatím ideální. Ale na mnoha příkladech nejrůznějších iniciativ na podporu otevřeného přístupu lze dokladovat, že autoři nejsou bezbranní a že získali daleko větší možnosti, jak ovlivnit snadnější přístupnost svého článku, a tím si zvýšit potencionální okruh čtenářů a citační ohlas. Nepostradatelným pomocníkem autorům při řešení problémů s (auto)archivací mohou být knihovny v jejich mateřských institucích. Knihovny mohou napomáhat vytvořením vhodného prostředí pro ukládání e-printů, tj. budováním digitálních repozitářů a zavedením efektivního workflow přispívání do repozitářů – včetně podpory při řešení autorsko-právních otázek a kurátorské role nad kvalitou metadatového popisu e-printů, která je bezpodmínečně nutná pro



snadné vyhledávání a pro interoperabilitu s vnějším informačním prostředím.

## Literatura

- [1] *SHERPA/RoMEO : publisher copyright & self-archiving* [online]. c2006-2008. [cit. 2009-11-21]. <http://www.sherpa.ac.uk/romeo>.
- [2] HARNAD, Stevan. Electronic preprints and postprints. In DRAKE, Miriam A. (ed.). *Encyclopedia of library and information science*. New York : Marcel Dekker, c2003. <http://www.ecs.soton.ac.uk/~harnad/Temp/eprints.htm>.
- [3] Wiley. *Copyright Transfer Agreement* [online]. [cit. 2009-11-21]. <http://www3.interscience.wiley.com/homepages/central/cta/UKscta.pdf>.
- [4] *Project RoMEO* [online]. [cit. 2009-11-21]. <http://www.lboro.ac.uk/departments/lis/disresearch/romeo>.
- [5] Project RoMEO. *The author, the publisher, their copyright agreement and its terms* [online]. [cit. 2009-11-21]. <http://www.lboro.ac.uk/departments/lis/disresearch/romeo/RoMEO-Advocacy-presentation.ppt>.
- [6] Česko. Zákon č. 121/2000 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon). In *Sbírka zákonů České republiky*. 2006, částka 36, s. 1658-1685. <http://business.center.cz/business/pravo/zakony/autorsky/cast1h1.aspx>.
- [7] *Taylor & Francis authors services : Taylor & Francis' position on copyright and author rights* [online]. c2009. [cit. 2009-11-21]. <http://journalauthors.tandf.co.uk/preparation/copyright.asp#link1>
- [8] *Open Access and Institutional Repositories with EPrints : self-archiving FAQ : what if the publisher forbids preprint self-archiving?* [online]. [cit. 2009-11-21]. <http://www.eprints.org/openaccess/self-faq>.
- [9] GADD, Elizabeth; OPPENHEIM, Charles; PROBETS, Steve. RoMEO studies 4: an analysis of journal publishers' copyright agreements. *Learned publishing*. 2003, vol. 16, no. 4, p. 293-308. <http://alpsp.publisher.ingentaconnect.com/content/alpsp/lp/2003/00000016/00000004/art00009>. DOI: 10.1087/095315103322422053. Online ISSN 1741-4857.
- [10] Science Commons. *Scholar's Copyright Addendum engine* [online]. [cit. 2009-11-21]. <http://scholars.sciencecommons.org>.
- [11] *DULCINEA : derechos de copyright y las condiciones de auto-archivo de revistas científicas españolas* [online]. 2008-11-29. [cit. 2009-11-21]. <http://www.accesoabierto.net/dulcinea>.
- [12] Queensland University of Technology. *OAKList database* [online]. Last modified 07-July-2009. [cit. 2009-11-21]. <http://www.oaklist.qut.edu.au>. □

## Ukládání a sdílení vědeckých dat

Luděk Matyska, ÚVT MU

Výsledky vědecké práce jsou publikovány především formou monografií, článků v časopisech a příspěvků na konferencích. Je proto přirozené, že aktivity *otevřeného přístupu* jsou primárně zaměřeny tímto směrem. Články ovšem představují jen určité završení etapy odborné práce vědců, postavené nejen na předcházejících publikacích, ale využívající také další zdroje, včetně experimentálních měření, výsledků výpočtů, statistických analýz apod. Ověření správnosti publikovaných výsledků se standardně provádí opakováním (schopnost opakovat experiment a dojít ke stejným závěrům, tzv. *reproducibility*, je základním kamenem soudobé vědy), to však není možné bez přístupu ke stejným datům a měřením, jaké použili autoři publikované práce.

