

Zpráva o české komunitě DSpace

Speech about Czech DSpace User Group

Mgr. Pavla Rygelová, Ústřední knihovna VŠB-TU Ostrava

Mgr. Vlastimil Krejčíř, Ústav výpočetní techniky MU

Příspěvek přináší základní informace o dění kolem systému *DSpace* v České republice. Zvláštní pozornost bude věnována dvěma uskutečněným ročníkům Setkání českých uživatelů systému *DSpace* a bude nastíněna představa o náplni příštího setkání.

Autoři příspěvku představí jednotlivé instalace systému *DSpace* na VŠB – Technické univerzitě Ostrava, Univerzitě Pardubice, Českém vysokém učení technickém v Praze, na Akademii múzických umění v Praze a na Masarykově univerzitě. Software *DSpace* využívají univerzity především pro budování repozitáře vysokoškolských kvalifikačních prací.

Závěrem shrnou přednosti tohoto systému, základní požadavky na jeho provoz a předají dostupné informace o budoucím vývoji systému *DSpace*.

The paper brings basic information about the activities of Czech *DSpace* user group. Special attention will be paid to the last two Czech *DSpace* User Group Meetings and the concept of the next meeting will be also presented.

Speakers will introduce the most significant installations of software *DSpace* in the Czech Republic. *DSpace* is used by VSB – Technical University of Ostrava, University of Pardubice, Czech Technical University in Prague, Academy of Performing Arts in Prague and Masaryk University. These universities have built repositories on *DSpace* platform which are used for presentation and archiving of electronic theses and dissertations.

At the end of the paper speakers will sum up the advantages of using *DSpace*, basic requirements for its installation and operation, and finally they will give information about future development of the *DSpace* software.

Úvod

DSpace patří v České republice k nejrozšířenějším otevřeným softwarům pro budování digitálních repozitářů. Vysoké školy využívají *DSpace* především pro řešení otázek zpřístupňování a archivování vysokoškolských kvalifikačních prací (dále jen VŠKP). S tím také souvisí fakt, že většina instalací *DSpace* spolupracuje různým způsobem se studijními informačními systémy, které využívá mj. jako informační zdroj pro metadatový popis VŠKP. *DSpace* poskytuje kvalitní uživatelské rozhraní pro vyhledávání a zpřístupňování VŠKP a zároveň zajišťuje mechanismy pro dlouhodobou archivaci dat. *DSpace* je oblíben v univerzitním prostředí, protože disponuje základními stavebními prvky repozitářů, podporuje metadatové standardy (Dublin Core, MODS), perzistentní identifikátory systému *Handle.net*, protokoly OAI-PMH a OAI-ORE, fulltextové vyhledávání aj., ale také vyhovuje ve způsobu, jak pracuje s uloženým digitálním obsahem. *DSpace* si digitální dokumenty organizuje do volitelně zanořených celků, které kopírují organizační strukturu instituce. Nejvyšší složky tvoří komunity (představující většinou jednotlivé fakulty nebo celoškolské ústavy), které se mohou dále členit na podkomunity (zastupující katedry, instituty apod.),

v rámci komunit a podkomunit se budují kolekce, do nichž je možné vkládat digitální dokumenty. (Samozřejmě není vyloučeno organizování digitálního obsahu podle časových, tematických kritérií, podle typu dokumentů aj.) Tato struktura napomáhá k rychlé orientaci v obsahu repozitáře.

Na Masarykově univerzitě a VŠB – Technické univerzitě vzniklo k tématu digitálních repozitářů několik diplomových a bakalářských prací. Za všechny jmenujeme diplomovou práci Mgr. Vlastimila Krejčíře¹, který srovnával několik softwarů pro budování repozitářů, a diplomovou práci Ing. Dušana Jalůvky², v níž se zaměřil na praktické zprovoznění softwaru *DSpace*. Ing. Jalůvka připravil v rámci své diplomové práce i českou lokalizaci *DSpace*, která však vyžaduje důkladnější revizi. Zatím si každá instituce upravuje lokalizaci podle vlastních potřeb.

V roce 2009 funguje sedm instalací *DSpace* na šesti českých univerzitách; na Masarykově univerzitě (MU) – dvě instalace, Vysoké škole báňské – Technické univerzitě Ostrava (VŠB-TUO), Univerzitě Pardubice (UPCE), Akademii múzických umění v Praze (AMU), Českém vysokém učení technickém v Praze (ČVUT) a na Univerzitě Jana Evangelisty Purkyně v Ústí nad Labem (UJEP). Masarykova univerzita, resp. její Ústav výpočetní techniky, se podílí na tvorbě *České digitální matematické knihovny* (DML-CZ) a pro účely testování si buduje repozitář studentských esejí, digitálních videonahrávek a fotografií z univerzitního archivu. Ostatní školy využívají *DSpace* pro již výše zmíněné řešení problematiky elektronických VŠKP. Instalace na ČVUT a UJEP jsou dostupné pouze v intranetu a *DSpace* na AMU prochází stále testovacím provozem.

