

Netextové objekty jako součást databáze kvalifikačních prací AMU

Autor: Radim CHVÁLA
Spoluautor: Iva HOROVÁ
Akademie múzických umění v Praze
radim.chvala@amu.cz
iva.horova@amu.cz

Příspěvek byl poprvé publikován pod názvem „Repositář audiovizuálních dokumentů Akademie múzických umění v Praze“ ve sborníku:

INFORUM 2009: 15. ročník konference o profesionálních informačních zdrojích, Praha 27.-29. května 2009 [online]. Praha: Albertina icome Praha, 2009. Dostupný z WWW: <http://www.inforum.cz/sbornik/>. ISSN 1801-2213.

Abstrakt:

Studenti a pedagogové Akademie múzických umění v Praze (dále AMU) produkují průběžně velké množství audiovizuálních materiálů zejména ve formě diplomových, magisterských, doktorandských, klauzurních a semestrálních prací. Specifikum přípravy budoucích profesionálních umělců je v tom, že v rámci závěru studia neodevzdávají pouze jednu kvalifikační práci, nýbrž prací několik, každou se samostatným tématem, vedoucím, oponenty, anotací a klasifikací. Je to jednak vždy povinně textová práce, avšak obvykle ještě záznam uměleckých výkonů (absolventský film, divadelní představení, grafické podklady, hudební koncert, soubor fotografií apod. Někdy se jedná o kolektivní dílo, sloužící jako příspěvek k závěru studia více studentům (např. absolventský film využije nastávající režisér, ale i scénárista, produkční, zvukař apod.). Uchování a snadné vyhledání a možnost náhledu do těchto děl je samozřejmě důležité stejně jako u klasických "textových" prací.

Kromě autorskoprávních problémů, že je zde řada dosud neřešených otázek metodické a technické povahy, které komplikují použití institucionálních repositářů v klasické podobě. V nich se kvalifikační práce obvykle katalogizuje jako jeden dokument s přílohami, což by však v našem případě vedlo k přiřazení různých původců k jednomu dokumentu a jejich zpětné přiřazení ke konkrétní práci by nebylo možné.

Tento problém byl na AMU vyřešen prostřednictvím úprav open source systému Dspace. Metadatový soubor na bázi Dublin Core byl rozšířen o potřebné kvalifikátory prvku relation. Editovací a zobrazovací moduly byly doprogramovány, aby umožňovaly vyhledat a zobrazit související dokumenty. Do systému byly integrovány přehrávače video a audio souborů, které jsou nyní přímo součástí stránky s dokumentem. Byl instalován server pro jednotný identifikátor Handle.

Zároveň byl vyřešen i automatizovaný přenos metadat ze Studijního systému školy do repositáře a dále protokolem OAI automatické sklízení do celonárodního registru Theses MU Brno a export do knihovního systému AMU.

Úvod:

Příspěvek popisuje jedno z možných řešení problému, který nějakým způsobem řeší všechny vysoké školy v České republice i v zahraničí. Archivace a zpřístupňování dokumentů produkovaných univerzitou je žádoucí proces, který mj. slouží jak k průběžnému zvyšování kvality této produkce, tak k racionálnímu a hospodárnému využívání intelektuálního kapitálu, který tato produkce obsahuje. Také proto je ošetřen zákonem (111/1998 Sb., § 47b).

V prostředí umělecké vysoké školy existují některá specifika, která bylo nutno vzít v úvahu a konkrétní řešení pro AMU patřičně modifikovat.

Tvorba dokumentů na AMU:

Studenti a pedagogové při studijní, pedagogické, umělecké a výzkumné činnosti produkuje průběžně během studia mnoho různých typů dokumentů. Jedná se o bakalářské, diplomové, disertační, ročníkové, semestrální, Klausurní a jiné práce. Vzhledem k uměleckému zaměření školy mají mnohdy audiovizuální charakter a povahu samostatných děl. Jejich archivace a zpřístupňování je úkolem, na kterém participují jak pedagogická pracoviště (fakulty, katedry), tak celoškolská informační pracoviště (Počítačové centrum a Knihovna AMU).