Při vědomí této souvislosti jsou odborné články zejména v renomovaných časopisech doprovázeny tzv. doplňkovým materiálem, který posuzovatelům a následně čtenářům přijatých článků umožňuje nahlédnout do původních dat, provést případnou nezávislou analýzu, a tak potvrdit (nebo vyvrátit) závěry publikované práce. Bohužel tento přístup již nedostačuje, protože autoři článku mohli do publikované kolekce primárních dat vybrat pouze ty údaje, které podporují jejich tvrzení, a ostatní údaje prostě „pominout“.

Existence těchto „pomínutých“ dat nemusí být snadno odhalitelná, posuzovatel či čtenář by museli zopakovat kompletně všechny experimenty, což samozřejmě není praktické a v řadě případů ani možné (např. při využití unikátních experimentálních zařízení). Potřeba přístupu k *úplným* sadám experimentálních dat, případně výsledkům rozsáhlých výpočtů se tak stává prakticky nezbytnou podmínkou pro udržení a další růst kvality publikovaných prací, zejména v prostředí standardní kontroly kvality formou peer review procesu.

Pro volný přístup k primárním experimentálním datům, laboratorním deníkům, výsledkům simulací a jiných výpočtů však hovoří další důvody. Řada vědeckých oblastí je ve stále větší míře závislá na unikátních, zpravidla extrémně drahých přístrojích. Příkladem mohou být velké urychlovače v CERNu, Fermilabu a dalších podobných pracovištích, velké genové sekvenátory schopné rychle analyzovat velký objem genetického materiálu a poskytnout sekvence nukleových kyselin, teleskopy a radioastronomická zařízení (např. Hubbleův teleskop) apod. Jejich použití vždy generuje enormní objemy dat (např. jeden běh sekvenátoru může generovat až desítky TB v jediném dni), přitom v drtivé většině případů je z těchto dat primárním vědeckým týmem, který inicioval měření, využit jen nepatrný zlomek. Pěkným příkladem může být sledování oblohy teleskopem. Přestože teleskop je například zaměřen pouze na jeden konkrétní sledovaný objekt (např. vzdálený kvasar), automaticky (jaksi „mimočodem“) zabírá určitý výřez oblohy a během pozorování shromáždí data o velkém množství dalších objektů. Tato data však primárním týmem (který zaměřil teleskop na kvasar) nebudou zpracována a bez vhodného přístupu k uchování a následnému zpřístupnění toto přesné pozorování určitého úseku oblohy nebude nikým nijak využito.

Jiným příkladem mohou být rozsáhlé klimatologické studie, které není možné udělat, pokud objem dat, které má konkrétní výzkumný tým k dispozici, nepřekročí určitou kritickou velikost (i numerické simulace je nutné kalibrovat s využitím historických záznamů). Jak ukazuje aktuální únik e-mailů z britského centra výzkumu kli-

matu, jsou přístup primárním datům a jejich interpretace umožněny zpravidla pouze poměrně úzké skupině vědců, kteří pak mohou – často v zájmu „dobré věci“ – podlehnout snaze ze svých studií vyloučit měření, která do jejich obrazu světa nezapadají a publikovat jen podpůrná data<sup>1</sup>,

