

ZPRÁVA O VÝSLEDKÁCH ŘEŠENÍ PROJEKTU V ROCE 2012

INTERPI – Interoperabilita v paměťových institucích

Program aplikovaného výzkumu a vývoje národní a kulturní identity (NAKI) (DF11P010VV023)

Předkládá:

Mgr. Marie Balíková, odpovědná řešitelka - koordinátorka
Národní knihovna České republiky
Klementinum 190
110 00 Praha 1

Spolupracovali:

Mgr. Jana Šubová, Ing. Naděžda Andrejčíková PhD.,
Ing. Miroslav Kunt

Ing. Tomáš Böhm
generální ředitel
Národní knihovna ČR



INTERPI

15. 11. 2012

Obsah

A	Konstatační část	3
A.1	Rešerše	3
A.2	Vstupní data a cíl	3
B	Analytická část	5
B.1	Analýza, ověřování, doplňování a úprava dat	5
B.1.1	Úprava a doplnění záznamů věcných autorit	5
B.1.2	Úprava a doplnění záznamů geografických autorit / entit ze třídy geografický objekt	6
B.1.3	Záznamy tematických autorit / entit ze třídy událost - podtřída války, bitvy	7
B.1.4	Záznamy geografických autorit / entit ze třídy geografický objekt - podtřída sídlo	8
B.1.5	Záznamy geografických autorit / entit ze třídy geografický objekt - podtřída historický správní celek	8
B.1.6	Úprava a doplnění záznamů jmenných autorit	9
B.1.6.1	Anonymní tvůrci	9
B.1.6.2	Šlechtické rody	10
B.2	Termíny z rejstříků Národního archivu	11
B.2.1	Korporativní entita - spolky	11
B.2.2	Záznamy o původcích typu korporace	11
B.2.3	Komparace rejstříků archivních pomůcek Národního archivu s bází národních autorit	11
B.3	Další zdroje termínů	13
B.3.1	Nemovité památky a souvisejí termíny	13
B.3.2	Tematické autority používané v Registru sbírek výtvarného umění a v Muzejních autoritách	15
B.3.3	Věcné/tematické entity (pojmy, objekty) v projektu INTERPI z pohledu regionálních muzeí	16
B.4	Nástroje sémantického webu a nástroje propojení různých terminologických systémů	17
B.4.1	Linked data	18
B.4.2	Potenciální aplikace klasifikačního systému MDT jako nástroje propojení různých systémů	19
B.5	Návaznost projektu na koncepční dokumenty a jiné národní projekty	20
B.5.1	Příprava metodických a metodologických postupů, jejich zavedení v archivech ..	20
B.5.2	Návaznost projektu na Koncepci rozvoje knihoven a na projekt CPK	21
B.6	Vymezení tříd entit	22

B.6.1	Třída entit osoba/bytost.....	23
B.6.2	Třída entit rod/rodina	23
B.6.3	Třída entit korporace.....	23
B.6.4	Třída entit geografický objekt	24
B.6.5	Třída entit událost	25
B.6.6	Třída entit dílo/výtvar	25
B.6.7	Třída entit pojem	25
B.7	Doplnění a úprava datové struktury.....	25
B.8	Funkční požadavky na softwarové řešení	26
B.9	Funkční koncept softwarového řešení a prototyp	26
B.9.1	Funkční koncept softwarového řešení	26
B.9.2	Funkční prototyp webového rozhraní.....	29
B.10	Propagace a informování o projektu	29
C	Návrhová část	34
D	Použití účelové podpory v členění dle rozpočtu projektu	35
E	Resumé.....	36
F	Přílohy	
F.1	Použití účelové podpory v členění dle rozpočtu projektu - komentář	
F.2	Dodatky, rozhodnutí	
F.3	Cestovní zprávy	
F.4	Funkční koncept webového rozhraní pro zpracování entit	
F.5	Zpráva o činnosti Národní knihovny v projektu INTERPI za rok 2012	
F.6	Zpráva o činnosti Národního archivu v projektu INTERPI za rok 2012	
F.7	Přehled požadavků paměťových institucí na zpracování entit pro projekt INTERPI - zpráva za rok 2012	
F.8	Nástroje sémantického webu a věcné autority - aplikace v projektu INTERPI	

A Konstatační část

A.1 Rešerše

BALÍKOVÁ, Marie. Věcné zpracování a zpřístupnění informačních zdrojů. In *36. celostátní seminář knihovníků muzeí a galerií při AMG v Uherském Hradišti*. 2012. Příspěvek na konferenci. Dostupné na www: <http://www.cz-museums.cz/UserFiles/file/komise/knihovnicka/seminare_prispevky/UherskeHradiste/Balikova_Vecne_zpracovani.pdf>.

BALÍKOVÁ, Marie. Czechia : interoperability in memory institutions in Czechia. In IFLA. *Classification and indexing section newsletter* [online]. 2012, nr. 45, p. 8-11. Dostupné na www: <<http://www.ifla.org/files/assets/classification-and-indexing/newsletters/IFLA-CI-newsletter-45-June2012.pdf>>.

BALÍKOVÁ, Marie. UDC MRF Translation Tool : how does it work for Czech edition? In IFLA. *World Library and Information Congress : 78th IFLA General Conference and Assembly : UDC Update Session*. 12. August 2012. Presentace na konferenci. Dostupné na www: <<http://authority.nkp.cz/interpi/aktivity/souvisejici-informace/souvisejici-projekty/udc-mrf-translation-tool-how-does-it-work-for-czech-edition/view>>.

KUNT, Miroslav. Příprava nových základných pravidel pre spracúvanie archívnych dokumentov a zmeny archívnej legislatívy v ČR vo vzťahu k medzinárodným štandardom. In *ICARUS@work : medzinárodné štandardy a ich aplikácia v archívnictve*. Bratislava, 17.2.2012. Příspěvek na mezinárodním pracovní setkání. Dostupné na www: <www.nacr.cz/Z-Files/sc/2012/sc-2012-01.pdf>.

BALÍKOVÁ, Marie - SLAVIC, Aida. Linking the web of knowledge : UDC's role in bridging information services and resources. In *12. konference Archivy, knihovny, muzea v digitálním světě*. 2011. Příspěvek na konferenci. Dostupné na www: <<http://www.skipcr.cz/akce-a-projekty/dokumenty/akm-2011/Slavic.pdf>>.

BALÍKOVÁ, Marie - KUNT, Miroslav - ANDREJČÍKOVÁ, Nadežda - ŠUBOVÁ, Jana. Interoperabilita v paměťových institucích : vize a realita. In *12. konference Archivy, knihovny, muzea v digitálním světě*. 2011. Příspěvek na konferenci. Dostupné na www: <<http://www.skipcr.cz/akce-a-projekty/dokumenty/akm-2011/Balikova.pdf>>.

A.2 Vstupní data a cíl

Základním cílem projektu, který naplňuje tematickou prioritu TP 3.6 - Infrastruktura pro vědeckou a technologickou podporu zachování národní kulturní identity, je tvorba znalostní databáze paměťových institucí na bázi sémantické interoperability.

V roce 2012 jsme navázali na výsledky a činnosti projektu z předešlého roku a s ohledem na cíle stanovené v projektovém záměru pro první etapu (2011-2012) jsme se v rámci základního výzkumu zaměřili na:

1. tvorba návrhů, ověřování a doplňování autoritních termínů v rámci stávající databáze národních autorit jako jednoho ze základních předpokladů jednoznačné identifikace entit,

2. ověřování termínů z rejstříků Národního archivu, jejich doplnění, úpravu a testování harmonizace s národními autoritami,
3. analýzu dalších zdrojů termínů pro potřeby projektu a možnosti propojení regionálních a specifických faktografických a terminologických databází s projektem INTERPI,
4. návaznost projektu na koncepční a legislativní dokumenty a jiné národní projekty,
5. nástroje sémantického webu a nástroje propojení různých terminologických systémů,
6. vymezení tříd a podtříd entit,
7. definování funkčních požadavků na software,
8. doplnění a úpravu datové struktury pro zpracování všech tříd entit.

Požadavky odborné komunity si vyžádaly částečné změny v plánovaném postupu při řešení projektu především v souvislosti s definováním funkčních požadavků na software pro zpracování entit. Kromě plánovaného detailního zaměření na základní skupinu entit jsme museli přistoupit k definování funkčních požadavků v jiném rozsahu. Funkční požadavky na software byly řešeny komplexně. Na základě terminologického vymezení projektu a vymezení tříd entit, byly definovány požadavky na zpracování základních údajů všech tříd entit. Současně jsme se zaměřili na vývoj prototypu softwarového řešení, který umožní odborné komunitě lépe pochopit a prakticky zhodnotit požadavky na rozsah a strukturu údajů entit.

B Analytická část

Analytická část zprávy obsahuje informace o výsledcích dosažené v roce 2012 v oblasti základního výzkumu a také vývoje v souladu s cíli projektu. Základní výzkum navazoval na výsledky v roce 2011.

Soustředili jsme se na zdroje termínů, které budou tvořit obsahovou náplň projektu a znalostní databáze INTERPI. Za tři hlavní zdroje lze považovat bázi národních autorit (viz B.1), rejstříky archivů (především Národního archivu - viz B.2) a jiné regionální a specializované terminologické a faktografické databáze (viz B.3).

Jako další se v rámci základního výzkumu řešila otázka nástrojů sémantického webu (především linked data) a nástrojů propojení různých terminologických systémů (MDT).