Nejdéle funguje *DSpace* na VŠB-TUO. Repozitář spravuje Ústřední knihovna od roku 2006 a ukládá do něj informace o VŠKP, článcích ze *Sborníku VŠB-TUO* a publikační činnosti akademických pracovníků univerzity. Od roku 2008 jsou do repozitáře ukládány plné texty některých VŠKP (830 souborů za rok 2008 a cca 2000 souborů za rok 2009), články ze sborníku jsou skenovány a připojovány k bibliografickým záznamům. Publikační aktivita akademických pracovníků je sledována zatím pouze na bibliografické úrovni. Na základě údajů z databáze *Web of Science* jsou do repozitáři vkládány záznamy článků z impaktovaných časopisů. Přístup k veškerým digitálním dokumentům je omezen na registrované uživatele z VŠB-TUO. Dalším záměrem Ústřední knihovny VŠB-TUO je propagovat na univerzitě princip otevřeného přístupu (*open access*), jmenovitě princip (auto)archivace eprintů vědeckých článků a repozitář *DSpace* využít jako technickou podporu.

Podobné záměry sleduje i Digitální knihovna Pardubice, spravovaná také tamější Univerzitní knihovnou. V digitální knihovně jsou již nyní volně přístupné VŠKP a přibudou také články z univerzitního sborníku. Zpřístupňování VŠKP se řídí podle licenčního ujednání

¹ KŘEJČÍŘ, Vlastimil. *Univerzální digitální repozitář*. Brno, 2005. 110 s. Diplomová práce (Mgr.). Masarykova univerzita, Fakulta informatiky.

² JALŮVKA, Dušan. *Úložiště digitálních dat pro potřeby ÚK VŠB-TU Ostrava I*. Ostrava, 2006. 64 s. Diplomová práce (Ing.). Vysoká škola báňská – Technická univerzita Ostrava, Fakulta elektrotechniky a informatiky.

uzavřeného mezi autorem práce a univerzitou. S novou směrnicí se však od této praxe ustupuje a volná přístupnost bude standardem.

Za třetí plně funkční instalaci DSpace můžeme označit Českou digitální matematickou knihovnu (DML-CZ). Projekt DML-CZ vznikl kooperací několika institucí a byl finančně podpořen Akademií věd ČR. DML-CZ obsahuje naskenované články z českých matematických časopisů, několika sborníků a matematických monografií. Obsah DML-CZ můžeme považovat za volně dostupný, omezený přístup se týká dokumentů novějšího data, které budou zpřístupněny po uplynutí určité doby.

Setkání českých uživatelů systému *DSpace*

Na podzim roku 2007 v rámci konference *Bibliotheca academica* v Ostravě proběhla neformální schůzka několika zástupců knihoven, kteří již pracovali se systémem *DSpace* a dohodli se na uspořádání *1. setkání českých uživatelů systému DSpace*, zaměřeného na výměnu praktických zkušeností s tímto softwarem. 1. setkání proběhlo v roce 2008, následovalo 2. setkání v roce 2009. Obě akce se uskutečnily v Ústřední knihovně VŠB-TU Ostrava. Záměrem setkání bylo předávání užitečných informací o implementaci a provozu *DSpace*. Do programu se vždy aktivně zapojily všechny instituce využívající *DSpace*. Příspěvky vycházely z konkrétních zkušeností a informovaly o instalaci, provozu a úpravách systému *DSpace*, provedených v jednotlivých institucích. Prezentace z obou ročníků jsou dostupné v repozitáři *DSpace VŠB-TUO*. V čísle 5/2009 časopisu *Ikaros* vyšla podrobná zpráva o druhém setkání³.

V současné době se uvažuje nad rozšířením obsahového zaměření setkání, neomezovaného na jedinou softwarovou platformu. Rozšíření by se mělo nést v duchu záměrů ÚK VŠB-TUO prosadit na univerzitě repozitář jako přirozené prostředí pro vědeckou komunikaci a zvýšit povědomí o modelu otevřeného přístupu k vědecké publikované literatuře. Setkání by zahrnovalo jak část technickou, tak část věnující se strategii či politické stránce prosazování tohoto principu. Setkání bude orientované na praktickou stránku věci.