Vedle otevřeného přístupu k publikacím a je bezprostředně doprovázejícím materiálům roste proto zájem o zpřístupnění primárních (nezpracovaných, a tedy ani nemanipulovaných) vědeckých dat. Tento zájem především vědecké komunity nalézá pozitivní ohlas i ve veřejných agenturách, které vědecký výzkum financují – neomezený přístup k výsledkům měření jiných snižuje náklady odstraněním duplicit a zvyšuje současně efektivitu vynaložených prostředků – výsledky měření prováděných za jedním konkrétním účelem mohou být využity pro úplně jiné studie. Zde je opět exemplárním příkladem astronomie, konkrétně Sloan Digital Sky Survey (<http://www.sdss.org>). V rámci této aktivity jsou dlouhodobě (od roku 2000) shromažďovány snímky oblohy ve vysokém rozlišení, které jsou zpřístupněny široké (nejen astronomické) veřejnosti. Jak ukázala analýza publikací za rok 2004, digitální archiv SDSS představuje fakticky nejúspěšnější světovou observatoř – články publikované s využitím dat ze SDSS byly citovány o 25% více než články využívající data Evropské jižní observatoře a o 60% více než články s daty z Hubbleova teleskopu. SDSS umožňuje nalézat velmi vzácné objekty, neboť obsahuje snímky stejných částí oblohy z různých let – data, která by jinak byla uložena v nepřístupných archivech nebo dokonce smazána poté, co je primární tým využije pro své původní záměry.

V posledních letech se stále více pro činnosti související s ukládáním a zpřístupněním primárních i odvozených vědeckých dat používá termín

---

<sup>1</sup>Abychom byli správně pochopeni: většina vědecké práce v experimentálních oborech spočívá ve schopnosti najít ve zdánlivě neuchopitelných a chaotických primárních datech nějaký řád, a ten pak dokázat. Je však nezbytné umožnit jiným studiím stejných primárních dat s možností, že najdou jiný řád a uspořádání. Špatné není ignorování některých primárních dat, špatná je pouze manipulace, která jiným vědcům znemožní přijít k jiným výsledkům.

*Data Curation* či *Digital Curation*<sup>2</sup>. *Data Curation* se věnuje udržení a zvyšování hodnoty důvěryhodné sady digitálně uložené informace (dat) tak, aby byly použitelné v současnosti i v budoucnosti [1].

Podstatnou součástí uvedené definice je pojem „důvěryhodná sada dat“, protože ta jasně předpokládá kontrolu (nejlépe opět veřejnou, tj. primárně vědeckou komunitou) kvality dat. Zpřístupnění v budoucnosti pak vyžaduje, aby primární data doprovázel jejich popis ve formě metadat, která často zahrnují i *provenanci*, tedy původ dat a veškeré předcházející manipulace s nimi. *Data Curation* také musí zajistit průběžnou transformaci dat do nových formátů, aby data byla přístupná (čitelná) bez ohledu na vývoj informačních technologií (přitom musí být zaručeno, že tyto transformace nemění původní obsah – riziko vzniku artefaktů může nastat např. u obrazové informace při nevhodné volbě nových kompresních formátů).

Zajímavým problémem je rovněž anonymizace dat, zvláště důležitá v lékařských či sociologických výzkumech – primární data nemohou být zpřístupněna sama o sobě, ale až poté, co jsou z nich odstraněny odkazy, které by mohly vést ke zcela konkrétním osobám. Nedílnou součástí *Data Curation* je i rozhodování o tom, kdy data ztrácejí hodnotu a je možné je nenávratně odstranit. Na rozdíl od knih, časopisů a dalších vědeckých publikací, které se snažíme uchovat trvale, je zde schopnost „ořezávání“ (pruning) kritickou a nezbytnou součástí práce s vědeckými daty. Posouzení dlouhodobé vědecké hodnoty uložených dat se musí postupně stát součástí přípravy nových vědců, protože jinak hrozí na jedné straně předčasná ztráta dat (současná situace ve velké většině disciplín), na druhé straně zavalení irelevantními a dále nepoužitelnými daty.

Primární data a procesy *data curation* budou stále významněji ovlivňovat budoucí vědeckou práci. Umožní její lepší kontrolu, přispějí ke zvýšení efektivity prostředků vložených do výzkumu (řada zemí začíná požadovat plné zpřístupnění dat, které byly získány v rámci reali-

<sup>2</sup>Viz. např. časopis *International Journal of Digital Curation*, UKOLN, University of Bath, UK, <http://ijdc.net>

zace projektů hrazených z veřejných rozpočtů) a povedou k rozvoji úplně nových metod vědecké práce (např. *data mining* dlouhých časových řad, jejichž shromáždění jde nad možnosti i zájem jednotlivých úzce zaměřených výzkumných týmů). Vedle otevřeného přístupu k publikacím se tedy musíme připravit i na otevřený přístup k našim primárním datům.

