

MASARYKOVA UNIVERZITA
FAKULTA INFORMATIKY

závěrečná esej do předmětu
PV070 Digitální knihovny

Portico

www.portico.org

Petr Haken

12.12.2007

Stručná charakteristika projektu

Portico je archív sloužící k dlouhodobému uchování odborných článků vydávaných v elektronické formě. Snaží se zajistit, aby tyto materiály zůstaly dostupné příštím generacím studentů a výzkumných pracovníků. Zajišťuje zabezpečený přístup k nim. Obsahuje články ze všech oborů. Umožňuje fulltextové vyhledávání. Pracuje s jednoduchým principem – vydavatel zašle článek, systém jej zpracuje a knihovny a univerzity umožní přístup uživatelům. Dnes mají přístup k tomuto systému knihovny z 11 zemí světa.

Cíle projektu

Projekt vznikl s několika cíli. Hlavní byl uchovat odbornou literaturu publikovanou v elektronické podobě a zajistit, aby tyto materiály zůstaly přístupné budoucím studentům a vědcům. Aby byly uchovány (převedeny) v jednotném formátu, který byl spolehlivě spravován po dlouhou dobu. Dále aby spolupracovali vydavatelé s knihovnami v uchování důležitých výsledků bádání a měli z toho oba prospěch.

Historie

Portico začalo jako iniciativa spuštěná organizací JSTOR v roce 2002. Byla financována nadací Andrewa W. Mellona. Měli za úkol vytvořit trvalý elektronický archív. Více než dva roky se pracovalo na vývoji nezbytných technologií a zjednávání mnoha rozhovorů s vydavateli a knihovnami, aby sjednali přístup, který by vyhovoval jak vydavatelům, tak knihovnám, dokud se nevytvoří dostatečný systém financování archívu. V roce 2004 se iniciativa stala součástí firmy Ithaka Harbors – nevýdělečné organizace s cílem urychlit používání informačních technologií v prospěch větší vzdělanosti po celém světě. Takto vzniklo Portico – systém, kde na vstupu vydavatelé poskytují komerční články, aniž by na tom vydělávali a ty byly malými, středními i velkými knihovnami vedeny k uživatelům. Tento systém byl spuštěn roku 2005 s podporou JSTORu, Ithaky, Kongresové knihovny a nadace Andrewa W. Mellona.

Aktuální stav

V současné době je do tohoto programu zapojeno 46 vydavatelů. Patří zde například Elsevier nebo Oxford University Press. Články jsou poskytovány 397 knihovnami. Nejsou to jen knihovny ze spojených států. Jsou zde zastoupeny země jako Bangladéš, Libanon, Nový Zéland, Indie, Kypr, Austrálie, Řecko, Švédsko, Velká Británie nebo Kanada. V archívu je přes 3,6 milionů článků z 7223 různých časopisů. Neobsahuje pouze text, ale i obrázky ve vyšší kvalitě, než v tisknutých vydáních. Také obsahuje vnitřní odkazy a podobné vymoženosti informační techniky.

Financování

Takovýto projekt potřebuje financování dvojího druhu. Prvním jsou peníze na vývoj a zavedení technické infrastruktury. Druhý je samotné udržování archívu po velmi dlouhou dobu. Grant od JSTORu, Mellonova nadace a Kongresová knihovna poskytli finance prvního druhu. Dlouhodobé financování je zajištěno každoročními příspěvky (poplatky) od vydavatelů i knihoven. Ty se počítají podle příjmů společností. U nakladatelů se pohybují od 250 do 75000 dolarů od knihoven od asi 1500 do 24000 dolarů.

Přístup do archívu

Přístup ke článkům v archívu, které nejsou dlouho dosažitelné od vydavatele, je zaručen knihovnám, když jsou splněny speciální podmínky, tzv. spouštěče událostí (trigger events). Těmito podmínkami jsou splněny, pokud vydavatel dlouho nevydává článek nebo pokud jej přestane publikovat nebo když kvůli přírodní katastrofě vydavatel trvale přestane dodávat články. 4 knihovníci z každé knihovny dostanou přístup pomocí hesla, aby mohli kontrolovat a ověřovat plnění závazků. Vydavatelé také dostanou hesla k jejich vlastnímu obsahu v archívu.