Pro dlouhodobé ukládání a zveřejňování těchto výsledků pedagogické, vědecko-výzkumné a umělecké práce studentů i pedagogů bylo rozhodnuto vybudovat na AMU institucionální repositář s některými archivními funkcemi. Bude sloužit především jako operativní nástroj, s dokumenty uloženými v úsporných náhledových formátech. Umožní rychlé nalezení dokumentů a snadné posouzení jejich dostupnosti a zajímavosti. Pokud uživatel vyhodnotí nalezený materiál jako užitečný či zajímavý, požádá např. katedru nebo autora apod. o poskytnutí dokumentu (díla) ve vyšší technické kvalitě. Repositář bude také propojen s knihovním systémem, který zůstane i nadále prvním místem pro vyhledávání informací. Dále bude propojen se Studijním informačním systémem a konečně musí být také schopen poskytovat metadata mimo AMU tj. nyní konkrétně např. do Národního registru VŠKP.

Závěr studia, při kterém vznikají kvalifikační práce, má na AMU řadu odlišností od situace na ostatních vysokých školách. Na většině škol odevzdávají studenti jednu práci, eventuálně doplněnou různými přílohami. Práce včetně příloh (byť třeba rozsáhlých) se vztahuje k jednomu tématu, má jednoho vedoucího, jednu skupinu oponentů, lze ji popsat jednou společnou anotací a je hodnocena jednou klasifikací. Pro takovou kvalifikační práci se pak obvykle vytváří jeden popisný metadatový záznam, v němž jsou v příslušném prvku (evskp:fileProperties) uvedeny základní údaje o existujících přílohách. Naproti tomu na AMU většina studentů v závěru studia odevzdává několik do značné míry samostatných prací, textových i jiných – hovoříme o tzv. „kvalifikačním výkonu“ či „diplomovém úkolu“. Terminologii v tomto směru by bylo zřejmě vhodné sjednotit na národní úrovni. Takovou práci může být např. absolventský film, soubor fotografií, interpretační výkon různé povahy (nástroj, dirigent, tanec) atd. apod.. Každá z těchto prací pak může mít své samostatné a jiné téma, jiného vedoucího, jiné oponenty, samostatnou anotaci a zejména má vlastní klasifikaci. Počet těchto prací určuje příslušná katedra v závěru studia. Každý student však musí vždy odevzdat povinně textovou práci, která má charakter podobný kvalifikační práci, jak je obvyklé na jiných vysokých školách. Metodika vytváření společného metadatového záznamu pro všechny součásti „kvalifikačního výkonu“ se při prováděných pokusech ukázala jako nedostatečná a nevyhovující – vede k nepřehlednosti a nemožnosti správně rozklíčovat příslušné vazby (např. kdo oponoval kterou práci apod.). Předpokládáme, že tato situace se netýká jen ryze umělecké VŠ, ale že se s ní bude možno setkat také univerzitách, které mají „umělecké“ fakulty či obory.

Všechny uvažované typy kvalifikačních prací byly proto pracovně na AMU rozděleny do několika základních typů. Teoretická práce – typ A, Hra, scénář – typ B, Film, videozáznam – typ C, Interpretační výkon – typ D, Skladba – typ E a je pro ně vždy vytvářen samostatný metadatový záznam.

Před zahájením celého procesu vedoucího k cílovému zpřístupňování a tedy k budování repositáře, bylo nutné zpracovat řadu vnitřních předpisů, výnosů rektora AMU, a

řešit otázku licenčních smluv, vyplývajících z vnější legislativy (zejména ze Zákona o vysokých školách a Autorského zákona). Dále byl na AMU jako norma přijat soubor národního metadatového standardu MS-EVSKP zpracovaný Odbornou komisí pro otázky elektronického zpřístupňování vysokoškolských kvalifikačních prací při AKVŠ.¹ Tyto normy jsou nyní na AMU závazné a určující pro tvorbu, archivaci a zpřístupňování kvalifikačních prací v elektronické podobě.

Při řešení této problematiky jsme samozřejmě nepostupovali izolovaně, ale průběžně jsme konzultovali s pracovníky ostatních vysokých škol, zejména zastoupených v Komisi eVŠKP; spoluautorka tohoto příspěvku měla řadu odborných vystoupení na téma práce s kvalifikačními pracemi audiovizuální povahy a v roce 2008 se AMU také úspěšně zúčastnila společného centralizovaného rozvojového projektu MŠMT „Národní registr VŠKP a systém na odhalování plagiátů“, v jehož rámci byl mj. vybudován nástroj na automatizované předávání popisných metadat z repositáře AMU do Národního registru VŠKP na Masarykově univerzitě v Brně. Prozatím jsou běžně předávána metadata textových prací, předávání ostatních typů se připravuje.