Webová podpora *DSpace CZ*

Od roku 2007 jsou udržovány webové stránky *DSpace CZ*⁴. S přispěním *Asociace knihoven vysokých škol ČR* stránky získaly doménu *dspace.cz*. V roce 2009 došlo k drobným grafickým úpravám a rozšíření obsahové náplně. Stránky nekopírují oficiální web komunity *DSpace*⁵, ale přinášejí základní informace pro zájemce z českého prostředí. Web je členěn do několika sekcí. Oddíl *Novinky* přináší informace o vydání nové verze *DSpace*, o pořádaných konferencích v zahraničí a setkáních uživatelů v Česku. Zveřejněny jsou kontakty a odkazy na jednotlivé instalace *DSpace* a tři soubory s českými lokalizacemi, které vznikly na VŠB-TUO, UPCE a UJEP. Užitečná je sekce *Dokumenty* s výběrovým přehledem prezentací a dalších dokumentů o systému *DSpace*. Zatím rozpracovaná zůstává část *Odkazy*, která by měla

³ <http://www.ikaros.cz/node/5431>

⁴ <http://www.DSpace.cz/index.html>

⁵ <http://www.dspace.org>

sloužit jako rozcestník pro další informace týkající se nejen software *DSpace*, ale také dalších otázek spojených s budováním repozitářů.

Sklizení českých repozitářů

Sekce *Odkazy* na webu *DSpace.cz* zahrnuje také nejvýznamnější registry a *service providers* sklizející obsahy repozitářů. K registraci nejčastěji slouží *OAI base URL* repozitáře. Následující tabulka zobrazuje přehled registrovaných českých repozitářů a digitálních knihoven. V registrech je indexován také jeden časopis vydávaný Ústavem mechaniky na Fakultě strojní ČVUT a sborník z konference Inforum. Registrace v *OAI data providers* by

Název repozitáře (poskytovatele dat)/Název registru (poskytovatele služeb)	DSpace	DSpace wiki	OAI data pr.	University Illinious OAI data providers registry	OAIster	Scientific Commons	ROAR	OpenDOAR	BASE
Bulletin of Applied Mechanics					✓	✓			✓
dKNAV			✓	✓		✓	✓		✓
Digitální knihovna NK Kramerius			✓	✓		✓			
Digitální knihovna UPCE	✓	✓					✓	✓	✓
DML-CZ	✓						✓		
DSpace AMU	✓								
DSpace MU		✓							
DSpace VŠB-TUO	✓	✓					✓	✓	✓
Katalog Vědecké knihovny Olomouc			✓	✓					
Sborník z konference INFORUM								✓	

měla dostačovat pro sklizení poskytovateli služeb *OAIster* a *ScientificCommons*.

Tab. 1 Poskytovatelé dat z ČR registrovaní a sklizení poskytovateli služeb

Registrace přináší větší viditelnost repozitáře, který se stává dostupnější pro uživatele doslova z celého světa. O přístupech se lze přesvědčit pomocí statistik *Google Analytics*. Žebříček *Ranking Web of World Repositories*⁶ čerpá z registrů ROAR a OpenDOAR údaje pro sestavování tabulek nejúspěšnějších repozitářů. *DSpace VŠB-TUO* figuruje jako jediný z českých repozitářů na solidním 233 místě v tabulce *TOP 400 Institutional Repositories*. Metoda měření úspěšnosti vychází z kombinace čtyř indikátorů, jejichž hodnoty jsou do celkového výpočtu zahrnuty v určitém procentním poměru.

Velikost	počet stránek pokrytých čtyřmi největšími vyhledávači Google, Yahoo, Live Search, Exalead
Viditelnost	celkový počet jedinečných externích odkazů získaných z Yahoo, Exalead
Počet souborů	pouze počet textových souborů (pdf) získaných z Google a Yahoo
Google Scholar	poměr celkového počtu článků a nových z let 2001-2008

Tab. 2 Webové indikátory pro výpočet úspěšnosti repozitáře

⁶ <http://repositories.webometrics.info>

Největší váha (50 procent) je přikládána indikátoru viditelnosti repozitáře (*visibility*), pak následuje indikátor velikost repozitáře (*size*) s 25 procenty a indikátory počet souborů (*rich files*) a Google Scholar (*scholar*) po 15 procentech. Do rankingu jsou počítány jen repozitáře s vlastní doménou nebo poddoménou.

Poznámky k instalaci a provozu

Technické požadavky na provoz *DSpace* jsou podrobně shrnuty přímo na stránkách *dspace.org*, kde si je může každý zájemce přečíst. Software jako takový je nabízen jako *open source* a je zdarma. Na tom místě je důležité si uvědomit, že zdarma je pouze samotný kód programu (a samozřejmě související dokumentace a návody). Distribuce *DSpace* nemá klasickou zákaznickou podporu, ale výborně fungují elektronické diskusní skupiny (*DSpace-general*, *DSpace-technical*, *DSpace-development*), systém hlášení chyb a systém pro spolupráci při vývoji (platforma *JIRA*). Celosvětová komunita *DSpace* je natolik velká (existuje přes 700 instalací), že reakce na hlášené chyby jsou velice rychlé, a problémy jsou odstraňovány prakticky okamžitě. (V zahraničí již fungují i komerční projekty, které nabízejí instalaci a správu *DSpace* např. @mire v Belgii a AePIC v Itálii).

