

Institucionální repozitář a role univerzitní knihovny

Prochásková, Iva

Univerzita Pardubice

Úvod

Ve Slovníku spisovného jazyka českého [Slovník, 1989] najdeme: *repositář* – z *lat. schránka, skříň, police apod. na úschovu knih, listin apod.* Heslo je tam dále klasifikováno jako slovo zastaralé, které již vyšlo z užívání. Termín teď zažívá slavný návrat. Samozřejmě se „z“. Stalo se označením pro systém, v němž jsou uloženy informace v digitální podobě.

Institucionální repozitáře (IR) na univerzitě mohou mít řadu podob. Mohou to být pouhá úložiště administrativních dokumentů určená pouze pro vnitřní potřebu instituce, jejichž účelem je zabezpečit především archivaci. Mohou to být úložiště pro dokumenty a další digitální objekty sloužící k výuce, které jsou určeny omezené komunitě uživatelů, a jejichž základním rysem je opakované využívání. Tyto typy repozitářů se však zásadně liší od repozitáře určeného pro zveřejňování a dlouhodobé ukládání výsledků vědecko-výzkumné práce, a to nejen druhem ukládaných dokumentů a účelem, ale také tím, že jejich softwarová platforma není tak standardizovaná a metadata často nelze získávat prostřednictvím protokolu OAI-PMH.

Předmětem tohoto příspěvku jsou právě jen repozitáře definované jako elektronický systém, který zachycuje, uchovává a zpřístupňuje výsledky vědecko-výzkumné práce [FOSTER, GIBBONS, 2005].

Proč budovat IR?

Protože IR je příležitost! Je to příležitost pro akademické pracovníky, kteří v něm mohou uplatnit svou publikační činnost a zpřístupnit tak výsledky své práce širokému okruhu vědců na celém světě. Je to příležitost pro univerzitu jako celek. Ta může jeho prostřednictvím zvýšit svůj kredit a konkurenceschopnost na poli vědecko-výzkumném a vzdělávacím a přispět zásadním způsobem ke změnám ve vědecké komunikaci směrem k otevřenému přístupu. Je to příležitost pro poskytovatele dotací, kterým umožní jednoduše sledovat, jakých výsledků bylo za jejich finanční prostředky dosaženo. A je to rovněž příležitost pro knihovnu, jejíž význam začíná být v digitálním věku zpochybňován. Dává jí možnost se zviditelnit, být nejen tím, kdo sbírá informace zvenčí a zprostředkovává je svým uživatelům, ale stát se tím, kdo informace produkované vlastní institucí zpracuje a zpřístupní světu. Je to šance získat v instituci nové a významné postavení. A mělo by být bytostným zájmem každé univerzitní knihovny tuto šanci nepromarnit.

Na univerzitní repozitář lze také pohlížet jako na soubor služeb, které univerzita nabízí členům své komunity pro správu a šíření vlastních elektronických dokumentů [LYNCH, 2003]. Tyto služby mohou být na různých univerzitách zajišťovány různými organizačními útvary, nicméně existuje několik dobrých důvodů pro to, aby se na tom velmi významnou měrou podílela knihovna. Knihovníci umí dokumenty standardizovaným způsobem zpracovat, pořádat a ukládat a zároveň zajistit jejich trvalou přístupnost. Jsou v kontaktu s akademickými pracovníky – autory, protože publikace potřebné pro vědeckou práci pro ně zprostředkovávají. Rychle vstřebávají nové technologie a jsou zvyklí pořádat pro uživatele informační školení.

Existují zde samozřejmě i určitá rizika. Jedním z nich může být konzervatismus knihovníků, kteří rádi aplikují nové technologie na zavedené postupy místo realizace tvůrčích změn v postupech pomocí moderních technologií. Dalším rizikem může být například nedostatek přesvědčivých argumentů při vysvětlování výhod IR autorům a někdy i vedení instituce.

Úvod do budování IR

Základní atributy charakterizující IR jsou: vazba na instituci, otevřený přístup, interoperabilita, standardizace a různorodost obsahu. Z dostupné zahraniční literatury o IR vyplývá, že je to často právě univerzitní knihovna, která je na univerzitě iniciátorem budování repozitáře. Od myšlenky k činu je však právě v tomto případě daleká cesta a je nezbytné být na ni velmi dobře teoreticky i prakticky připraven.

Autoři knihy *The Institutional Repository* [JONES, 2006] uvádějí:

...je několik zásadních věcí, které je třeba důkladně promyslet a rozhodnout ještě před vlastním založením repozitáře. Za nejdůležitější považujeme tyto tři: rozhodnutí o struktuře repozitáře, stanovení formátu a úrovně uchovávaných dat a vyřešení otázky metadat. Pokud se později některé z těchto rozhodnutí ukáže jako špatné, může to ohrozit úspěch celého projektu.