Významnou část základního výzkumu představovalo také vymezení tříd a podtříd entit a ustanovení jejich označení pro potřeby projektu. V návaznosti na specifikaci funkčních požadavků na základní skupinu entit z roku 2011 a vytvoření datové struktury byly stanoveny funkční požadavky pro softwarové řešení a zpracován funkční koncept. Jak již bylo zmíněno výše předpokládaný rozsah funkčního konceptu byl rozšířen z původní základní skupiny entit na všechny entity. Také jsme přistoupili ke zpracování funkčního prototypu softwarového řešení.

B.1 Analýza, ověřování, doplňování a úprava dat

Cílem činnosti v této oblasti v roce 2012 bylo ověřit a doplnit údaje, které jsou vyžadovány pro projekt na základě vymezení potřeb paměťových institucí a z pohledu specifikace vlastností jednotlivých typů entit.

B.1.1 Úprava a doplnění záznamů věcných autorit

V roce 2012 se v bázi národních autorit ověřovaly a doplňovaly záznamy především geografických termínů, záznamy událostí typu válka a záznamy jmenných autorit.

Vzhledem ke kooperaci v rámci paměťových institucí bylo nutné provést úpravy zápisu některých údajů v bázi věcných autorit.

Zrušení indikátorů v poli 55X

V bázi věcných autorit byly zrušeny indikátory v polích 5XX, a to:

- první indikátor s hodnotou 1 označující podřazený termín a s hodnotou 9 označující nadřazený termín.

Zdůvodnění: jde o duplicitní vyjádření hodnoty podřazenost a nadřazenost nestandardním způsobem.

Více informací v příloze F.5.

Aplikace standardního pole pro zápis klasifikačního znaku MDT - pole 080

Regulérní pole pro zápis znaku klasifikačního systému MDT bylo připojeno ve formátu MARC v roce 2009¹. Pole 080 nahrazuje národní pole 089 a 099.

Více informací v příloze F.5.

B.1.2 Úprava a doplnění záznamů geografických autorit / entit ze třídy geografický objekt

Připojení pole 052

Pole obsahuje kód detailnějšího geografického třídění, které souvisí s autorizovaným selekčním prvkem zapsaným v poli 151. Uvádění detailnějšího geografického třídění v AUT záznamech je nezbytné pro kooperaci s ostatními paměťovými institucemi.

V poli 052 se zapisuje kód `cz_retro`, který je součástí databáze CZ_RETRO. Tato databáze zahrnuje všechna města, městečka a vesnice, která existují či existovala na území Čech, Moravy a Slezska, a rovněž jejich čtvrti a díly. Výběrově obsahuje i některé další lokality charakteru místních částí či samot (např. izolovaně situované hrady či dvory). Každá tato lokalita je jednoznačně identifikována pomocí kódu CZ. To je stabilní číselný kód v číselné řadě 1 až 99999².

Příklad použití

```
151 $a Albrechtice (Most, Česko : zaniklá obec)
052 7 $a 22 $d Albrechtice $2 cz_retro
151 $a Albrechtice nad Orlicí (Česko)
052 7 $a 26 $d Albrechtice $2 cz_retro
```

V kombinaci s polem 856 umožní jednoduchým způsobem obohatit záznam GEO autority o další významné informace.

Aplikací těchto nástrojů bude také zajištěna propojitelnost databáze národních autorit a znalostní databáze INTERPI s dalšími významnými projekty a databázemi, např.:

- <https://docs.google.com/file/d/0B-GMFFu9DT-pbnphWTZMdDBBMEU/edit>
- http://sovamm.wz.cz/o_kodech.htm

Více informací v příloze F.5.

Rozšíření údajů v poli 951

Geografické jméno označující kraj a okres bude rozšířeno o kvalifikátor označující stát- Česko, např.:

¹ <http://www.loc.gov/marc/authority/ad01x09x.htm>

² http://sovamm.wz.cz/o_kodech.htm

951 \$a Česko \$c Královéhradecký kraj \$d Rychnov nad Kněžnou : okres \$e Liberk
951 \$a Česko \$c Královéhradecký kraj (Česko) \$d Rychnov nad Kněžnou (Česko :
okres) \$e Liberk

Sjednocení zápisu doplňků/kvalifikátorů u geografických autorit / entit ze třídy geografický objekt

Pro současný stav je charakteristická značná nejednotnost v zápisu obecných doplňků/kvalifikátorů, které se zapisují:

- po čárce, např. Morava, řeka;
- po pomlčce, např. Karlovy Vary-oblast (Česko);
- po dvojtečce, např. Kažezské rybníky (Česko : přírodní památka).

Z důvodů jednoznačnosti geografických autorit i entit, které považujeme za geografický objekt, se bude uvádět obecný kvalifikátor vyjadřující typ geografické autority a entity vždy, nejenom v tzv. konfliktních případech, tj. tehdy slouží-li tato forma jména k identifikaci několika informačních objektů, např. „Morava“ označuje zemi, řeku, vinařskou oblast, albánské pohoří, název místní části města Desná, údolí na Marsu. Navíc toto jméno se používá i k identifikaci jiných informačních objektů (které se nepovažují za geografické autority), např. označuje hudební skupinu, palác v Brně, často bývá i jako příjmení součástí osobního jména.

V důsledku v současné době zřetelně deklarované a oprávněné potřeby strukturovaného, strojem zpracovatelného, algoritmizovaného zápisu jednotlivých částí selekčních prvků (přístupových bodů) je standardizovaný zápis jednotlivých komponentů selekčních prvků (přístupových bodů) i u geografických autorit řešen následovně:

1. Obecný kvalifikátor obsahující typ autority se bude u geografických autorit používat vždy a bude se zapisovat v kulatých závorkách
2. V kombinaci geografického a obecného kvalifikátoru, se obecný kvalifikátor zapisuje jako druhý v pořadí a je od geografického kvalifikátoru oddělen konvencí mezerou dvojtečkou mezerou
3. Zápis geografického kvalifikátoru se nemění, zapisuje se v závorkách, jednotlivé kvalifikátory jsou odděleny čárkou
4. Chronologický kvalifikátor se zapisuje v kulatých závorkách

Více informací v příloze F.5.

B.1.3 Záznamy tematických autorit / entit ze třídy událost - podtřída války, bitvy

V tomto roce řešení projektu bylo vyřešeno zařazení těchto entit do třídy entit události. V současné době pokračuje shromažďování návrhů autoritních záznamů ve speciálním souboru dostupném na portálu INTERPI v rámci Národních autorit. Je nutné vyřešit strukturu zápisu jednotlivých komponent selekčních prvků (přístupových bodů) téže třídy entit a sjednotit je se způsobem zápisu dalších entit (konferencí) zařazených do této třídy entit. Poté budou návrhy vřazeny jako regulérní autoritní záznamy do souboru národních věcných autorit.

Více informací v příloze F.8.

B.1.4 Záznamy geografických autorit / entit ze třídy geografický objekt - podtřída sídlo

Podobně jako v roce 2011 pokračují práce na pořizování návrhů geografických autorit / entit ze třídy geografický objekt - podtřída sídlo ze dvou hlavních zdrojů³. V případě jmen sídel v Čechách, resp. jejich historických jmen od středověku do 19. století se ukazuje náročnost jednoznačné identifikace jména uvedeného v citacích dobových pramenů (často latinsky a německy), kdy citace obsahuje celé nebo zkrácené věty, často s několika jmény, za ní následuje citace pramene atd. To vše znesnadňuje identifikaci jména dotčené obce z dané doby. Obdobně se také některá jména od doby, kdy byl zdroj (Profous, Antonín: Místní jména v Čechách) publikován, změnila. Z těchto důvodů byla práce - ve srovnání s rokem 2011 - rozdělena mezi více osob. Předpokládáme, že práce s pořizováním tohoto typu dat budou ukončeny v průběhu roku 2013.

Zpracování jmen sídel Slovenska podle vlastivědného slovníku z druhé poloviny 70. let⁴ bylo v roce 2011 dokončeno a připraveno pro databázi INTERPI. Další území bývalého Československa, Podkarpatská Rus, bylo zpracováno podle dobových lexikonů, takže zachycuje nejnověji stav z roku 1930.

Více informací v příloze F.6.

B.1.5 Záznamy geografických autorit / entit ze třídy geografický objekt - podtřída historický správní celek

Byla vypracována metodika pro tvorbu autoritních záznamů / záznamů entit ze třídy geografický objekt - podtřída historický správní celek, týkající se oblasti veřejné správy Česka (okresy, kraje) a historických území Německa potřebné termíny jsou postupně vřazovány do databáze Národních autorit.

Více informací v příloze F.5.

