

DSpace

<http://www.dspace.org/>

Stručná charakteristika

DSpace je digitální úložiště pro vědecké knihovny a instituce. Je vyvíjen na Massachusetts Institute of Technology společně s Hewlett-Packard Labs a je celosvětově volně k dispozici jako open source výzkumným institucím tak, aby si ho mohli rozšířit a upravit dle vlastních potřeb nebo ho používat tak jak je.

Doba řešení, aktuální stav

Na přelomu dubna a května 2000 se MIT Labs a firma Hewlett-Packard rozhodli spolupracovat na projektu za 1.8 miliónu dolarů. Vytvořit a etablovat elektronické úložiště, které bude sbírat a uchovávat a poskytovat intelektuální výstup z výzkumů fakulty a laboratoří, a které by mohlo sloužit jako model pro ostatní univerzity.

Po dvouletém vývoji byla v dubnu 2002 uvolněna první beta-verze několika institucím (Sloan School of Management, the Department of Ocean Engineering, Center for Technology, Policy and Industrial Development a Laboratory for Information and Decision Systems), k testování a po úspěšném otestování byla 4. listopadu uvolněna ostrá verze.

Nyní je volně k dispozici ke stažení verze 1.2 na SourceForge.net distribuovaná pod Berkeley Software Development licencí. V současnosti DSpace používají desítky institucí a organizací pro sběr a správu svých digitálních výstupů z výzkumů a jiné činnosti.

Cíl projektu

Cílem projektu bylo vytvořit stabilní a dlouhodobě udržovatelné digitální úložiště, které poskytuje možnost zkoumat problémy týkající se řízení přístupu, přístupových práv, verzování, vyhledávání, komunikace mezi komunitami uživatelů a možnosti publikování dokumentů.

Popis projektu

DSpace je digitální úložiště navržené pro sběr, ukládání, indexování a zálohování a redistribuci intelektuálních výstupů z univerzitních výzkumů v digitální podobě běžící na libovolném UNIXu či LINUXu.

DSpace spravuje a distribuuje digitální data v podobě digitálních souborů a umožňuje vytváření, indexování a vyhledávání souvisejících metadat k jejich vyhledávání. Je navržen pro dlouhodobé uložení a uchovávání digitálních materiálů v úložišti.

DSpace je snadno přizpůsobitelný pro jednotlivé DSpace komunity (*DSpace Communities*), které plní své úložiště daty. Jejich úložiště mohou být dále členěna na tzv. sbírky (*Collections*). Komunity mohou být jednotlivé školy, univerzity nebo oddělení na univerzitách, či jednotlivé či skupiny laboratoří.

DSpace sbírky patří jednotlivým nebo více komunitám (např.: 2 výzkumné laboratoře mohou mít společnou sbírku) a obsahují jednotlivé soubory dat.

Čím se DSpace liší od ostatních datových úložišť?

Je to digitální úložiště, které se zabývá různými problémy obsaženými v multi-disciplinárních archívech jako jsou:

- různé postupy a praktiky zavedené jednotlivými vědními disciplínami
- různé digitální formáty používané v dnešních výzkumech
- složitost standardů pro tvorbu metadat nutných pro získávání a udržování přístupu k digitálním formátům podporovaných systémem

DSpace je navržen jako flexibilní úložiště pro různé digitální formáty a pro různé vědní obory - komunity uživatelů. Každá z komunit má svůj portál používající postupy a terminologii příslušné komunity.

Funkce systému

- Systém lze propojit s ostatními systémy používanými v instituci, protože poskytuje Java API rozhraní.
- DSpace má webové rozhraní takže jeho použití a správa jsou jednoduché a na platformě nezávislé.
- Pro implementaci perzistentních ukazatelů na digitální položky úložišť využívá CNRI Handle System.
- DSpace zajišťuje dva způsoby zachování dat:
 - *bitové zachování dat* - tzn. že žádný bit dat se nezmění zatímco se budou měnit v čase datová média
 - *funkční zachování dat* - tzn. zajištění změn formátu dat v čase podle toho jak se budou vyvíjet digitální formáty

DSpace rozděljuje formáty do 3 úrovní a pro všechny tyto skupiny zajišťuje bitovou perzistenci, pro "digitální archeology" budoucnosti.:

- *podporované (supported)* - kde je perzistence zajištěna převodem mezi formáty a jinými verzemi formátů z duvodu znalosti jejich specifikaci tj. např.: TIFF, XML, SGML, AIFF, PDF, ...
- *známé (known)* - což jsou hojně používané populární formáty komerčních firem, a tam není perzistence slíbena a je zajištěna pouze pomocí převodních programů třetích firem př.: Microsoft Word, PowerPoint, Lotus 1-2-3, WordPerfect
- *nepodporované (unsupported)* - které nejsou dost známé na to aby byla zajištěna funkční perzistence, to jsou formáty jednoúčelových aplikací, apod.