Při stanovení struktury IR je třeba vycházet z dobré znalosti lokálního prostředí. Není možné obecně přejímat příklady z jiných univerzit. Samozřejmě by měla být spolupráce se všemi součástmi univerzity, které budou, třeba i v poměrně daleké budoucnosti, informace do IR ukládat.

Formáty pro ukládání dat je třeba volit zvláště uvážlivě s přihlédnutím k tomu, že bude třeba zajistit dlouhodobé (trvalé) uchování obsahu a zároveň jeho zpřístupňování. Lze předpokládat, že tento požadavek lépe splní formáty založené na otevřených standardech. Co se týče úrovně uchovávaných dat, je třeba rozhodnout, zda budou do IR ukládána také pouze metadata bez připojených souborů s plnými texty. Na základě tohoto rozhodnutí je pak nutné někdy změnit v knihovně celou koncepci ukládání sekundárních informací.

Metadata se stanou knihovníkovým denním chlebem. I když se předpokládá, že základní metadata budou vkládat sami autoři a na knihovníkovi zůstane jen kontrola, opravy a doplnění chybějících polí, tj. zajištění kvality, právě to může být zpočátku problém. Je velmi nebezpečné nechat repozitář zahlit spoustou nekvalitních metadat, která knihovníci nestačí průběžně opravovat. Cesty řešení mohou být různé: od intenzivního školení autorů, přes implementaci různých pomůcek a automatických kontrolních mechanismů, které sníží možnost chyb, až po zaměstnání dostatečného počtu vyškolených katalogizátorů.

Technické řešení

Za neméně důležité rozhodnutí považuji volbu software. Účelem tohoto příspěvku není podrobně popsat techniky a technologie používané pro IR, ale zmínit základní rysy, které by měl mít zvolený software. Výběr je již poměrně široký, a to jak v komerční sféře, tak v oblasti open source software (OSS). Autoři výše zmíněné knihy se vyslovují ve prospěch použití OSS právě pro IR budované na univerzitách. Zdůrazňují, že se jedná o software vyvíjený a podporovaný silnou aktivní komunitou, který respektuje specifické potřeby univerzitního prostředí. U tohoto typu software je také předpoklad trvalého dynamického rozvoje a nezanedbatelné jsou samozřejmě také nulové pořizovací náklady. Příloha knihy obsahuje popis šesti typů OSS.

IR se začíná budovat v momentě, kdy již na univerzitě existuje určitá informační infrastruktura. Nemělo by tedy být příliš složité nový systém do ní integrovat. Důležitá je modulární architektura, která umožní snadnou lokalizaci, a striktní oddělení aplikačních vrstev umožňující propojení jednotlivých komponent s dalšími systémy a vytvoření rozhraní pro web. Bezproblémové by mělo být propojení s lokálním autentikačním systémem. Výhodou je, když software má svůj vlastní nástroj pro řízení autorizovaného přístupu. Důležité je, aby software obsahoval administrativní nástroje pro snadnou a rychlou údržbu, zajišťoval bezpečnost a integritu dat a podporoval použití persistentních identifikátorů.

Vzhledem k obsahu budoucího IR, který může být značně různorodý, je výhodné flexibilní metadatové schéma. Systém může mít své vlastní schéma, v každém případě je však nutné, aby bylo transformovatelné do formátu kompatibilního s nekvalifikovaným Dublin Core, který je podporovaným formátem pro službu OAI-PMH. Pokud uvážíme, že do IR budou vkládat metadata uživatelé a nikoli zkušení katalogizátoři, je vhodná u některých polí autoritní kontrola a předem nastavené hodnoty. Systém musí také umožňovat snadnou kontrolu a opravu dat kvalifikovaným knihovníkem.