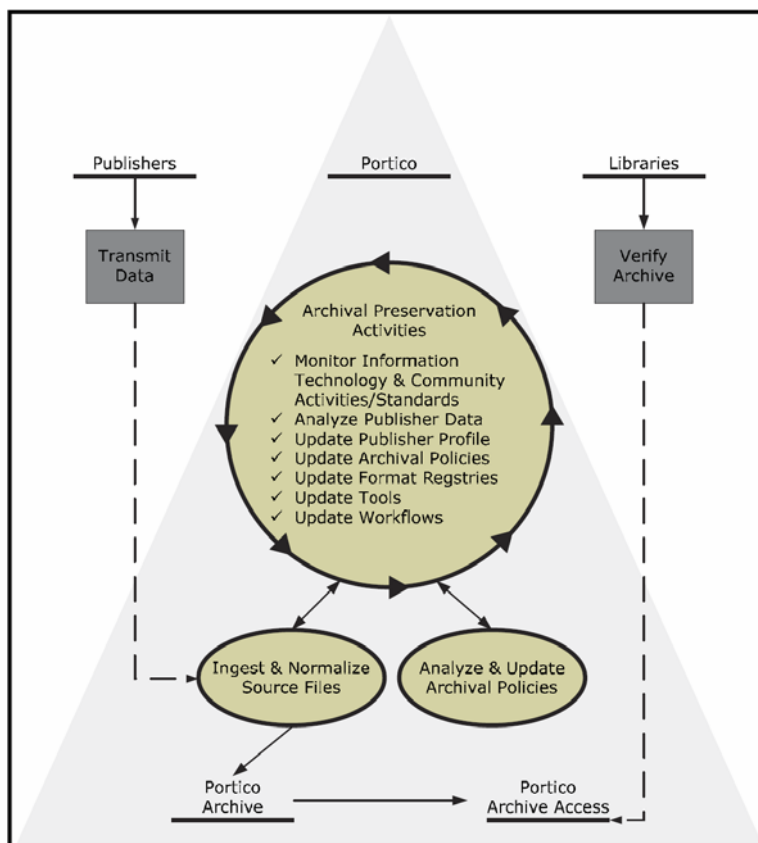
Politika uchování

Portico zavedlo množinu principů, které plní pro přístup a uchování elektronických materiálů. Tyto principy vyplynuly z toho, jak současná vědecká komunita chápe uchování digitálních materiálů a z diskusí s knihovnami a vydavateli.

- *musí být zachována neporušenost vědeckých záznamů* – originální materiály jsou převáděny do jiného formátu. Původní i převedená verze musí být trvale uchována
- *záměr je archivování vědeckého obsahu* – tento projekt vznikl právě kvůli uchování vědeckých prací. Proto nebude ustupovat od tohoto svého cíle.
- *zdrojové soubory mohou spolehlivě zachytit intelektuální obsah* – zdrojové soubory jsou často PDF, které lze použít k tisku a zobrazení, XML nebo SGML soubory a metadata, mnoho obrázků, hudby a videa. Některé druhy zdrojů nelze převést, zaznamenat nebo použít bez systému vydavatele. Některé soubory proto nebudou uvolněny pro archivaci.
- *Pro dlouhodobé uchování bude použita technika migrace*

Proces archivace

Proces začíná, když vydavatel zašle zdrojové soubory do archívu. Pak se spustí automatizovaný proces, jehož výsledkem je sbírka souborů ve formátu nezávislém na formátu od vydavatele. Tento balík je poté vložen do archívu. Co se děje uvnitř samotného procesu?