A na závěr jedno ohlédnutí: Protokol *http*, základ světového webu, byl původně vynalezen v CERNu jako technický nástroj pro sdílení publikací, primárních experimentálních dat i výsledků výpočtů co nejširší vědeckou komunitou částicových fyziků. Současný Internet tak za podstatnou část svého úspěchu vděčí potřebě vědecké komunity sdílet data a současně nežádat ochranu intelektuálních práv na software, který sdílení umožňuje.

## Literatura

- [1] N. Beagrie, *Digital Curation for Science, Digital Libraries, and Individuals*, IJDC, Vol 1(1), 3-16, 2006, <http://ijdc.net/index.php/ijdc/article/viewFile/6/2>. □

## Nástroje Google.

### 7. Google Analytics

Roman Dušek, ÚVT MU,

Nikol Kokešová, FI MU

#### 1 Co je to Google Analytics

Google Analytics (zkráceně GA) je dalším zástupcem rozsáhlé rodiny webových aplikací Google. Jeho prostřednictvím je možné měřit návštěvnost webových stránek a zejména zkoumat trendy v chování návštěvníků na sledovaných webových stránkách. Výsledky těchto výzkumů jsou zajímavé zejména pro provozovatele webových stránek, pro webmastery a pro marketingové pracovníky.

Nezbytnou součástí GA je služba, která zaznamenává a archivuje údaje o přístupech návštěvníků k jednotlivým webovým stránkám. Kvalita a spolehlivost měření se několikaletou praxí

osvědčila natolik, že se GA zařadila mezi uznávané nástroje v oblasti měření návštěvnosti.

Google Analytics je bezplatná služba s velkým potenciálem, kterou Google neustále rozvíjí. GA najdeme na <http://www.google.com/analytics/>.

## 2 Analýza návštěvnosti

Cílem analýzy návštěvnosti je získat odpovědi. Analýza návštěvnosti prostřednictvím Google Analytics nám může odpovědět pouze na otázky spjaté se souhrnným chováním návštěvníků našich webových stránek; neumožňuje monitorovat chování konkrétních návštěvníků. Příklad relevantních otázek:

- Kolik návštěvníků navštívilo naše webové stránky minulý týden?
- Jaký je trend v návštěvnosti za poslední měsíc?
- Ve kterém týdnu v roce jsme zobrazili nejvíce stránek?
- Které hypertextové odkazy uživatelé na dané stránce nejčastěji využívají?
- Jaký webový prohlížeč používají naši návštěvníci? Jakou verzi?
- Jaké jsou nejčtenější webové stránky posledního měsíce?
- Kolik návštěvníků se minulý týden zaregistrovalo?
- Na které stránce tráví návštěvníci nejvíce času?
- Kolik % návštěvníků se na naše stránky pravidelně vrací?
- Přes která klíčová slova použitá ve vyhledávači se dostal návštěvník na naši stránku?

GA umožňuje sledovat různé ukazatele související s návštěvníky, s webovými stránkami, s klíčovými slovy apod. Tyto ukazatele postupně v tomto článku představíme.

### 2.1 Způsob měření návštěvnosti

Pokud chcete využívat možnosti, které GA přináší, je potřeba si nechat pro svůj web vygenerovat unikátní měřicí kód a umístit jej do patičky všech stránek, které chcete monitorovat. Po uplynutí několika hodin budete mít k dispozici ve svém účtu první údaje o návštěvnosti.

Měření návštěvnosti je realizováno pomocí kódu Javascriptu, využívajícího cookies. Ke korektní inicializaci měřicího kódu GA a zaslání reportu GA serverům ke zpracování je zapotřebí, aby návštěvníci měli v prohlížeči povolen Javascript a cookies.

Jestliže některý z těchto elementů není povolen nebo nefunguje správně, nebude daný návštěvník v GA reportech zaznamenán. Další situací, která způsobuje nepřesnost v GA reportech, je mazání cookies, ať již uživatelem nebo automaticky po vypnutí prohlížeče. V tomto případě bude návštěvník při každé návštěvě stránky identifikován jako nový unikátní návštěvník, a tudíž příslušné reporty nebudou plně relevantní.