Obhajoba účelnosti

Aby IR plnil své základní funkce, musí být plněn věrohodným obsahem. Ukazuje se však, že toto je jeden z nejtěžších úkolů, před kterým „budovatelé“ stojí. Mnozí akademičtí pracovníci nejsou přesvědčeni o tom, že je publikování v univerzitním repozitáři s otevřeným přístupem prospěšné jak pro komunikaci vědeckých poznatků obecně, tak pro ně osobně. Pracovníci knihovny se musí pokusit nalézt správné argumenty (příklady volného vědeckého publikování v různých disciplínách, zvýšení citovanosti dokumentů uložených v IR s otevřeným přístupem, dlouhodobá archivace) a srozumitelným způsobem je předložit. Řada autorů také například neví, kteří vydavatelé, a v jaké fázi publikování dokumentu, umožňují autorům vlastní archivaci [SHERPA, 2006]. Velmi podnětná studie k problému „přijetí“ IR akademickou obcí byla provedena na University of Rochester [FOSTER, GIBBONS, 2005]. Autoři studie provedli rozsáhlá šetření, aby poznali potřeby vědeckých pracovníků v souvislosti s procesem tvorby výstupů vědecké práce a publikování. Přiblížili jim účel a funkci repozitáře srozumitelnou řečí (při komunikaci s nimi totiž zjistili, že pod řadou odborných termínů běžně používaných v knihovnictví si nedokáží nic konkrétního představit) a na základě získaných poznatků upravili použitý software (DSpace) a postupy pro vkládání dokumentů do IR.

Výše uvedené zdaleka nepostihuje všechny aspekty, které je třeba při budování IR předem promyslet. Mnohem více lze nalézt v již zmíněné knize [JONES, 2006], kterou je možno považovat za skutečnou „kuchařku“. Další poučení lze získat, podíváme-li se na již realizované projekty na řadě evropských i světových univerzit.

Repozitáře zemích Evropské unie

IR začaly vznikat na univerzitách ve vyspělých zemích po roce 2000. Ve světovém registru ROAR (Registry of Open Access Repositories) je jich registrováno již více než 840 (únor 2007). Existuje také adresář OpenDOAR budovaný v rámci projektu SHERPA. Nejvíce IR v Evropě je ve Velké Británii a v Německu.

V rámci Šestáho rámcového programu Evropské unie byl 1. června 2006 zahájen projekt DRIVER (Digital Repository Infrastructure Vision for European Research). Jednou z prvních aktivit tohoto projektu byla souhrnná studie o stavu digitálních repozitářů. Repozitář byl pro účely této studie definován takto: (1) obsahuje výsledky vědeckého výzkumu, (2) je institucionální nebo tématický a (3) umožňuje sklizení dat pomocí OAI-PMH. Zjišťování bylo provedeno pomocí poměrně rozsáhlého dotazníku. Dotazník byl rozeslán na 248 e-mailových adres získaných dílem z adresáře OpenDOAR a dílem dalším pomocným zjišťováním v zemích EU, které nemají repozitáře v OpenDOAR registrovány. V prvním kole se vrátilo 95 vyplněných dotazníků a 18 institucí odpovědělo, že IR

nebudují. Tyto výsledky byly předběžně zpracovány a zveřejněny [DRIVER, 2006]. Na základě toho, bylo vyplněno ještě dalších 19 dotazníků. Pracovní verze studie byla zveřejněna počátkem února 2007. Výsledky vycházejí ze 114 vyplněných dotazníků ze 17 zemí EU.

Výsledky studie

Obsah

Většina obsahuje plné texty různých typů textových materiálů (především časopisecké články, vysokoškolské kvalifikační práce, knihy nebo kapitoly z knih, working papers a proceedings). V menší části repozitářů se objevují pouze metadata textových materiálů. Netextové materiály jsou uloženy ojedinelé. Zajímavé je, že články jsou častěji ukládány jako publikované verze nebo postprinty než preprinty. Z hlediska oborového jsou nejvíce zastoupeny humanitní a společenské vědy, následují přírodní vědy a inženýrské obory.

Institucionální politika a způsob ukládání

Ukládání materiálů do repozitáře je ve více než polovině institucí dobrovolné, i když ve třetině z nich jsou k tomu autoři důsledně pobízeni. Ve čtvrtině institucí je ukládání povinné, většinou se jedná o povinné ukládání disertačních prací. Zbývající čtvrtina nemá ukládání nijak oficiálně upraveno. Do 28% repozitářů vkládají materiály sami autoři a knihovníci provádějí pouze kontrolu. Ve 26% autoři materiály zasílají, ale ukládají je knihovníci. V 7% případech plní repozitář pouze knihovníci bez účasti autorů. Ve zbylých případech se jedná o kombinaci všech předchozích metod.

Software a další technické údaje

Nejvíce používán je GNU Eprints (24%) a DSpace (20%). V 17% jsou používány lokálně vyvinuté programy. V 75% repozitářů jsou připojeny persistentní identifikátory ke každému dokumentu a u 70% jsou k dispozici statistická data o počtu přístupů a využití jednotlivých materiálů. Více než polovina repozitářů používá jako metadatový standard kvalifikovaný Dublin Core a více než třetina nekvalifikovaný Dublin Core. Většinu repozitářů lze vyhledat pomocí obvyklých vyhledávačů (Google, Yahoo, MSN) a prostřednictvím OAIster a Google Scholar.