Proces uchování

Od vydavatele přijde balík souborů v předem definovaném formátu. Je očekáván určitý formát, což je analýzou zkontrolováno. Poté jsou vytvořena metadata, pro uchování v Porticu. Používá se XML soubor v upraveném METS standardu, takzvaný PMETS nebo Portico METS. Dále se z dat odstraní odkazy na majitelovy data, zkontroluje se, zda výsledek odpovídá tomu, co jsme chtěli a monitorujeme, zda je tento nový formát dat akceptován uživateli. Formát a politika uchování je pravidelně kontrolována, a když je to vhodné, tak se provedou změny.

Moje zhodnocení projektu

Na závěr bych uvedl, že jsem si tento systém nemohl vyzkoušet ani v nějaké zkušební nebo jinak omezené verzi, protože veškerý přístup je jen pro zaregistrované knihovny. Je vhodné, jaký tento projekt zavedl politiku financování, která je pro dlouhodobé uchování dat velice potřebná. Platí jak knihovny, tak i vydavatelé. Navíc neplatí všichni stejně, ale poplatky jsou závislé na finančním zabezpečení knihovny. Podobně i na straně vydavatelů. Velmi užitečná je i úprava ukládaných dat, když je to potřeba, a flexibilní politika jejich zpracování, což jistě má za důsledek trvání, ale i akceschopnost archívu. Navyšování pravidelných (ročních) poplatků není nijak přehnané, navíc pro již déle zaregistrované uživatele (vydavatele i knihovny) je nižší, než pro nově příchozí. Životnost projektu ukazuje nejen, že je zaštitěn Kongresovou knihovnou, ale i zúčastněním velkých vydavatelů, jako je například Elsevier, Springer, Oxford University Press a mnoho dalších.

Na internetu se mnoho článků navždy ztratí již po dvou letech, proto archív, kde se uchovávají všechny články, je téměř nezbytností. Zúčastnění téměř všech univerzitních knihoven v (nejen) anglicky mluvícím světě umožňuje snadný přístup vědců i studentů k potřebné – a jinak někdy prakticky nedosažitelné – literatuře.

Projekt již funguje dva roky a za poslední týden psaní této eseje přistoupily další 4 knihovny k tomuto projektu, přidalo se dalších 15 zdrojů článků více a přibylo neuvěřitelných 600 tisíc článků (o 20 procent více) z dalších asi 750 nově archivovaných děl (celkem asi 4000). Do budoucna by bylo vhodné rozšířit tento projekt i do dalších zemí. Problémem bych neviděl legislativu, ale spíše domluvu, aby se knihovny a vydavatelé různojazyčných zemí připojili do projektu, který je zatím anglicky. Ale země jako Řecko nebo Švédsko tuto překážku již překonaly.

Zdroje, literatura

1. Portico (<http://www.portico.org>)
2. JISC Collections (http://www.jisc-collections.ac.uk/catalogue/coll_portico.aspx)
3. DigitalPreservationEurope
(http://www.digitalpreservationeurope.eu/publications/portico_briefing_paper_def.pdf)
4. Ithaka (<http://www.ithaka.org/incubated-entities/portico>)

Metadata v Dublin Core

```
<link rel="schema.DC" href="http://purl.org/dc/elements/1.1/" />
<meta name="DC.Title" content="Portico" />
<meta name="DC.Creator" content="Petr Haken" />
<meta name="DC.Subject" content="Portico" />
<meta name="DC.Description" content="PORTICO je systém pro dlouhodobou
archivaci vědeckých publikací." />
<meta name="DC.Date" content="5.12.2007" />
<meta name="DC.Type" content="Text" />
<meta name="DC.Type" content="esej" />
<meta name="DC.Format" content="application/pdf" />
<meta name="DC.Format" content="computerFile" />
<meta name="DC.Identifier"
content="http://www.fi.muni.cz/~xhaken/PV070/Portico.pdf" />
<meta name="DC.Source" content="http://www.portico.org/" />
<meta name="DC.Language" content="cze" />
```