Tyto nepřesnosti nikterak nesnižují sílu tohoto nástroje. Znalost, že návštěvnost dané stránky vzrostla o 40 % po spuštění internetové kampaně je podstatnější než informace o tom, že danou stránku navštívilo přesně 120 uživatelů.

### 2.2 Základní terminologie

Abychom byli schopni analyzovat informace, které nám GA předkládá, je potřeba nejprve porozumět základním pojmům, které se prolínají skrze všechny nástroje GA. Do této množiny patří pojmy jako návštěvník, návštěva, zobrazená stránka nebo poměr návratů.

**Návštěvník** (Visitor) je osoba, kterou GA identifikuje prostřednictvím speciálního cookie souboru. Tento soubor je vygenerován při první návštěvě webové stránky sledovaného serveru. Platnost této identifikace je 2 roky od poslední návštěvy serveru. Pokud by došlo k vymazání cookie souboru s identifikací návštěvníka, nebyl by při příští návštěvě rozpoznán a byla by mu přidělena identifikace nová.

GA klasifikuje návštěvníky jako nové a vracující se. S touto klasifikací úzce souvisí pojem *unikátní návštěvník*. Při opakovaných přístupech návštěvníka na web v rámci různých návštěv je unikátní návštěvník započítáván v rámci daného časového období vždy pouze jednou. V tomto případě se tedy jedná o jednoho unikátního vracujícího se návštěvníka.

**Návštěva** (Visit) je GA definována jako 30minutová interakce mezi prohlížečem a webovou

stránkou. Pokud návštěvník vypne okno prohlížeče nebo je neaktivní více než 30 minut, je stávající návštěva ukončena a při dalším zobrazení stránky nebo pohybu na stránce se započítává návštěva nová. Mechanismus, který stojí v pozadí identifikace návštěv, je založen opět na cookie souborech.

**Zobrazená stránka** (Page view) je načtení stránky v prohlížeči. Ačkoliv je tento údaj podstatný, např. pro marketingové kampaně, pokud ho nebudeme zkoumat v širším kontextu, snižujeme tím jeho význam. Častou chybou začátečníků je soustředění pozornosti na právě jednu metriku, a to na počet zobrazených stránek.

**Poměr návratů** (Bounce rate) je metrika vyjadřující procentuální zastoupení návštěvníků, kteří opustí webovou stránku, aniž by v rámci sledovaného webu navštívili stránku jinou. Čím nižší hodnota poměru návratů, tím vyšší pravděpodobnost, že návštěvníci na dané stránce našli zajímavý relevantní obsah a pokračovali v procházení webu. Stránky s hodnotou poměru návratů blízké se 100 % bychom měli podrobit analýze, abychom zjistili příčinu nízkého zájmu návštěvníků.

### 3 Popis aplikace Google Analytics

Po přihlášení do GA prostřednictvím uživatelského jména a hesla ke službám Google máme k dispozici výčet GA účtů.

#### 3.1 Účty a profily Google Analytics

Účty představují nejvyšší stupeň hierarchie v GA. Jejich úkolem je seskupit množinu souvisejících domén. Např. spravujeme-li firemní webové stránky, je doporučeno založit GA účet s názvem naší firmy a v rámci účtu pak spravovat jednotlivé firemní domény, subdomény apod. U každého účtu můžeme navíc nastavit míru sdílení informací s ostatními GA produkty, a tím si zpřístupnit další analytické nástroje. Druhým stupněm hierarchie jsou tzv. *profily*.

Profilem chápeme konkrétní sledované domény, subdomény nebo námi specifikovanou část webu. V rámci jednoho účtu můžeme spravovat až 50 profilů. Na úrovni profilů můžeme nadefinovat uživatele, kteří budou mít přístup k našim

reportům, nebo nastavit filtry zobrazovaných informací. Ke každému profilu pak GA zpracovává reporty se statistikami.