³ Profous, Antonín: Místní jména v Čechách. Jejich vznik, původní význam a změny. Nakladatelství Československé akademie věd, 1954-1960, díl I-IV + dodatky (V); Hosák, Ladislav - Šrámek, Rudolf: Místní jména na Moravě a ve Slezsku I, II. Academia Praha, 1970, 1980

⁴ Kol.: Vlastivedný slovník obcí na Slovensku I-III. VEDA, vydavateľstvo Slovenskej akadémie vied, Bratislava 1977-1978

Obr. 1: Ukázka zpracování historického správního celku

Záhlaví	Budějovický kraj (Česko : 1850-1855)	
Kód geogr. oblasti	e-xr--- e-xr-jc	
Odkaz. forma	kraj Budějovický (Česko : 1850-1855)	
Podřazený termin	Planá (Česko : okres : 1850-1855)	
	Krumlov (Česko : okres : 1850-1855)	
	Budějovice (Česko : okres : 1850-1855)	
	Sedlčany (Česko : okres : 1850-1855)	
	Sedlec-Prčice (Česko : okres : 1850-1855)	
	Soběslav (Česko : okres : 1850-1855)	
	Tábor (Česko : okres : 1850-1855)	
	Veselí nad Lužnicí (Česko : okres : 1850-1855)	
	Milevsko (Česko : okres : 1850-1855)	
	Mladá Vožice (Česko : okres : 1850-1855)	
	Neveklov (Česko : okres : 1850-1855)	
	Nová Bystřice (Česko : okres : 1850-1855)	
	Pacov (Česko : okres : 1850-1855)	
	Pelhřimov (Česko : okres : 1850-1855)	
	Počátky (Česko : okres : 1850-1855)	
	Nové Hradky (Česko : okres : 1850-1855)	
	Trhové Sviny (Česko : okres : 1850-1855)	
	Kanice (Česko : okres : 1850-1855)	
	Vyšší Brod (Česko : okres : 1850-1855)	
	Jindřichův Hradec (Česko : okres : 1850-1855)	
	Volice (Česko : okres : 1850-1855)	
	Bechyně (Česko : okres : 1850-1855)	
	Třeboň (Česko : okres : 1850-1855)	
	Týn nad Vltavou (Česko : okres : 1850-1855)	
	Hluboká (Česko : okres : 1850-1855)	
Asoc. termin	České Budějovice-oblast (Česko)	
	Budějovický kraj (Česko : 1855-1868)	
	Písecký kraj (Česko : 1855-1868)	
	Táborský kraj (Česko : 1855-1868)	
MDT	437.119	
Angl. ekvivalent	Budějovický kraj (Czechia : 1850-1855)	
	České Budějovice Region (Czechia : 1850-1855)	
GPS	48°58'26.1"N, 14°28'30.1"E	Zobrazit na mapě
Poznámka	Jmenný popis: používá se u korporací zapsaných pod správním celkem s působností pro Budějovický kraj v letech 1850-1855 . Věcný popis: používá se pro události odehrávající se na území Budějovického kraje v letech 1850-1855 . Jako geografické zpřesnění/podheslo se nepoužívá.	
Status	správní celek	
Další informace	 Wikipedie (Historie krajů v Česku)	

B.1.6 Úprava a doplnění záznamů jmenných autorit

V oblasti jmenných autorit se v roce 2012 řešily otázky vyplývající z požadavků jiných typů paměťových institucí týkající se způsobu zpracování anonymních tvůrců a šlechtických rodů (rodin).

B.1.6.1 Anonymní tvůrci

Pravidla AACR2 a RDA se věnují problematice neznámého tvůrce pouze okrajově. Personální záhlaví se vytváří jen v případě, kdy v popisovaném díle samotném nebo v příručkových pramenech existuje výraz, který by mohl být použit pro identifikaci autorství. Tímto výrazem může být buď nějaké charakterizující slovo nebo fráze (kromě neabecedních či nečíslných symbolů), nebo jím může být fráze uvádějící název jiného díla téže osoby. Ve všech ostatních případech se personální záhlaví nevytváří, hlavním záhlavím je název díla. Takovéto zpřístupnění je dostačující v případě knih či dalších tištěných dokumentů, k jejichž popisu jsou AACR2 i RDA primárně určeny, ale selhává v případě popisu děl jiné povahy (výtvarná díla a další předměty hmotné kultury).

CCO se v případě tzv. anonymního tvůrce (anonymous creator) s oběma knihovnickými pravidly shoduje, ale v případě tzv. neznámého tvůrce (unknown creator) přináší něco navíc. Záhlaví tohoto typu AACR2 ani RDA neznají. Knihy a další tištěné dokumenty jsou v jistém slova smyslu „samopopisující“. Mají název, který je možné použít pro jejich zpřístupnění, pokud je jejich tvůrce neznámý. Předměty hmotné kultury tuto vlastnost nemají. Model uměle vytvořených záhlaví („neznámý“ + kulturní (národnostní) příslušnost + široce pojatá data) jsou určitým pokusem nenechat pole autorské odpovědnosti prázdné.

Více informací v příloze F.5.

B.1.6.2 Šlechtické rody

Pro potřeby zpracování šlechtických rodů byly stanoveny zdroje doporučené k verifikaci jmen rodů. V současné době lze nalézt dostatečné množství literatury slovníkové povahy, v konkrétních případech ovšem nelze vyloučit, že bude nutno vytvářet záznamy také na podkladě konkrétní monografie, zejména tvorba autoritních forem pro rody v období středověku je problém *sui generis*. V případě tvorby autoritních záznamů u nečeských rodů, je třeba sledovat a analyzovat způsob zpracování v příslušných národních bázích a tato zjištění následně zohlednit při tvorbě a přebírání těchto záznamů. Současná podoba tvorby autorit pro šlechtické rody nezahrnuje rozlišení rodových linií, které se neprojevovaly do názvu rodu (např. orlická větev rodu Schwarzenbergů atd.), bude zřejmě nutné zpracovat metodiku pro tento případ.

Stávající záznamy s ohledem na knihovnické potřeby nezpracovávají vazby ani na členy rodu, ani na rozrody či alianční následnictví. Zatím ani jinde v knihovnách nelze najít záznamy rodů zpracovaných formou odkazů (provazeb, linků) vzájemně mezi sebou, pouze formou poznámky v neselekčních polích. Členění na jednotlivé rozrody následně může způsobovat problémy z hlediska správnosti přiřazování k jednotlivým popisovaným zdrojům. Je třeba také dobře rozvažovat tvorbu jednotlivých autoritních záznamů na základě jediného zdroje bez jeho důkladné analýzy (už ze samotného podtitulu je jasné, že publikace je věnována celému rodu erbu vrchních pruhů, zpracování poznámky formou obecného konstatování typu český či moravský šlechtický rod, má nulové využití - rodové majetky byly často v průběhu času na různých místech).

Doporučované řešení - spojit jednotlivé rody s polem geografická vazba (sídlo rezidence, název rozrodu dle určité lokality, používat geografické autority, v poznámce se snažit určit dobu působení konkrétního šlechtického rodu a vytvořit okruh vazeb k předchůdcům či pokračovatelům - možnost vytvářet řady typu: Předchůdce: Ronovci - Pokračovatel: Lichtenburkové, Berkové z Dubé, z Lipé - model zpracování vazeb u tiskařských dílen, či korporací; nebo formou hierarchizace Nadřazený termín: Ronovci - Podřazený termín: Lichtenburkové, Berkové z Dubé, z Lipé - model zpracování věcných autorit, není zcela kompatibilní se souborem národních jmenných autorit, kam se zpracování rodů a rodin v rámci AUT řadí)

Způsob zpracování - v mnohem větší míře než dosud používat odkazové varianty, většina šlechtických rodů působících na našem území používala 2-3 varianty svého jména (minimalizovat pravděpodobnost vzniku duplicit - v případě použití provazeb mezi jednotlivými záznamy je odstraňování duplicit navíc spojeno s opravami provázaných záznamů, časová náročnost tedy stoupá geometrickou řadou)

Vazba mezi šlechtickým rodem a konkrétními jednotlivci - **doporučované řešení** - propojovat pouze jedince s konkrétním rodem (jednostranná vazba, do souboru AUT by se toto pole nedostávalo), při zakládání personálních autorit ponecháno na volbě katalogizátora, zda bude zároveň zakládat záznam na konkrétní šlechtický rod.

Více informací v příloze F.5.

B.2 Termíny z rejstříků Národního archivu

V oblasti terminů z rejstříků archivů (především Národního archivu) spočívala hlavní pozornost v analýze obsahu terminů a jejich struktury.

B.2.1 Korporativní entita - spolky

Pro účely budoucího zpracování entit korporace typu spolek je upravován soupis spolků (jména spolků), který vznikl v Národním archivu v minulých 20 letech. Data ve stávající podobě nebyla použitelná, proto se přikročilo k jejich čištění. V této souvislosti byl doplněn i funkční model v případě korporací o data působnosti. U spolků totiž fakticky nelze stanovit ani jejich vznik, nehledě na nemožnost definování tohoto okamžiku v různých historických obdobích (je jím ustavující schůze, schválení stanov, ustavení přípravného výboru apod.).

B.2.2 Záznamy o původcích typu korporace

V průběhu roku probíhaly práce na pořizování záznamů o původcích - částečně na základě prací na Základních pravidlech pro zpracování archiválií, částečně za účelem praktického odzkoušení datové struktury budoucí databáze INTERPI a jejího funkčního modelu. Tak byly zpracovány podrobné záznamy o regionálních korporacích různých typů z oblasti východních Čech, současných korporacích v oblasti státní správy. Současně byly shromážděny podklady pro další práci (zejména korporace 1848-1918), jejich zpracování je však závislé na stabilizaci funkčního konceptu a datové struktury, protože pozdější úpravy by byly neefektivní.

V souvislosti s přípravou různých souborů dat k napojení na sdílenou databázi INTERPI byl proveden předběžný zkušební test kompatibility rejstříkových hesel archivních pomůcek Národního archivu s dosavadní databází národních autorit ČR s cílem umožnit harmonizaci s národními autoritami. V návaznosti na předešlý rok projektu se hledaly společné a rozdílné prvky terminů.