Náš projekt od začátku neměl ambici suplovat v plném rozsahu archiv audiovizuálních dokumentů, neboť je zde ještě mnoho i na národní úrovni nedořešených metodických otázek a AMU zatím nemá technické prostředky pro ukládání, zálohování a zpřístupňování značných objemů dat, jaké by regulární digitální archiv vyžadoval. Hledali jsme možnost jak zpřístupnit netextový obsah „informativním“ způsobem, který by umožňoval rychlou orientaci a poskytl uživatelům možnost rozhodnout se, zda budou chtít následně primární dokument skutečně dohledat v plné kvalitě. Repositář má tedy fungovat především jako operativní nástroj s možností prohlížení obsahu dokumentů v úsporných náhledových formátech, které je obtížně komerčně zneužít a lze s nimi pracovat i na méně výkonných systémech. Byly proto stanoveny základní pracovní formáty pro jednotlivé typy dokumentů. Pro ukládání textových a statických obrazových prací byl zvolen formát PDF/A-1a, pro ukládání zvukových dokumentů formát MP3 a pro ukládání videozáznamů formát FLV. Tento výčet není definitivní, počítáme s tím, že bude průběžně korigován, např. v souladu s usnesením Vlády ČR č. 1338 ze dne 3. 11. 08.²

Budování repositáře:

Na problematice zpřístupňování VŠKP se podílejí 3 systémy: Studijní informační systém KOS, vlastní repositář a knihovní systém. V každém z nich bylo na základě provedených analýz nutné provést řadu úprav a doplňků.

Příslušné moduly Studijního informačního systému byly doplněny o potřebné formuláře pro založení záznamů prací a vyplnění metadat podle standardu MS-ESVSKP. Byl vytvořen též modul pro exportování zkompletovaných metadat ve formátu XML. Práce na úpravách systému byly do značné míry koordinovány s ČVUT, které používá stejný studijní systém.

Jako vlastní repositář byl vybrán systém Dspace: pro malé finanční náklady (Open source), jednoduchou instalaci, a administraci, snadné možnosti interních úpravy a již dalšími univerzitami v ČR zvládnutou lokalizaci. Systém nativně podporuje nejdůležitější standardy

¹ Viz <http://www.evskp.cz/standardy/evskp/>

² Viz www.mvcr.cz/soubor/vv-0608-pdf.aspx

(XML, DC, MODS, METS). Pro potřebnou interoperabilitu využívá vestavěný OAI-PMH server. Podporuje i zabezpečený přístup napojením na systém identifikace uživatelů sítě AMU (LDAP). Obsahuje výkonný vyhledávací mechanismus, včetně fulltextového vyhledávání v uložených dokumentech textové povahy. V neposlední řadě jsou zde již k dispozici pozitivní zkušenosti více vysokých škol, které systém reálně provozují (VŠB-TUO Ostrava, Univerzita Pardubice, ČVUT Praha, MU Brno aj.). V ČR již také existuje komunita, která pořádá setkání uživatelů, kde je možné konzultovat problémy a systém se tak dále vyvíjí.³

Úpravy na systému Dspace se týkaly v první řadě struktury metadat. Dspace v originální instalaci pracuje s rozšířeným metadatovým schematem Dublin Core. Ten bylo potřeba rozšířit o prvky ID autora, Datum narození autora, Kód katedry, Název katedry, ID studia, k jehož závěru se práce vztahují, a Typ práce. Přidáním těchto prvků byl pokryt metadatový standard MS-ESVKP.

Dále však bylo nutno vyřešit hlavní problém – vazby mezi souvisejícími samostatnými záznamy, tedy mezi pracemi, které tvoří „kvalifikační výkon“ tj. vztahují se k ukončení jednoho studia jednoho studenta. Pro tento účel byl využit prvek „dc:relation“ s atributy „hasPart“ a „isPartOf“. Textová práce (představuje ji typ formuláře A) je v dané množině souvisejících prací vždy jediná a „nadřazená“. Je tedy označena „dc:relation hasPart“. Ostatní (B, C, D, E) jsou „podřízené“. A jsou označeny elementem „dc:relation isPartOf“.

Dále byl upraven vzhled uživatelského rozhraní Dspace a byl důsledně přizpůsoben grafickému korporátnímu stylu AMU a byla vytvořena anglická mutace.