Open source

Systém je vyvíjen jako open source a dá se použít tak jak je vytvořen, nicméně některé jeho části (jako je autorizace a autentizace) je vhodno upravit pro konkrétní využití. Také některé části na kterých je postaven lze vyměnit (např. databázi).

Podmínky použití

Systém byl vyvinut na UNIXu a je napsán v Javě a JSP, což implikuje možnost použití i na jiných platformách (MS Windows, Solaris, ...).

Současná implementace používá open source nástroje: Apache - web server, Tomcat - servlet container, PostgreSQL – databáze.

Metadata - "data popisující data"

Pro inteligentní popis dat používá DSpace kvalifikovaný Dublin Core. Povinná jsou jen tři pole: nadpis, jazyk a datum vložení dokumentů, ostatní pole jsou volitelná. Dalšími poli, která je možno vyplnit jsou klíčová slova, abstrakt dokumentu, technická metadata, metadata popisující oprávnění, atd. Tato metadata jsou zobrazována u jednotlivých položek sbírek komunit v DSpace úložišti. Metadata jsou indexována pro jednoduché vyhledávání v úložišti.

Mé hodnocení projektu

DSpace je celkem robustní a jednoduše použitelné digitální úložiště. Implementací v Javě a JSP se elegantně vyřešil problém s jeho použitím napříč platformami. Díky možnosti vyhledávání v dokumentech sbírek i administraci těchto sbírek přes rozhraní webového prohlížeče je jeho použití snadné pro většinu dnešních uživatelů, zejména pak zaměstnanců a studentů univerzit. DSpace umožňuje seskupovat uživatele do skupin a řídit oprávnění jejich přístupu k jednotlivým dokumentům a sbírkám a vytvářet hierarchie komunit a jejich sbírek.

Nejprve jsem se pokusil nainstalovat DSpace v operačním systému Windows firmy Microsoft, což se mi bohužel nepodařilo. Důvodem ale nebyla implementace DSpace,

ale problémy s technologiemi a komponentami, které používá (Jakarta Tomcat, PostgreSQL), a které nemusí vždy tak dobře fungovat.

Na Linuxu proběhla kompilace a instalace DSpace hladce, takže jsem si mohl vyzkoušet jak lze digitální úložiště spravovat a plnit přes webový prohlížeč.

Projekt hodnotím jako přínosný a vysoce použitelný pro různé instituce a organizace, zejména pro otevřenost kódu implementace, a tím možnost upravit digitální úložiště podle vlastních potřeb, za vynaložení minimálních nákladů na vývoj, i když si myslím, že DSpace je použitelný již tak jak je.

Použité zdroje:

- Domovská stránka projektu DSpace - <http://dspace.org>
- DSpace Project Hits Milestone: Early Adopter Phase Begins - <http://libraries.mit.edu/about/news/early-dspace.html>
- Digitální úložiště postavené na DSpace na MIT - <https://dspace.mit.edu/index.jsp>
- DSpace : An Open Source Dynamic Digital Repository- <http://www.dlib.org/dlib/january03/smith/01smith.html>
- Domovská stránka Apache Jakarta Tomcat - <http://jakarta.apache.org/tomcat/index.html>
- Domovská stránka PostgreSQL - <http://www.postgresql.org/>
- Handle system - <http://www.handle.net>
- Java Development Kit – <http://java.sun.com>

Metadata v Dublin Core:

<i>DC attribute</i>	<i>Scheme (if any)</i>	<i>Value</i>
DC.Title		DSpace
DC.Creator		Martin Kotlář
DC.Creator.Address		xkotlar@fi.muni.cz
DC.Description		Esej na téma digitální úložiště DSpace.
DC.Subject		úložiště
DC.Subject		knihovna
DC.Subject		digitální
DC.Date	ISO8601	2005-01-27
DC.Type		Text.Article
DC.Format	IMT	application/pdf
DC.Identifier		http://www.fi.muni.cz/~xkotlar/pv070/dspace.pdf
DC.Language	ISO639-1	cs
DC.Source	URL	http://dspace.org