B.2.3 Komparace rejstříků archivních pomůcek Národního archivu s bází národních autorit

Zadání předpokládalo následující postup:

- strojově porovnávaná vstupní data jsou v MS Excell nebo databázi MS SQL 2007. Obsahují identifikaci a heslo - slovo nebo skupinu slov (textový, resp. alfanumerický řetězec). Pro účely snazšího porovnání v následných krocích (uživatel) zadavatel vstupní data rozšíří o další informace (popis archiválií, ke kterým se data váží);
- referenční data jsou dostupná pomocí protokolu Z 39.50 z databáze Národních autorit ČR (viz <http://autority.nkp.cz/zakladni-informace/pristup-do-baze-autorit-pres-z39.50>);
- z referenčních dat se zpracovávají všechny podoby jména (preferovaná i nepreferovaná podoba);
- cílem je porovnat zmíněná hesla ze vstupních dat s hesly dat referenčních, najít úplnou shodu, částečnou shodu nebo podobnost (např. na základě začátku jednotlivých slov) a nabídnout párování vstupních a referenčních dat k odsouhlasení zadavateli;

- po odsouhlasení jednotlivých propojení vstupních a referenčních dat pracovníkem zadavatele dojde k uložení informace o referenčních datech do dat vstupních tak, aby vazba mezi porovnávanými a referenčními daty byla jednoznačná a trvalá (identifikátor i řetězec-jméno);
- nástroj nebo skripty, které vzniknou, budou (pokud vyhovují definici autorského díla) licencovány jako Open-source a zpřístupněny zadavateli, který je oprávněn s nimi v rámci licence Open-source nakládat.

Termíny z rejstříků byly rozděleny do těchto skupin:

1. tematická hesla
2. jména osob
3. geografická jména
4. korporace

Termíny v každé skupině byly analyzovány a pro každou skupinu termínů byl stanoven postup pro hledání odpovídajících autorit v bázi národních autorit. Očekávaným výsledkem byl soubor termínů doplněný o informace o autoritě nebo autoritách, které termínu potenciálně odpovídají. Byly doplněny informace: záhlaví, kód záznamu v bázi národních autorit a případně poznámka (podle toho, o kterou skupinu termínů se jednalo). Součástí byl také výraz v selekčním jazyku PQF, který se použil pro vyhledání.

Jako hlavní problém byl při vyhledávání prostřednictvím protokolu Z39.50 hned v úvodě stanoven obsah indexů pro vyhledávání. V bázi národních autorit není dostupný index, který by obsahoval pouze záhlaví. Současné indexy jak pro jmenné, tak pro věcné autority obsahují všechny odkazované formy jména. Při této struktuře indexu je velmi obtížné hledat termíny podle jednotlivých slov, protože mezi výsledky jsou zařazeny také autority, které jedno z hledaných slov obsahují v záhlaví a druhé v odkazované formě. Hledání podle jednotlivých slov by bylo výhodné především v případě tematických hesel z rejstříků Národního archivu, kde se používá invertovaný zápis hesel.

Dalším problémem byla skutečnost, že indexy pro osobní jména a korporace odkazují do společného indexu jmenných autorit a indexy pro geografické, věcné autority a autority pro formu a žánr odkazují do společného indexu věcných autorit, přesto že všechny mají přidělené samostatné atributy protokolu Z39.50. Výsledkem na dotaz na osobní jméno může být ve výjimečných případech autorita korporace podobně jako na dotaz na geografické jméno autorita věcná.

Odhalení těchto problémů je důležité pro vymezení postupů pro spolupráci jiných paměťových institucí.

Navržené propojení na národní autority se bude postupně posuzovat pro každou skupinu termínů individuálně. Na základě výsledků se bude přistupovat k dalším úpravám dat. Z výsledku v následujících podkapitolách plyne především již dříve předpokládaná skutečnost, že největší aktuální význam z hlediska interoperability mají geografické entity (zeměpisný rejstřík), kde lze docílit automatizovaného propojení. Dosavadní nedostatek, nedostatečná identifikace geografické entity pouze záhlavím bude ještě v databázi NAČR odstraněna uváděním hierarchického začlenění entity do okresu, kraje atd. (týká se sídel).

Naopak u entity typu osoba je propojení obtížné zejména z důvodu potenciální chybné identifikace osoby. Ukazuje se tak nezbytné uvádět bližší charakteristiku důsledně jak v archivních pomůckách, tak v databázi a používat tyto údaje k porovnání. Překvapením bylo nízké procento propojených jmen korporací. To lze vysvětlit tím, že zejména historické korporativní entity nejsou v databázi národních autorit dostatečně zastoupeny a také stejnými problémy jako u věcného rejstříku: formalizace je jiná než v databázi národních autorit, zpravidla je heslo doplněno podheslem (např. „OSN - Evropská hospodářská komise“).

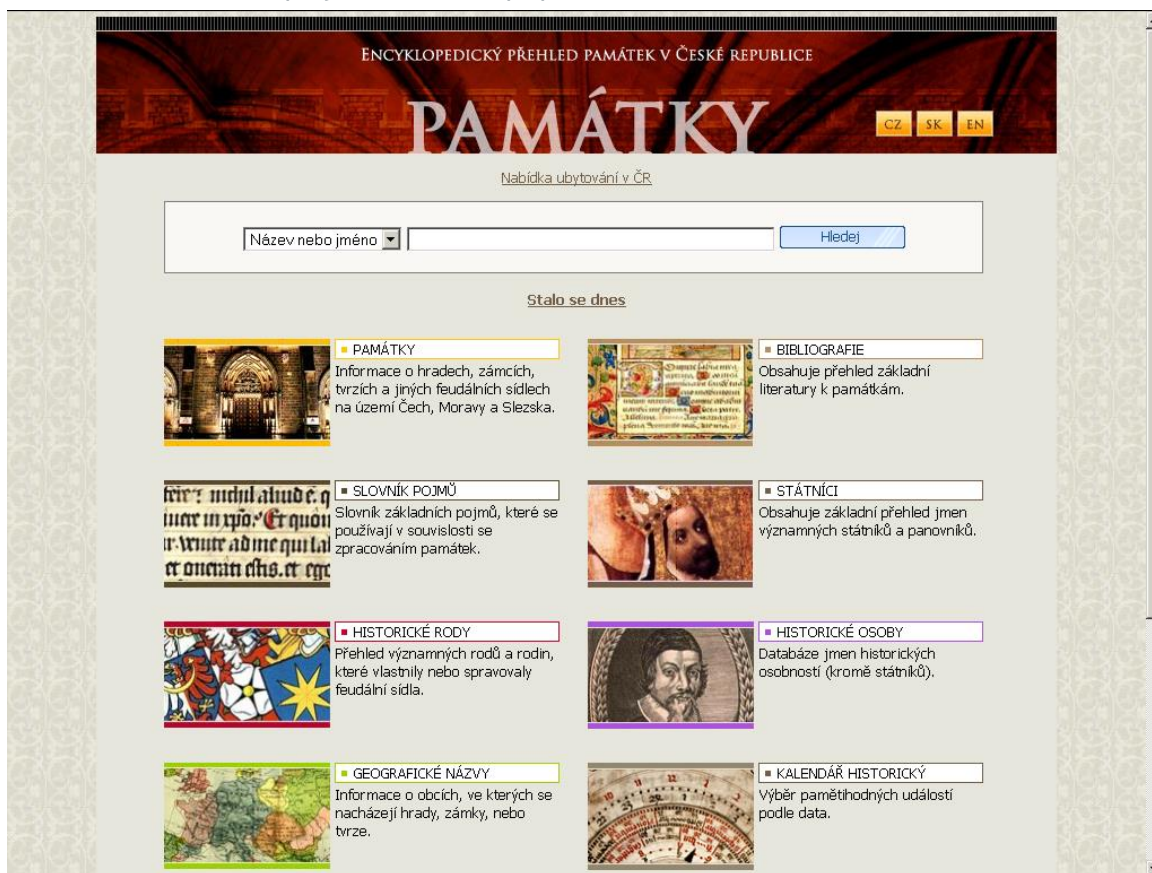
Více informací v příloze F.6.

B.3 Další zdroje termínů

Tato kapitola obsahuje požadavky ostatních paměťových institucí ve formě, ve které byly definovány zástupci těchto paměťových institucí. Cílem bylo zjistit reálné současné požadavky těchto paměťových institucí na vedlejší skupinu entit, jejímiž parametry se budeme v rámci projektu INTERPI zabývat v příštím roce. Sladění všech těchto požadavků, jejich transformace a namapování na definované třídy a podtřídy entit bude tématem příštího roku. Tyto definované požadavky zástupců jednotlivých paměťových institucí jsou důvodem a zároveň dokladem toho, proč bylo nutné částečně pozměnit a upravit původně plánovaný harmonogram projektu a specifikovat třídy a podtřídy základní a vedlejší skupiny entit a realizovat vývoj dílčí verze prototypu již v tomto roce (viz část A.2).

B.3.1 Nemovité památky a souvisejí termíny

V roce 2012 byly do projektu INTERPI zapojeny také informace z projektu www.soupispamatek.cz, který je zaměřen na nemovité památky hradiště, hrady, tvrze, zámky, paláce, letohrádky a pod. na území Česka.