Veškeré náklady na instalaci, údržbu a provoz systému si tedy instituce nese sama. Vyžaduje to především systémového administrátora – programátora se znalostí jazyka *JAVA*, administrace databáze *SQL*, (*PostgreSQL*, *Oracle*) a webového serveru, který umí zpracovat servlety, např. *Apache Tomcat*. Vzhledem k použité technologii je možné *DSpace* provozovat na operačních systémech unixového typu i na platformě *Microsoft Windows*. Instalace a konfigurace celého systému je poměrně nenáročná a lze ji se základním nastavením provést během jednoho dne (dokumentace k *DSpace* je dostatečně kvalitní a podrobná).

Jak již bylo zmíněno, *DSpace* je software pro budování univerzálního repozitáře, tedy pravděpodobně nebude úplně přesně vyhovovat všem požadavkům dané instituce, která jej nasadí, a je pravděpodobné, že bude nutné *DSpace* mírně upravovat dle konkrétních požadavků. Na mnohé změny a úpravy je již *DSpace* připraven, umožňuje snadno měnit design, konfigurovat metadatový formát pro popis objektů apod., jiné funkcionality je nutné doprogramovat. S tím *DSpace* také počítá, neboť je budován ve vrstvách, z nichž každá má přesně zdokumentované *API*, kterého může každý programátor podle potřeb využít. Samozřejmě je možné se zapojit do samotného vývoje *DSpace* a prosadit vlastní změny a rozšíření do oficiálního vydání *DSpace*.

Uživatelé, kteří by chtěli využít jen některých funkcí *DSpace*, např. repozitáře samotného a nad ním si vybudovat vlastní rozhraní s vlastní funkcionalitou, mají k dispozici systém *Manakin*. Ten umožňuje kompletně změnit celý vzhled včetně funkcionality bez přímého zásahu do kódu *DSpace* a to pouhým rozšiřováním již připravených šablon (na bázi *XML* a *XSLT*). Na systému *Manakin* je například postavena celá *DML-CZ*.

Vývoj a budoucnost *DSpace*

Velkou výhodou používání hotového software je to, že budoucí vývoj (implementace nových standardů, opravy chyb atp.) leží na bedrech třetí strany. Je možné říci, že komunita *DSpace* je v současnosti natolik velká a silná, že budoucnost *DSpace* je zajištěna na léta dopředu, zárukou je i to, že za *DSpace* stojí *MIT*, *Hewlett-Packard* aj.). Vývoji *DSpace* se nyní plně věnuje více než desítky programátorů a další desítky přispívají drobnějšími úpravami. V

rámci konference *Open Repositories* dochází i k setkávání vývojářů *DSpace* s konkurencí – vývojáři systémů *Fedora* a *EPrints*. Poměrně nedávno pak vznikla skupina *DuraSpace*, která se věnuje společným prvkům *DSpace* a *Fedora* a bude na jejich vývoji spolupracovat.

Díky rozsáhlosti komunity kolem *DSpace* je k dispozici i celá řada rozšíření, která pro své potřeby již někdo implementoval a dal komunitě k dispozici. Často jsou to natolik rozsáhlé úpravy, že zásadně mění *DSpace* nebo mu přidávají rozsáhlou novou funkcionalitu. Můžeme zmínit systém *Tapir* (rozšíření *DSpace* o podporu pro vysokoškolské práce v UK), statistický modul univerzity v Minho (umožňuje podrobně sledovat provoz v repozitáři, vyhodnocovat jej, vytvářet grafy apod.), *Dpubs* (buduje nad *DSpace* publikační systém) a mnoho dalších.

Paralelně se pracuje na dvou vývojových větvích *DSpace* – základní větev se věnuje inovacím a opravám stabilní verze *DSpace* (1.x, nyní verze 1.5). Vedle ní probíhá živý vývoj verze 2.0, která by měla mít plně zapracovanou celou řadu nových funkcí a prvků. Verze 2.0 by měla posílit univerzálnost používaných metadatových formátů, struktur pro ukládání objektů a systému Manakinu při zachování nenáročnosti - jedné z hlavních předností systému *DSpace*. I když se systém *DSpace* neustále vyvíjí, vždy bude existovat snaha zachovat pravidlo instalace a okamžitého provozu „*out of the box*“, tedy garance plně funkčního systému ihned po jeho instalaci bez žádných dalších nároků na jeho uživatele.