

Návrh rozšíření aplikace Resolver

Změna datového modelu

Stávající Resolver vznikl jako modul Registru digitalizace a jako takový je s tímto systémem úzce propojen. Metadata jsou nyní získávána přímým přístupem do jeho databáze a editována mohou být administrátorem přes webové rozhraní. Ač mohlo být datový model pro dřívější účely dostatečný, dnes je rozvoj Resolveru jako samostatného systému tím velmi limitován. Nové požadavky si vynucují jeho radikální změnu. Mezi problematické části patří zejména:

- Identifikátory Intelektuální entity (dále jen IE) i digitální reprezentace (DR) jsou pevně dané a není možné použít jiné nebo některé pro určitý typ dokumentu nepoužít vůbec. IE sice umožňuje kromě cCNB, ISBN a ISSN použít i jinyId, i to však nemusí být dostačující. DR má povinný URN:NBN a dále cislo_zakazky a cisloRDCZ. Místo těchto atributů budou IE a RD v relaci 1:M s entitami IEIdentifier a RDIdentifier. Takto lze druhy identifikátorů specifikovat pro různé typy dokumentů. Podobně budou odstraněny atributy cisloPeriodika a rocnikPeriodika a nahrazeny
- Na úrovni databáze není možné určit jen jednoho původce, IE má atributy autor, akce i korporace. To bude nahrazeno vazbou na entitu Originator s atributem typ, který může nabývat hodnot author/action/corporation.
- Instituce musí mít atribut SIGLA, což v realitě neplatí. Tento atribut byl nahrazen atributem id, který může ale nemusí být siglou.

Další rozšíření datového modelu

- Přibyla entita catalogue. Ta reprezentuje katalog nebo jiný systém, který pokud bude zdrojem přístupu na resolver pod některým urn:nbn, má se resolver zachovat jinak a pokusit se přeměřovat na url Digitální Instance (DI).
- URN:NBN bylo přesunuto do vlastní entity, protože musí být zachováno i v případě smazání DR a o tomto stavu tak informovat.
- Kvůli vytvoření uživatelských účtů pro instituce s právy importu a editace metadat bude vytvořena entita User se vztahem "has access to" na entitu Institution.
- DR přijde o atributy pridelenoDne a pridelenoKym, jelikož nemají smysl a nevyžívají se.
- Přibudou Entity používané novým modulem OAI Adapter

Stávající data v Resolveru obsahují nepřesné a chybné údaje. Proto budou muset být vytvořeny skripty pro jejich opravu a transformaci do nových tabulek. Změna datového modelu pravděpodobně ovlivní i způsob vyhledávání podle identifikátorů. Výrazy budou prefixovány typem identifikátoru, tj. např. ISBN:1803-4217 místo dnešního 1803-4217.

Generování URN:NBN

Velkou změnou bude způsob generování URN:NBN. Ty nyní vytváří Resolver, nové varianty digitalizačních procesů budou v některých případech vyžadovat, aby urn:nbn vytvářela instituce a byly tyto identifikátory byly přítomny v importovaných metadatech. Několik možností realizace této změny je posáno v příloze priloha_Pridelovani_URN-NBN.pdf.

Účty institucí

Instituce budou moci samy importovat a editovat metadata svých digitalizovaných objektů, proto je potřeba přidat správu uživatelských účtů. Kvůli zabránění zahlcení serveru a kontrole těchto účtů však bude administrátor jednotlivé účty povolovat. Po aktivaci účtu budou instituce importovat metadata buď přes RESTové rozhraní, nebo přes modul OAI Adapter.

Import a aktualizace metadat

Importovat a editovat metadata budou moci instituce dvěma způsoby:

1. Pokud mají k dispozici OAI Provider a v záznamech potřebná metadata včetně url na digitalizované dokumenty, tak použijí modul OAI Adapter.
2. jinak přes RESTové rozhraní s tím, s tím, že si případně vyrobí modul do vlastního systému, který s rozhraním komunikuje.
3. případně ručně přes webové rozhraní.