Obr. 2: Titulní stránka projektu www.soupispamatek.cz

Tento zdroj byl zvolen pro svoji komplexnost, protože pokrývá existující i zaniklé nemovité památky vyjmenovaných typů z území Česka a současně poskytuje relativně komplexní údaje o každé památce (minimálně název, variantní názvy, lokaci, stav, přístupnost a typologizaci památky). Projekt kromě informací o nemovitých památkách (celkem 6502 záznamů) zpracovává faktografii šlechtických rodů a rodin (232 záznamů), státníků a panovníků (76 záznamů), historických osobností (259 záznamů), informace o geografických lokalitách (celkem 1602 záznamů) a terminologii (celkem 2457 záznamů) v souvislosti s památkami.

Hlavní činnost v souvislosti s projektem www.soupispamatek.cz spočívala v úpravě a doplňování dat. Především byly k památkám doplňovány lokační údaje (původní koncept lokalizace památky pouze pomocí okresu v případě památky ležící mimo katastru obce byl opuštěn a doplňovány byly údaje o nejbližší obci). Vznikaly tak také nové záznamy obcí, místních částí a zaniklých obcí. Perspektivně se počítá s harmonizací záznamů geografických objektů se záznamy v bázi národních autorit. Záznamy památek byly také doplněny o faktografické informace a obrazovou dokumentaci.

Současně byly doplňovány záznamy do tzv. slovníku - tj. záznamy pojmů z oblasti památkové péče. V tomto směru může být projekt v dalších etapách zajímavým terminologickým zdrojem. Pro úplnost lze zmínit také tzv. kalendárium, což je samostatná báze událostí, které se vážou ke konkrétnímu datu.

Přehled změn v datech projektu www.soupispamatek.cz je v následující tabulce. Z celkového počtu upravených záznamů tvořily cca 2/3 úpravy faktografických informací a 1/3 úpravy struktury dat (včetně doplnění zmíněné lokace, preferovaného označení památky a pod.).

Tab. 1: Projekt www.soupispamatek.cz - statistika změn v záznamech v roce 2012

Typ záznamu	Počet záznamů	
	nových	upravených
bibliografické záznamy	116	19
památky	842	5683
geografické názvy	282	
osobní jména	92	
rody	14	
kalendárium	419	
slovník	1925	

V čase zpracování této zprávy probíhá analýza datové struktury, metodiky zpracování, informačního potenciálu a technologických možností propojení a využití dat z projektu www.soupispamatek.cz v projektu INTERPI.

B.3.2 Tematické autority používané v Registru sbírek výtvarného umění a v Muzejních autoritách

V Muzejních autoritách a v Registru sbírek výtvarného umění převládají termíny z oblasti umění. Jde především o termíny představující názvy uměleckých směrů, hnutí a proudů, názvy uměleckých skupin, názvy označující tendence, principy a tvůrčí pojetí projevující se ve výtvarném umění. Vyskytují se zde i termíny pro umění určitých oblastí, termíny označující typické výtvarné pojmy, témata zobrazení, ikonografické pojmy, typy uměleckých děl, termíny označující výtvarné techniky a názvy výrobku podle materiálu, pojmy související se vznikem uměleckého díla, pojmy z oblasti užitého umění a termíny příbuzných oborů

Vyskytují se zde termíny obecné, oborové, úzce odborné (specifické pro určitý směr daného oboru), příležitostné a individuální. Velmi často se používají termíny spojené s tvorbou jednoho umělce tvořené příponou -ismus.

V Registru sbírek výtvarného umění a v Muzejních autoritách se projevuje tendence častěji zařazovat termíny pro umění určitých oblastí, např. „olomoucké baroko“. Tyto termíny podtrhují význam umění v dané oblasti, zařazují toto umění jako svébytný kulturní fenomén a představují v uvedených systémech důležitý řadící a jednotící prvek pro celou skupinu umělců a dalších entit svázaných s tímto fenoménem.

Z hlediska univerzálního slovníku souboru věcných autorit se mnohdy jedná o termíny situační charakterizující specifický fenomén, často související s jedním informačním zdrojem a s lokálním prostředím. V těchto případech je nutné provést důkladné rešerše, zkoumat výskyt termínu v uměleckých oborech úzce souvisejících atp.

Této problematice, tj. teoretické přípravě napojování terminologie oborových systémů na univerzální databázi INTERPI, bude v příštím roce věnována zvýšená pozornost, aby byla umožněna pokud možno bezproblémová integrace těchto oborových systémů.

Praktická ukázka termínu specifického pro danou oblast

„Olomoucké baroko“

Výstava Olomoucké baroko | Výtvarná kultura let 1620-1780 (2010/11) je největším projektem v historii olomouckého muzejnictví, jehož příprava trvala více než pět let. Dokumentuje jedno z nejvýraznějších období kulturního vzepětí města Olomouce a regionu. Doprovází ji vícesvazková publikace, na níž se podílelo více než osm desítek odborníků z různých oborů. Cílem projektu je mimo jiné rovněž určení pojmu „olomoucké baroko“ jako svébytného kulturního fenoménu.

V databázi MA je k pojmu přiřazeno zatím 53 záznamů. Téma se dále zpracovává.

Více informací v příloze F.7.

B.3.3 Věcné/tematické entity (pojmy, objekty) v projektu INTERPI z pohledu regionálních muzeí

Regionální muzea dokumentují určitý region, určitou oblast. V podmínkách Česka mohou být tyto oblasti různě velké, od obce, města, přes okres, kraj až po stát. Jiným pojmem pro regionální muzeum je muzeum vlastivědné. Vlastivědná muzea jsou zdaleka nejpočetnější skupinou muzeí. Sbírky těchto muzeí pokrývají prakticky veškeré přírodní i humanitní obory.

Z těchto premis vyplývá, že požadavky vlastivědných muzeí na tematické autority, které by měly stejně jako ostatní autority sloužit jako jednoznačné selekční prvky (přístupové body) při běžné katalogizační praxi i při zpětném vyhledávání, musí být velmi komplexní, řekněme maximalistické.

Pojmy

Pod „pojmy“ můžeme - dle názoru některých paměťových institucí - zařadit vše mimo osobní jména, geografická jména, objekty a události.

Patří sem méně početné skupiny jako abstraktní pojmy (smutek, radost), ikonografické výjevy (Poslední soud) či vědní a umělecké obory; a dále především - z pohledu muzejního zcela nejpodstatnější skupina - názvy předmětů.

Některý „pojem“ lze jednoznačně charakterizovat a systematizovat menším počtem kategorií informací, u jiného pojmu je nutné užít více kategorií. Pro pokrytí veškerých potřeb by měly být součástí struktury záznamu tematické autority následující kategorie informací:

- název
- nadřazené a podřazené termíny
- synonyma
- oborové zařazení
- charakteristika
- neverbální informace
- weblink

Objekty

Pod „objekty“ můžeme - dle názoru některých paměťových - zařadit stavby v nejširším slova smyslu - od měst, silnic či domů až po sochy na náměstích nebo drobná boží muka v krajině.

Objekty jsou v současných autoritách zařazeny do dvou skupin - obce jsou v geoautoritách (nejsou chápány jako „objekty“), vše ostatní je v tematických autoritách, příp. jmenných(?). Myslím, že všechny objekty by měly být v geoautoritách. Důvodem je zejména jednotnost zápisu, jednotnost druhů informací. Tytéž druhy informací potřebujeme k městu i k hradu.

Některý „objekt“ lze jednoznačně charakterizovat a systematizovat menším počtem kategorií informací, u jiného pojmu je nutné užít více kategorií. Pro pokrytí veškerých potřeb by měly být součástí struktury záznamu objektu (tedy geoautority) následující kategorie informací:

- název
- nadřazené a podřazené termíny - územní jednotky
- synonyma - historické názvy
- oborové zařazení
- charakteristika
- geolink

Více informací v příloze F.7.

B.4 Nástroje sémantického webu a nástroje propojení různých terminologických systémů

Rozvoj informačních a komunikačních technologií jde stále rychle vpřed a dnes se již ukazuje potřeba spolupráce nejenom na národní a mezinárodní úrovni knihoven, ale také potřeba mezioborové spolupráce.

V současné době, kdy téměř všechna data knihoven jsou přístupná také v prostředí webu, by se mohlo zdát, že všechny problémy s dostupností dat jsou již vyřešeny. Ukazuje se však, že vyvstává jiný mnohem závažnější problém, tj. problém se strojovým zpracováním těchto dat publikovaných na webu.

Tuto potřebu identifikoval jeden z tvůrců samotného webu již v prvních letech jeho existence, jako potřebu publikovat nejenom samotné informace, ale informace rozšířené o jejich význam umožňující jejich vnímání a sdílení nejenom lidmi ale i vyhledávacími stroji, čímž vyslovil potřebu vzniku tzv. „sémantického webu“. Postupně vznikalo množství iniciativ, které byly později označeny jako Web 2.0. Ten však nenaplnil veškerá očekávání sémantického webu, protože jde spíše o sociální web, který podstatně usnadnil publikování dalších informací a tím zároveň zvýšil i potřebu strojového zpracování informací publikovaných na webu.

Důležitou roli v standardizaci prostředí webu sehrává konsorcium W3C. Jde o mezinárodní organizaci na členském principu se stálými zaměstnanci, kteří spolupracují na vytváření standardů pro web. Standardizační aktivity W3C zasáhly také problematiku sémantického webu a zároveň je W3C organizací, která popisuje a svým způsobem usměrňuje vývojové tendence webu.