#### 3.2 Grafické rozhraní reportů

Každý report se obecně skládá z levého a pravého panelu. Levý panel obsahuje 5 hlavních navigačních kategorií a odkazy pro rychlý přístup k nápovědě. Pravá část je tvořena již samotnými statistikami. Tato informační sekce zahrnuje kalendář pro vymezení časového období, sumari-zaci nejpodstatnějších informací a hlavní obsahovou část. GA umožňuje zobrazit data různým způsobem. Standardní forma zobrazení je prostý výpis informací nebo tabulka, ale data lze vizuálně zpřehlednit také pomocí koláčových nebo sloupcových grafů.

První report, který se nám zobrazí po kliknutí na konkrétní profil, je tzv. *řídící panel*, který obsahuje výčet nejdůležitějších přehledů se statistikami o naší stránce. Hlavní výhodou řídicího panelu je jeho přizpůsobitelnost. Přehledy v řídicím panelu můžeme přesouvat pouhým přetažením myši, editovat, přidávat nebo mazat dle našich aktuálních preferencí. Počet panelů nebo jejich uspořádání není nijak limitován.

#### 3.3 Navigační kategorie

Levý panel obsahuje pět hlavních navigačních kategorií (Řídící panel, Návštěvníci, Zdroje provozu, Obsah, Cíle). Na konci článku jsou možnosti těchto kategorií podrobně popsány. Nyní následuje krátký přehled.

V kategorii **Návštěvníci** nalezneme odpovědi na otázky týkající se návštěvníků našich webových stránek. Nabízí široké spektrum přehledů jako je např. poměr nových a vracejících se návštěvníků, délka a hloubka jednotlivých návštěv nebo přehledná vizualizace geografického rozmístění našich návštěvníků.

Druhou neméně důležitou kategorií jsou **Zdroje provozu** s přehledy placených i neplacených klíčových slov a s informacemi, jakým způsobem návštěvníci přichází na naše stránky. Analýzou zdrojů provozu můžeme odhalit např. nedostatky našich internetových kampaní, následně eliminovat některá klíčová slova z placených pozic ve vyhledávači nebo upravit cílové stránky.

Chceme-li zjistit, které jsou nejčastější vstupní stránky našeho webu, jak návštěvníci prochází náš web a z jakých stránek naopak odchází, musíme navštívit kategorii **Obsah**.

Poslední standardní kategorie GA se věnuje **Cílům**. Cíl je aktivita, která souvisí s úspěchem našeho businessu. Pro někoho může cíl představovat přihlášení se k odběru newsletteru, vyplnění ankety nebo stažení pdf souboru s nabídkou našich služeb. V této kategorii lze nalézt např. vizualizaci cesty k cíli nebo míru konverze.

## 4 Příklady užití Google Analytics

GA poskytuje uživatelům řadu analytických nástrojů. Od jednodušších, které umožňují sledovat konkrétní hodnoty nebo trendy různých kritérií v různých časových intervalech, až po vysoce sofistikované nástroje, které pomáhají uživatelům odhalovat závislosti či souvislosti mezi různými ukazateli.

GA využívají z našeho pohledu tři základní skupiny uživatelů – provozovatelé webů, webmasteři a marketingoví pracovníci orientující se na internetovou reklamu. Každá skupina požaduje po GA jiné výstupy, využívá různý mix analytických nástrojů a upřednostňuje odlišný způsob práce. V této kapitole představíme požadavky jednotlivých skupin a způsob, jakým je GA naplňuje.

### 4.1 Provozovatel webu

Cílem provozovatele webu je mít k dispozici aktuální informace o vývoji vybraných ukazatelů webu v čase. Provozovatel webu nemusí umět pracovat s analytickým nástrojem; je pro něj dostačující, aby pravidelně dostával reporty do své e-mailové schránky. Typické otázky provozovatele webu:

- Kolik návštěvníků navštívilo naše webové stránky minulý týden?
- Jaký je trend v návštěvnosti za poslední měsíc?
- Kolik % návštěvníků se na naše stránky pravidelně vrací?

Pro tyto potřeby nabízí GA možnost nechat si zasílat e-mailem v pravidelných intervalech (denních, týdenních, měsíčních, čtvrtletních) výsledky jednotlivých metrik, ukazatelů či vlastní

přehledy. Přílohou e-mailu může být buď pdf dokument, který odráží strukturu stránky v GA nebo dokument ve formátu xml, csv nebo tsv sloužící pro další zpracování.