Další, poněkud nečekaný problém, se vyskytl s fulltextovým prohledáváním českých textů, které je pokládáno za jednu z hlavních přidaných hodnot celého systému. Při ukládání souboru pdf si Dspace pomocí mediafiltru s knihovnou pdfbox vytváří pomocný textový soubor pro indexování a fulltextové hledání. V některých případech nebyly korektně extrahovány znaky s diakritikou a fulltext potom nefungoval. Ověřili jsme si, že záleží na tom, jaký nástroj pro tvorbu pdf bude použit. Proto jsme přijali závěr, že do repositáře budou vkládány pouze ty soubory typu PDF/A které vznikly standardizovaným způsobem. Ve spolupráci s firmou Software602 byl pro AMU upraven a nasazen jejich produkt Print2PDF. Uživatel otevře na intranetu školy dialogové okno a vloží do něj svou práci. Automaticky se vytvoří správný PDF soubor a vrátí jej uživateli. Byl vytvořen mediafilter (java plugin) který místo knihovny „pdfbox“ používá pro extrakci textu utilitu „pdftotext“.

Pro trvalou identifikaci záznamů byl v repositáři nainstalován a zprovozněn server pro perzistentní identifikátor HANDLE. ([hdl.handle.net/10318/...](http://hdl.handle.net/10318/)) se kterým již má mnoho uživatelů Dspace dobré zkušenosti.

Dále bylo naším úkolem zajistit přenos metadat z repositáře do elektronického katalogu tj. provést migraci pro potřeby bibliografických záznamů do knihovního systému Tinlib. Uvnitř TinLibu nebylo nutné dělat žádné systémové úpravy. Technologií XML/XSLT bylo však třeba vytvořit šablony, které pomocí procesoru SAXON transformují soubory, exportované z Dspace, na soubory importovatelné do Tinlibu. Metadatové záznamy jsou přenášeny včetně svých vazeb tj. ze záznamu textové práce je vytvářen záznam na monografické úrovni, zatímco ostatní související záznamy netextových prací, které tvoří jeden „kvalifikační výkon“, se přenášejí na úroveň analytickou. Jedině tak je možné zajistit jejich vzájemné propojení a zachovat „nadřazenost“ textové práce. V knihovním katalogu jsou poté ručně knihovníky doplňovány chybějící údaje bibliografického popisu, a dalším importem přeneseny do Dspace.

³ Viz www.dspace.cz

Interporebilita systému Dspace je v základní instalaci zajištěna serverem OAI-PMH, který zajišťuje bezproblémové vystavení metadat pořízených v Dublin Core. Zde bylo potřeba upravit příslušný Java plugin, aby byla zpracována též přidaná metadata dle standardu MS-ESVKP. Tato úprava se netýká stabilní části kódu, nebude tedy ovlivněna případnými upgrade.

Vazba repositáře na okolí:

Jedním z cílů budovaného repositáře je předávání metadat do jiných systémů mimo AMU. V rámci již zmíněného centralizovaného projektu se AMU zavázala přispívat svými záznamy do Národního registru VŠKP theses MU. Vedení AMU rozhodlo, že budeme poskytovat do registru pouze metadata, nikoli plné verze prací. V první fázi jsou zatím předávány pouze textové práce (typ A). Vlastní předávání probíhá tak, že záznamy jsou na určené adrese vystaveny pro sklizeň (harvest). Přebírající strana každých 24 hodin kontroluje, zda již záznam existuje a podle data úpravy u každého záznamu určí, zda jej přebere jako nový, či nahradí stávající nebo neudělá nic. Zatím bylo z AMU tj. ze systému DSpace předáno do Národního registru cca 300 záznamů o VŠKP obhájených v roce 2008. Nadále připravujeme předání záznamů kvalifikačních prací starší produkce a počítáme i s předáváním záznamů o netextových pracích. V tuto chvíli by však bylo možné předávat pouze samostatné záznamy, v nichž by se neprojevovaly vazby na příslušné textové práce. Zde však prozatím narážíme na nepřipravenost českého metadatového standardu, který s touto eventualitou původně nepočítal. Rádi bychom se podíleli na dalším dopracování standardu MS_EVSKP tak, aby standardně obsahoval také vazebné prvky. Dále plánujeme také dohodnout s Národním registrem na MU v Brně možnost rozšíření importního formátu na jejich straně tak, aby NR mohl přijímat a hlavně zobrazovat provázané záznamy.