Zdá se, že sémantický web s jeho technologiemi pro propojená data se vynořil naprosto nečekaně a nenadále vystřelil odnikud a přerušil veškeré aktivity v oblasti metadat. Knihovnický svět s jeho dobře propracovanými a relativně stabilizovanými postupy v oblasti metadat nevyjímá. Knihovní katalogy ztrácejí své výjimečné postavení v procesu vyhledávání, které nyní probíhá současně v různých katalogích i mimo ně. Data z katalogů se přesouvají z knihovních databází na web „propojených dat“ (web of linked data). Data z katalogů mohou odkazovat na jiné zdroje a obráceně, což znamená, že data z katalogů jsou vyhledatelná z různých míst na webu. Uživatel nemá k dispozici pouze jeden přístupový bod ke katalogům (kontrolován knihovnou), naopak data zveřejněná v knihovních katalogích se včleňují do aktuálního prostředí, ve kterém se uživatel pohybuje⁵.

V rámci aktivit zaměřených na sémantický web jsme zkoumali technologie sémantického webu a hledali příklady aplikací koncepce linked data na údaje věcného přístupu. Součástí bylo také zkoumání možností MDT jako nástroje pro propojení různých terminologických systémů.

B.4.1 Linked data

Termín linked data (propojená data) použil v souvislosti se sémantickým webem jako první Tim Berners-Lee. Koncept propojených dat představuje soubor metod k publikování a vzájemnému propojení strukturovaných dat v prostředí webu s cílem zvýšit efektivitu jejich využitelnosti. Tento způsob publikování staví na standardních webových technologiích, jako je protokol HTTP pro zabezpečení přístupu, HTML formát pro zpřístupnění obsahu a unikátní identifikátor URI. Na rozdíl od klasických webových stránek, které umožňují publikovat a propojovat dokumenty určené pro čtení lidmi, koncept propojených dat rozšiřuje možnosti sdílení informací tím, že stroje jsou schopny tyto informace číst, rozumět jim a mohou je tak i zpracovávat.

Tim Berners-Lee definoval 4 základní principy pro propojené data⁶:

1. využívání URI jako jména pro objekty,
2. využívání HTTP URI, to znamená, že uživatelé mohou jména objektů vyhledat,
3. pokud někdo URI vyhledá, poskytnout užitečné informace za využití standardů (RDF, SPARQL),
4. doplnit odkazy na jiné URI - takže uživatelé mohou vyhledávat a objevovat další věci (objekty).

Propojená data vs. propojená otevřená data (linked data vs. linked open data)

Koncept propojených dat (linked data) se příliš neodlišuje od konceptu propojených otevřených dat (linked open data). Hlavní rozdíl je v tom, že propojená otevřená data jsou publikována pod tzv. otevřenou licenci - to znamená, že jsou volně použitelná při dodržení daných podmínek. Princip otevřených licencí (open licence) vychází z vymezení autorských práv ve vztahu

⁵ COYLE, Karen. Linked Data Tools : connecting on the Web : introduction. In *Library technology reports*. May/June 2012, p. 7-8.

⁶ <http://www.w3.org/DesignIssues/LinkedData.html>

k různým typům děl a též k webovým datům a publikování na webu. V obecné rovině poskytuje princip otevřené licence povolení k přístupu k dílu (dokumentu, softwaru atd.), jeho opakované využití, distribuci s malými nebo žádnými omezeními.

Je ale potřeba si uvědomit, že ne každá propojená data jsou zároveň otevřenými a obráceně, ne každá otevřená data jsou současně propojenými.

Místo věcných autorit v sémantickém webu

Existuje několik příkladů, na kterých lze prezentovat zveřejnění termínů věcného zpracování (nebo samotných souborů věcných autorit) v podobě propojených dat (linked data). Tyto projekty jsou důkazem, že termíny věcného přístupu mohou být významným soběstačným zdrojem a zároveň nástrojem pro kontrolu obsahu jiných zdrojů. Pomocí slovníků je možné strojově kontrolovat platnost termínů, doplnit termíny o jejich význam a také harmonizovat termíny jedné komunity s termíny používanými v jiné komunitě.

Věcné autority se v sémantickém webu uplatňují jako kontrolované slovníky nejen pro propojení termínů různých odborných komunit, ale také pro propojení termínů různých jazykových oblastí. Vyhledávání může probíhat v přirozeném jazyku uživatele bez ohledu na původní jazyk zdroje.

Projekty pro prezentaci a publikování termínů věcného přístupu jako propojených dat (linked data)⁷ je možné rozdělit do několika skupin:

1. tvořené mimo působnost knihoven:
 - 1.1. obecné slovníky a tezaury (např. DBpedia Ontology, New York Times Subjects),
 - 1.2. seznamy (např. GeoNames, BBC Wildlife Ontology);
2. tvořené knihovnami:
 - 2.1. soubory věcných autorit (např. Library of Congress Subject Headings, REMEAU, Deutsche Nationalbibliothek Subject Heading, National Diet Library List of Subject Headings),
 - 2.2. soubory jmenných autorit (např. Virtual International Authority File, REMEAU),
 - 2.3. seznamy (např. MARC Term Lists, RDA Vocabularies),
 - 2.4. specializované tezaury (např. National Agriculture Library, AGROVOC),
 - 2.5. klasifikační schémata (např. Dewey Summaries, Universal Decimal Classification).

Více informací v příloze F.8.

B.4.2 Potenciální aplikace klasifikačního systému MDT jako nástroje propojení různých systémů

V tomto roce (2012) řešení projektu jsme se zabývali i teoretickým průzkumem a potřebnou úpravu klasifikačního systému MDT jako potenciálního propojovacího jazyka při řešení

⁷ COYLE, Karen. Linked Data Tools : connecting on the Web : vocabularies. In *Library technology reports*. May/June 2012, p. 27-35.

terminologické různorodosti termínů aplikovaných v jednotlivých paměťových institucích a jako nástroje pro tvorbu ontologií.

V roce 2012 byla vytvořena v kooperaci NK ČR a UDC Konsorcia v Haagu česká verze databáze pro Kapesní vydání, v současné době probíhá překlad a potřebné úpravy záznamů MDT. Probíhají práce na tvorbě abecedního rejstříku k této bázi.

V roce 2012 byl také učiněn významný krok z pohledu prezentace a aplikace knihovních údajů v prostředí sémantického webu: byl zveřejněn systém MDT ve formě linked open data (prozatím výběr 2 500 znaků ve 48 jazycích). Dále probíhá aktualizace a překlad standardní úplné verze systému MDT.

Všechny práce spojené s překladem a aktualizací jednotlivých verzí klasifikačního systému MDT a s publikací klasifikačního systému MDT probíhají v úzké kooperaci s UDC Konsorciem v Haagu, které vytváří prostředí pro jejich realizaci, např. vývoj a hosting a údržbu multilingvární překladové databáze (pro různé verze vydání MDT - UDC Summary, UDC pocket edition, UDC MRF), porovnání a propojení různých verzí národního souboru klasifikačních znaků MDT - v našem případě porovnání verzí z let 2002 a 2003 s verzemi z let 2010 a 2012, publikaci klasifikačního systému jako linked open data (LOD) apod.

Více informací v příloze F.5.

B.5 Návaznost projektu na koncepční dokumenty a jiné národní projekty

B.5.1 Příprava metodických a metodologických postupů, jejich zavedení v archivech

Rok 2012 byl rokem novelizace právních předpisů v oblasti archivnictví a spisové služby. Hlavním důvodem změn (kromě opravy chyb) bylo nastavení pravidel pro elektronickou archivaci v návaznosti na projekt národního digitálního archivu. Postup projektu INTERPI se přímo promítal do následujících činností, které INTERPI zpětně ovlivňovaly.

Zadávací dokumentace technologie národního digitálního archivu

Jedním z úkolů celostátního archivního portálu je správa informací o původcích (entity třídy osoba nebo korporace). Proto bylo již při tvorbě zadávací dokumentace zajištěno, aby výsledky projektu INTERPI byly v oblasti archivů aplikovány (že tomu tak není v případě projektu národní digitální knihovny, lze litovat).

Novela zákona 499/2004 Sb., o archivnictví a spisové službě ve znění pozdějších předpisů

Do zákona byla při přípravě novely včleněna ustanovení o evidenci původců, z nichž nejvýznamnější je § 18c odst. 3, který kodifikuje povinný popis původců v elektronické podobě jednotlivými archivy. „Popis a evidence původců“ je postavena na roveň (paralelně vedle ní) dosavadní evidenci Národního archivního dědictví. Při aplikaci tohoto ustanovení je zamýšleno

užívání databáze INTERPI, pro niž komunikaci bude zprostředkovávat celostátní archivní portál, jak je zřejmé z citovaných ustanovení zadávací dokumentace národního digitálního archivu.

Novela vyhlášky 645/2004 Sb., kterou se provádí zákon o archivnictví a spisové službě

Novelizovaná vyhláška uvádí podrobnosti k zákonem stanoveným povinnostem. Příprava těchto ustanovení byla již plně navázána na práce v rámci projektu INTERPI. V souladu s legislativními pravidly, zvyklostmi a postupy nebylo však možné některé pojmy vhodně upravit. Jedná se zejména o následující problémy:

1. u osob při popisu původců vypuštěn na zásah ve vnitrorezortním připomínkovém řízení „životopis“, který ovšem zůstal jako povinná součást úvodu inventáře,
2. nebylo možné použít pojem „rod“ (nežádoucí vazba na občanský zákoník a zákon o rodině) ani „korporace“,
3. při legislativně-technické úpravě došlo k dodefinování „sídla“ na rozsah celé adresy, i když to nebylo původním úmyslem.