Jak již bylo zmíněno, je možné si z jednotlivých metrik a ukazatelů sestavit vlastní přehled (custom report). Pro vytvoření přehledu využívá uživatel šablonu, do jejíchž jednotlivých předem vymezených polí přetahuje ze stromu metrik a ukazatelů položky, které chce sledovat. Vytvoření vlastního přehledu je z uživatelského hlediska jednoduché a velice intuitivní.

Automaticky generované pdf dokumenty jsou z hlediska monitoringu návštěvnosti uznávaným informačním zdrojem, a jsou proto často sdíleny či vystavovány přímo na webových stránkách pro potřeby prokázání návštěvnosti daného webu.

### 4.2 Webmaster webu

Cílem webmastera je, kromě správy obsahu a struktury vlastního webu, reflektovat aktuální technologické a uživatelské trendy. Webmaster požaduje nástroj, prostřednictvím kterého může sledovat chování uživatelů na webových stránkách. Na základě této zpětné vazby může webmaster provést úpravy ve struktuře či obsahu webu. Prostřednictvím nástroje může zjistit, jak moc úspěšná daná úprava byla. Typické otázky webmastera:

- Které hypertextové odkazy uživatelé na dané stránce nejčastěji využívají?
- Jaký webový prohlížeč používají naši návštěvníci? Jakou verzi?
- Ze kterých stránek webu uživatelé nejčastěji odchází?
- Jaké jsou nejčtenější webové stránky?
- Na které stránce tráví návštěvníci nejvíce času?
- Kolik % návštěvníků se na naše stránky pravidelně vrací?

GA nabízí řadu předpřipravených pohledů s různými metrikami, se kterými mohou webmasteři pracovat. Jsou rozřazeny do kategorií Návštěvníci a Obsah. Tyto kategorie jsou součástí levého menu GA. Následuje přehled metrik, které je možné v jednotlivých kategoriích nalézt.

Kategorie Návštěvníci:

- geografické umístění návštěvníků (přehled, z jakých měst či států návštěvníci pocházejí),
- trendy návštěvníků (počet návštěv, počet návštěvníků, počet absolutně jedinečných návštěvníků, počet zobrazených stránek, průměrné zobrazení stránek, průměrná doba strávená uživateli na stránkách, počet lidí pravidelně se vracějících na stránky),
- věrnost návštěvníka (počet návratů návštěvníka na web, pravidelnost návratu návštěvníka, průměrná délka návštěvy, počet zobrazených stránek při jedné návštěvě),
- možnosti prohlížeče (typ prohlížeče, verze prohlížeče, typ operačního systému, verze operačního systému, barevná hloubka obrazovky, rozlišení obrazovky, verze Flash playeru, podpora Javy),
- vlastnosti sítě (název hostitele, rychlost připojení).

#### Kategorie Obsah:

- přehled nejnavštěvovanějších stránek,
- nejlepší vstupní stránky (stránky, které uživatel v rámci své návštěvy navštíví jako první),
- hlavní výstupní stránky (stránky, které uživatel v rámci své návštěvy navštíví jako poslední),
- překryvná stránka (viz níže),
- přehled výsledků vyhledávače na stránkách (jak moc je využíván na stránkách fulltextový vyhledávač, jaká klíčová slova návštěvníci webu zadávají).

Zajímavým analytickým nástrojem je tzv. *překryvná stránka*. Při spuštění nástroje se v novém okně webového prohlížeče zobrazí úvodní stránka monitorovaného webu. Každý hypertextový odkaz je rozšířen o ukazatel vypovídající o využití daného odkazu na dané webové stránce pro přechod na jinou webovou stránku téhož webu. Tento nástroj je možné využít jednak pro sledování pohybu uživatelů po webu, ale také pro odhalení nejužívanějších (nebo naopak nejméně užívaných) hypertextových odkazů na stránce.