Překážky dalšího rozvoje

Za největší překážky dalšího rozvoje repozitářů byly označeny: (1) v instituci není stanoveno povinné ukládání, (2) copyright, otázky s ním spojené a nedostatečné povědomí autorů o tomto problému, (3) grantové agentury nevyžadující povinné ukládání výsledků výzkumu.

Shrnutí

Z výsledků studie vyplynulo sedm faktorů, které se jeví jako nejdůležitější pro další rozvoj digitálních repozitářů obsahujících výsledky vědeckého výzkumu v EU. Tyto faktory budou využity jako základ programu aktivit na evropské úrovni v souladu s projektem DRIVER. Jedná se o: (1) větší zviditelnění repozitářů zlepšením jejich vyhledatelnosti, tj. přístupnosti pro vyhledávací služby a zdokonalením předmětového indexování, (2) harmonizaci procesů ukládání, (3) povinné ukládání, (4) otázky copyrightu a s nimi spojenou flexibilitu ve formě přístupu k uloženým materiálům, (5) aktivity podporující zájem o budování repozitářů ze strany řídicích složek univerzit a vysvětlující kampaně mezi akademickou obcí, (6) přidané služby – např. citační a (7) technickou harmonizaci.

Pozice České republiky

Byly odevzdány čtyři vyplněné dotazníky a na základě toho jsme byli zařazeni do zemí nacházejících se v „raném stádiu“. (Slovensko bylo zařazeno mezi země, které digitální repozitář dosud nebudují). Nemáme k dispozici žádnou studii, která by stav IR na českých univerzitách mapovala, nicméně z kontaktů s kolegy knihovníky z českých univerzit vím, že to je u nás skutečně ještě pole neorané. Ze čtyřech nahlášených odpovídá parametrům pro IR pouze repozitář vytvořený a provozovaný Ústřední knihovnou VŠB – TUO. V ostatních případech se jedná o úložiště vysokoškolských kvalifikačních prací. Možná existují někde další repozitáře, které tato studie nezachytila, protože si nikdo nedal práci s vyplněním dotazníku (dotazník jsem rozeslala na kontaktní adresy do knihoven všech veřejných vysokých škol s prosbou o předání kompetentní osobě, pokud knihovna sama IR nebuduje). Zařazení do kategorie „rané stádium“ tedy skutečností neodpovídá, patřili bychom spíše do kategorie „na startu“. Bude určitě ještě nějakou dobu trvat, než budou v České republice funkční IR, které bude možno s plnou zodpovědností registrovat ve světových registrech.

Závěr

Před vysokoškolskými knihovníky v České a Slovenské republice (ale podobně jsou na tom i v Estonsku, Irsku, Portugalsku, Polsku nebo Španělsku) je tedy nový velký úkol. Možná je pro nás o něco těžší v tom smyslu, že postavení knihovníka na vysoké škole není u nás tak silné a prestižní, jako tomu je v západních zemích a zvláště v USA. Nicméně pustit se do toho budeme muset.

Bibliografické odkazy

1. DRIVER inventory study digital repositories in EU countries [online]. Pleiade, 2006. [cit. 2007-02-26]. Dostupný na [www: http://www.pleiade.nl/wiki/index.php?title=Main_Page](http://www.pleiade.nl/wiki/index.php?title=Main_Page).

2. FOSTER, N. F., GIBBONS, S. Understanding faculty to improve content recruitment for institutional repositories. D-Lib Magazine [online]. 2005, vol. 11, no. 1. [cit. 2007-02-23]. Dostupný na www: <http://www.dlib.org/dlib/january05/foster/01foster.html>.
3. LYNCH, C. A. Institutional repositories : essential infrastructure for scholarship in the digital age. Portal : Libraries and the Academy [online]. 2003, vol. 3, no. 2. [cit. 2007-02-26]. Dostupný na www: <http://proquest.umi.com/pqdweb?did=350571071&sid=1&Fmt=3&clientId=45144&RQT=309&VName=PQD>.
4. JONES, R., ANDREW, T., MacCOLL, J. The Institutional repository. Oxford : Chandos Publishing, 2006. ISBN 1 84334 138 7.
5. SHERPA : RoMEO databases [online]. University of Nottingham, 2006. [cit. 2007-02-26]. Dostupný na www: <http://www.sherpa.ac.uk/romeo.php>.
6. Slovník spisovného jazyka českého. 2. nezm. vyd. Praha : Academia, 1989.