Přes zmíněné nedostatky se podařilo prosadit úpravu, která zaručuje řádné pořízení informací o původcích a je kompatibilní s INTERPI. Zcela novým způsobem byly definovány částečně korporace, přičemž byl opuštěn v dosavadní archivní legislativě přítomný koncept pouze právnických osob: „u archiválií, jejichž původcem je skupina osob vystupujících pod stejným názvem a sdružených k dosažení shodného účelu“.

Návrh nových Základních pravidel pro zpracování archivního materiálu

Nová Základní pravidla pro zpracování archiválií fakticky znamenají po 50 letech převrat v metodice základních archivních činností s výjimkou předarchivní péče. Jejich koncepce vychází, jak bylo naznačeno v loňské průběžné zprávě, z mezinárodních archivních standardů, především ISAD(G) a ISAAR(CPF). Právě druhý jmenovaný standard se stal základem pro popis původců archivních fondů, který se promítl i do legislativy (viz výše). Zároveň musela být stanovena pravidla pro tvorbu názvů - a to nejen v případě základních entit.

Práce na Základních pravidlech pro zpracování archiválií byly v případě kapitol „popis původců“ a „tvorba názvů a jmen“ koordinovány a konzultovány i v rámci projektu INTERPI. Tyto konzultace významně přispěly k definici entit (nejen základních) a také k zamyšlení nad udržitelností či neudržitelností některých metodických postulátů ať již v oblasti archivnictví nebo knihovnictví.

Protože se jedná o základní metodiku pro archivy, byla její přípravě věnována velká pozornost.

B.5.2 Návaznost projektu na Koncepti rozvoje knihoven a na projekt CPK

Jednou z priorit Koncepce rozvoje knihoven pro roky 2011-2015 je i vytvoření jednotného rozhraní systému knihoven s cílem zprostředkovat všechny nabízené služby z každé knihovny. Toto silné centrální rozhraní je prezentováno pod názvem Centrální portál českých knihoven (CPK) - www.knihovny.cz.

Za jakéhosi předchůdce jednotného rozhraní CPK, které je cílem této priority, je možné chápat Jednotnou informační bránu (www.jib.cz). Jednotná informační brána (JIB) integruje řadu českých i zahraničních volných i licencovaných zdrojů a umožňuje paralelní vyhledávání v heterogenních zdrojích založené na technologii Z39.50.

Projekt CPK chce stavět na technologiích pro "web-scale discovery systems", které jsou uživatelsky přívětivé, respektují požadavky Webu 2.0 a Webu 3.0 a především v jednom rozhraní integrují pokud možno všechny zdroje knihovny, počínaje knihovními katalogy a konče elektronickými zdroji.

Mezi požadavky projektu CPK⁸ patří také vyhledávání pomocí ontologií (existence nástrojů pro rozlišení shodných jmen, názvů...), kumulace návrhů na doplnění/úpravy ontologií, využití vícejazyčných ontologií, vizualizace klíčových slov, našeptávač a pod.

V této souvislosti lze v projektu CPK najít uplatnění výsledků projektu INTERPI pro našeptávače, tvorbu faset, doporučování souvisejících dokumentů nebo zdrojů. Svou roli hrají i při vypočítávání relevance nalezeného záznamu, kde se určuje, která pole mají pro sestavení míry relevance větší váhu a ověřená metadata mívají zpravidla vyšší váhu než např. informace nalezené v plném textu. Soubory věcných autorit mohou však sloužit i jako základ k vytvoření nových nástrojů pro efektivní a uživatelsky přívětivé vyhledávání, navigaci v množství informací, které se může uživatelům zdát obtížně zvládnutelné. Od tematické mapy fondu je možné pokročit např. k různým vizualizacím. Příkladem je např. projekt WorldCat Identities (www.worldcat.org/identities/), prezentace metadat v platformě HighWire, která využívá hyperbolický strom při hledání dokumentů podle tématu nebo katalog nové generace AquaBrowser, který uživateli zobrazuje vybrané téma a k němu se vážící vztahy. Soubor národních věcných autorit může potenciálně fungovat jako základ jakéhosi „překladače“ pro vyhledávání v jiných jazycích.

Více informací v příloze F.5.

B.6 Vymezení tříd entit

V návaznosti na závěry terminologické analýzy z předešlého roku projektu jsme se v tomto roce zaměřili na vymezení a označení tříd entit. Termín *třída entit* označuje skupinu entit se stejnými vlastnostmi. Pro jejich vymezení bylo tedy nezbytné sumarizovat vlastnosti, které mají jednotlivé entity. Výsledkem je vymezení sedmi tříd entit:

1. osoba/bytost,
2. rod/rodina,
3. korporace,
4. geografický objekt,
5. událost,
6. dílo/výtvar,

⁸ http://www.knihovny.cz/wp-content/uploads/2012/09/KNIHOVNY-CZ_RFI_2012_09_102.pdf

7. pojem.

Při označení jsme se rozhodli nevyužívat vazbu s přídavným jménem (například třída osobních entit), protože vytvořením slovního spojení se komplikuje jeho interpretace v budoucnu. Jako schůdná cesta se jeví využití pojmenování třídy formou těsného přístavku a uvedením alternativních termínů oddělených lomítkem (osoba/bytost). V tom případě budeme hovořit o:

- třídě entit osoba/bytost, třídě entit korporace,
- entitě ze třídy osoba/bytost, entitě ze třídy korporace.

Každá třída je rozdělena do podtříd. Členění do podtříd vychází ze společných znaků, resp. ze společné metodiky pro zpracování entit v rámci podtřídy.

Vymezení podtříd nemusí být konečné a může se v průběhu projektu změnit na základě požadavků odborné komunity. V příštím roce se bude pozornost věnovat také pojmenování podtříd.

Označení třídy a podtřídy entity tvoří základ pro zpracování entity - předpokládá se, že pro zpracování každé třídy entit bude vytvořena samostatná metodika a samostatné softwarové rozhraní nebo jeho samostatná část (např. pracovní list, formulář). Taktéž můžou být vyčleněny metodické pravidla pro každou podtřídu entity.

Pro bližší zařazení entity do nějaké skupiny entit se používá při zpracování samostatný datový prvek, proto bylo možné zachovat pouze dvouúrovňovou kategorizaci tříd a podtříd.

B.6.1 Třída entit osoba/bytost

Třidu entit osoba/bytost tvoří tyto podtřídy:

- fyzická osoba, fiktivní fyzická osoba,
- bytost,
- zvíře.

B.6.2 Třída entit rod/rodina

Třidu entit rod/rodina tvoří tyto podtřídy:

- rod, rodina, fiktivní rod, fiktivní rodina,
- větev rodu, rodiny.

B.6.3 Třída entit korporace

Třidu entit korporace tvoří tyto podtřídy:

- administrativně vymezená území s vlastní správou (například obce, panství, velkostatky, farnosti),
- orgány, organizační složky a příspěvkové organizace veřejné správy (vláda, Parlament České republiky, Poslanecká sněmovna, Senát, prezident, ministerstva, ústřední úřady, krajské úřady, okresní úřady, obecní úřady, soudy, státní fondy),

- sdružení korporací (sdružení organizací, mezinárodní, mezivládní, vládní organizace, pakty, mezinárodní smlouvy, mikroregiony/svazky obcí),
- vojenské a bezpečnostní jednotky (včetně teroristických a partyzánských skupin),
- organizace založené za účelem podnikání (podniky, firmy, fyzické osoby mající charakter korporace - živnostníci),
- politické organizace (strany, hnutí),
- náboženské organizace a instituce (církve a náboženské společnosti, farní úřady náboženská sdružení a spolky, historické právnické osoby - podle kanonického práva),
- kulturní, výchovné a výzkumné organizace a instituce (organizace a instituce pro výchovu, vědu a kulturu) ,
- nadace a nadační fondy,
- profesní a zájmové organizace (profesní komory, odborové organizace),
- spolky, společenské organizace,
- pojmenované části korporací (komise, oddělení, útvary, poradní orgány).

B.6.4 Třída entit geografický objekt

Třidu entit geografický objekt tvoří tyto podtřídy:

- administrativně či jinak lidmi vymezená území:
 - státní útvary (stát, země, závislé území),
 - historické oblasti, regiony a území (nadnárodní i v rámci jednoho státního útvaru),
 - útvary v rámci státu/země (kraj, provincie, prefektura, department, panství, vojenský újezd, území farnosti, kraj, okres, (Bezirk, district, county), obec, městská část, městský obvod, městská čtvrť, místní část obce, prostranství (náměstí, ulice, nábreží),
 - území vymezená pomístními jmény,
- administrativně vymezené části přírody (národní park, přírodní rezervace, chráněná krajinná oblast, přírodní památka, významný krajinný prvek, ptačí oblast, lesopark, park, sad, zahrada),
- vymezené obvody a správní celky - nespádající do kategorií uvedených výše (národopisná oblast, památková zóna, archeologická lokalita, rekreační území),
- geomorfologické útvary na dně moře (pevninský šelf, pevninský svah, hlubokooceánská pánev, hlubokooceánský příkop, oceánský hřbet, podmořská hora, podmořská sopka),
- geomorfologické útvary na zemském povrchu,
 - kontinent, světadíl, litosférická deska, zlomové pásmo, ledovec,
 - horizontální (souostroví, ostrov, poloostrov, atol, mys, útes),
 - vertikální (pahorkatina, vrchovina, (vele)hornatina, pohoří, kopec, hora, vrch, skála, sopka, rokle, průsmyk, sedlo, údolí, jeskyně, rovina, plošina, nížina),
- vodstvo a vodní plocha, vodní tok (oceán, moře, mořský proud, záliv, jezero, pleso, vodopád, gejzír, řeka, potok, pramen, peřeje, povodí, meandr, rybník, údolní nádrž, vodní nádrž jiné přírodní útvary a jevy,
- pojmenované útvary typu rašeliniště, prales, les, skupina stromů, strom, poušť, oáza),
- pojmenované trvalé klimatické jevy (např. Jihovýchodní obchodní vítr),
- Vesmír a jeho části (mlhovina, hvězda, planeta, měsíc).