### 4.3 Marketingový pracovník webu

Jedním z důležitých cílů marketingových pracovníků je přilákat na webové stránky vhodnou

skupinu návštěvníků. Google nabízí několik služeb, prostřednictvím kterých je možné propagovat své webové stránky. Ve všech případech si musí uživatel vymezit návštěvnickou skupinu pomocí množiny klíčových slov, která je pro ně typická. Prostřednictvím přehledu klíčových slov (v záložce Zdroje provozu) má uživatel možnost sledovat klíčová slova, která zadávají návštěvníci do vyhledávačů a následně se prokliknou až na náš web. Kromě samotného přehledu klíčových slov je zde uvedeno, které vyhledávače odkazují na náš web nejvíce návštěvníků. Tyto informace lze využít zejména při plánování reklamní kampaně.

Dalším důležitým přehledem pro marketingové pracovníky je přehled webů a webových stránek, přes které přechází návštěvníci na naše stránky. GA poskytuje v záložce Zdroje provozu přehled odkazujících stránek. Součástí přehledu je také počet takto realizovaných návštěv v průběhu času. Je tedy možné jednoduše sledovat trend návštěvnosti přes daný hypertextový odkaz.

Kromě uvedených analytických přehledů, které slouží zejména jako podklady pro optimalizaci stávající propagační strategie, je možné přes GA sledovat také vývoj reklamních kampaní, které jsou realizovány prostřednictvím Google AdWords.

Jelikož webové stránky často slouží pouze jako nástroj pro dosahování vyšších cílů (například u elektronických obchodů může být tímto cílem registrace nového zákazníka nebo nákup zboží), nabízí GA možnost definovat si své marketingové cíle a poměřovat úspěšnost naplňování cílů (tzv. konverze). Definovat cíl znamená určit cestu po webu, která vede k určité akci návštěvníka (např. k registraci uživatele, koupi zboží, stažení demoverze programu). GA nabízí v záložce Cíle tyto přehledy:

- přehled celkové míry konverze (kolik bylo za daný časový interval úspěšně naplněno cílů),
- procentuální vyjádření míry konverze jednotlivých cílů,
- reverzní cesta k cíli (výčet alternativních cest k cíli),
- hodnota cíle (přepočtení dosažení cíle na peníze),

- vizualizace cesty (tzv. Funnels; jedná se o vizualizaci chování návštěvníků vzhledem k definovaným cílům; je zde možné zjistit, kde se návštěvníci nejčastěji odkloní od předem definované cesty k cíli).

## 5 Závěrem

GA je uznávaný profesionální nástroj, prostřednictvím kterého je možné získávat odpovědi na otázky o chování uživatelů na vybraných webových stránkách. Google tento nástroj neustále rozvíjí a přidává nové analytické nástroje, přehledy a metriky.

Oblast analýzy návštěvnosti vyžaduje znalost základních termínů a vzájemných souvislostí. Bez těchto znalostí je možné chybně nebo

nepřesně interpretovat výsledky, které analytické nástroje poskytují. Proto Google nabízí zájemcům na svých stránkách <http://www.google.com/analytics/education.html> bezplatný on-line kurz Google Analytics, ve kterém se podrobně vysvětlují jednotlivé analytické pojmy, souvislosti a dílčí nástroje. Tento kurz je možné zakončit také zkouškou a získat tak certifikát o absolvování kurzu; tato varianta je ale již zpoplatněna.

GA pomáhá uživatelům pochopit hlubší vzájemné souvislosti a závislosti mezi jednotlivými činiteli webu, které jsou pro každý web unikátní. Takto nabyté zkušenosti lze zpětně zúročit využitím maximálního potenciálu, který nabízí veškeré služby Google, ve prospěch svého webu. □

## Obsah

<b>Open access - otevřený přístup k vědeckým informacím. Úvod do problematiky, Miroslav Bartošek, ÚVT MU .....</b>	<b>1</b>
<b>Licence Creative Commons a jejich využití v OA, Lukáš Gruber, Národní knihovna ČR .....</b>	<b>8</b>
<b>Služba SHERPA/RoMEO - další život vědeckého článku, Pavla Rygelová, VŠB-TU Ostrava .....</b>	<b>12</b>
<b>Ukládání a sdílení vědeckých dat, Luděk Matyska, ÚVT MU .....</b>	<b>17</b>
<b>Nástroje Google. 7. Google Analytics, Roman Dušek, ÚVT MU, Nikol Kokešová, FI MU .....</b>	<b>19</b>