RESTové API

Je určeno pro přístup dalších systémů. Specifikováno je v příloze priloha_REST_API.pdf. Na základě rozhodnutí o způsobu přidělování URN:NBN (priloha_Pridelovani_URN-NBN.pdf) bude umožňovat jeden nebo více způsobů, jak zapojit import do URN:NBN Resolveru do procesu digitalizace a zpřístupnění.

OAI Provider

Modul umožní instituci spravovat tzv sledování OAI. Správa sledování bude přístupna přes webové rozhraní. Ke každému sledování se váží:

- oai base url
- název oai setu/seznam oai identifikátorů
- typ dokumentů
- transformační šablona
- seznam metadatových formátů
- frekvence aktualizací

Po vytvoření sledování a v pravidelných intervalech se pak stahují ze setu (podle seznamu identifikátorů) metadatové záznamy ve specifikovaných formátech. Všechny záznamy pro některou položku (item) se šablonou transformují v jedno xml pro import/update Resolveru. Vniklá xml budou validována a importována do Resolveru. Pro každý import/aktualizaci bude vytvořen log přístupný přes webové rozhraní. Pokud bude urn:nbn přiděleno, bude ho možné zobrazit v tomto logu anebo přes RESTové rozhraní na základě oai identifikátoru.

Pokud by se oai identifikátor v budoucnu změnil, ale metadata již obsahovala urn:nbn, OAI Adapter podle urn:nbn identifikuje DR a změní její alternativní identifikátor - oai id. Viz diagramy OAI Adapter - conceptual, Register Oai watch a Initial WatchedOaiSet import v příloze priloha_UML_diagramy.pdf.

Spojování různých DR pod stejnou IE

Mluvím-li o importu metadat, myslím tím hlavně import DI. Xml dokument, který je importován do Resolveru ale obsahuje i metadata IE. S těmi se zachází takto:

- Pro typ Book Resolver zjistí, zda již v databázi existuje IE se stejným cCNB, jako právě

importovaná IE. Pokud shodu nenajde, zkusí ISBN. Pokud neuspěje ani zde, vytvoří z poskytnutých metadata novou IE. Je-li shoda nalezena, poskytovaná metadata IE nejsou použita a importována je jen DR případně DI. DR se pak napojí na nalezenou IE.

- Pro typ Periodical se hledá shoda na atributu ISSN, pro typ PeriodicalVolume na ISSN a zároveň na volumeName.
- Pro PeriodicalIssue na ISSN, volumeName a zároveň issueNumber.
- Pro nové typy dokumentů budou definovány jiné identifikátory.
- Pro některé typy dokumentů se shoda nehledá. Jde třeba o digital born typy, např. Thesis nebo Analytical.

Dodatečné DI

Protože se předpokládá, že url do digitálních knihoven nemusí být v době importu do Resolveru známy, není nutné rovnou uvádět DI. To je možné udělat až dodatečně editací DR.

Zapojení externích systémů

Národní Technická knihovna

Tato instituce má funkční OAI Provider a může tak pro import a aktualizace používat modul OAI Adapter.

Repozitář Univerzity Karlovy

Mají funkční OAI modul a budou schopni specifikovat set, případně seznam identifikátorů pro sledování modulem OAI Adapter.

Ikaros

Jimi používaný systém Drupal umožňuje provozovat OAI provider, nemají ho však instalovaný. Plánují upgrade Drupalu do konce roku, takže poté snad zprovozní i ten OAI Provider. Zvažují také ruční vkládání metadat.

OAI Provider

Výstupní rozhraní Resolveru pro sklizení jeho záznamů jinými institucemi. Bude použit software proai, který sklízí záznamy z databáze systému (zde Resolver) a ty si ukládá do cache, čímž umožňuje provádět sklizeň i v situaci, kdy samotná aplikace Resolveru není přístupná.