B.6.5 Třída entit událost

Třidu entit událost tvoří tyto podtřídy:

- organizované akce a události,
 - krátkodobé (sjezd, konference, seminář, workshop, slavnost, oslava, festival, soutěž, veletrh, výstava, dobročinná akce),
 - „historické“ (válka, bitva, vojenská operace, obléhání povstání, revoluce, hnutí, poprava),
 - rozvojové plány (Marshallův plán, akce Z),
- lidové zvyky, významné dny a svátky,
- dočasné přírodní jevy (Kyrill, Katrina).

B.6.6 Třída entit dílo/výtvar

Třidu entit dílo/výtvar tvoří tyto podtřídy:

- autorská a umělecká díla (divadelní, hudební, literární, náboženské, naučné, rozhlasové, televizní, vědecké, výtvarné dílo),
- názvy všeobecně známých dokumentů, zákonů, předpisů, norem,
- vyznamenání, ceny, soutěžní trofeje,
- výrobky a jejich typová označení, obchodní značky, projekty korporací, odrůdy, plemena vytvořená člověkem,
- názvy společenských, dětských her,
- stavby, trasy, zásahy do přírodních útvarů s vlastním jménem nebo jinou identifikací,
 - budovy, komplexy budov, domy,
 - místa těžby surovin,
 - liniové stavby a trasy (dopravní..., plavební kanály, průplavy).

B.6.7 Třída entit pojem

Třída entit pojem bude rozpracována v dalším roce projektu.

B.7 Doplnění a úprava datové struktury

Formulace funkčních požadavků na softwarové řešení a vytvoření funkčního konceptu webového rozhraní si vyžádalo doplnění a úpravy datové struktury. Především byly doplněny prvky související s administrací záznamů entit, prvky odkazující na záznamy dokumentů a prvky pro zpracování informací o vyobrazení souvisejícím s entitou. Přehodnocení některých vztahů mezi prvky umožnilo ve větší míře využít princip dědičnosti a zjednodušit strukturu ve skupině prvků pro popis různých vlastností a událostí souvisejících s entitou.

Schéma datové struktury je na Obr. 5.

Použitelnost datové struktury v praxi byla ověřena při zpracování funkčního konceptu webového rozhraní. Jednotlivé prvky webového rozhraní jsou namapovány na příslušné prvky datové struktury. Popis struktury formulářů webového rozhraní je v příloze F.4.

B.8 Funkční požadavky na softwarové řešení

Základní funkční požadavky na softwarové řešení definovali řešitelé projektu jako:

- webové rozhraní přístupné po přihlášení,
- vyhledávání zpracovaných entit,
- vytváření nových entit a editace stávajících, pro každou třídu entit musí existovat osobitní pracovní list nebo formulář s definovanými údaji,
- dvojúrovňové zpracování (zpracovatel a supervizor), rozdílné úrovně přístupu pro zpracovatele a supervizor a též rozdílný přístup k záznamům z různých tříd entit,
- nástroje pro správu - rozesílání informací o změnách v záznamech e-mailem a pod.

Rozsah údajů pro každou třídu entit byl definován v tabulce. Více informací o funkčních požadavcích je v samostatném dokumentu.

B.9 Funkční koncept softwarového řešení a prototyp

Na základě stanovení základních funkčních požadavků na softwarové řešení a stanovení nezbytných údajů pro zpracování jednotlivých tříd entit byl vypracován funkční koncept webového rozhraní. Po schválení funkčního konceptu se začalo s vývojem funkčního prototypu.

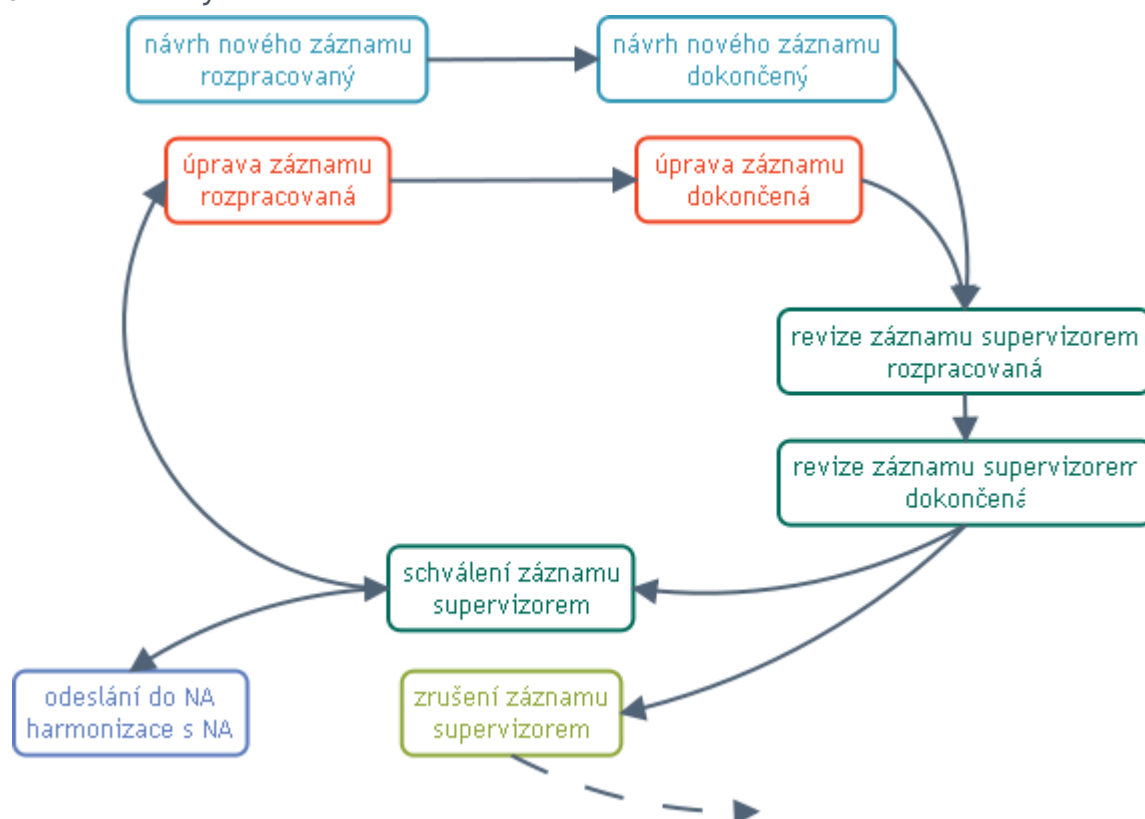
B.9.1 Funkční koncept softwarového řešení

Ve funkčním konceptu bylo potřebné vymezit rozsah a účel řešení, čímž byly stanoveny základní podmínky pro webové rozhraní:

- životní cyklus záznamu,
- řízení přístupu,
- funkce webového rozhraní,
- pracovní postup.

Životní cyklus záznamu vychází z požadavků odborné komunity především v podobě dvojúrovňového zpracování. Životní cyklus záznamu byl navržen tak, aby umožňoval vytvoření, nebo změnu záznamu zpracovatelem a jeho následné schválení nebo odmítnutí supervizorem.

Obr. 3: Životní cyklus záznamu



Životní cyklus záznamu úzce souvisí s určením pracovního postupu, řízením přístupu k záznamům a vymezením základních funkcí webového rozhraní.

Řízení přístupu znamená ve funkčním konceptu stanovení úrovně přístupových práv a skupiny údajů, které je potřebné uchovávat o každém uživateli a které určují jeho přístupové práva. Webové rozhraní je určeno pouze autorizovaným uživatelům - tj. uživatelům, kteří zadali platné přístupové jméno a heslo. Anonymní uživatelé mají přístup pouze ke stránce s přihlašováním.

Obr. 4: Prototyp webového rozhraní - stránka s přihlášením



Každý autorizovaný uživatel musí mít definované:

- uživatelské jméno
- heslo
- e-mail
- plné jméno
- pracoviště
- typ přístupu:
 - zpracovatel
 - supervizor - schvaluje záznam a podle potřeby ho odesílá do NA ČR a zároveň má všechny práva zpracovatele
- třída entity s určenou úrovní přístupu, pro každou třídu se určuje (v budoucnu se přístup bude určovat až na úroveň podtříd entit):
 - nový záznam
 - editace
 - schválení
 - zrušení
 - předdefinováno pro vyhledávání

Pracovní postup je ve funkčním konceptu stanoven v těchto základních krocích:

1. Přihlášení uživatele
2. Otevření okna „Editace“
3. Vyhledání záznamu
4. Otevření záznamu/ů pro editaci
5. Editace záznamu
6. Uložení záznamu

Pro webové rozhraní byly stanoveny základní funkce, funkce pro supervizora a podpůrné funkce. Základními funkcemi jsou: přihlášení, vyhledávání a editace. Základní funkce jsou přístupny zpracovateli i supervizorovi. Supervizor má kromě základních funkcí k dispozici také schválení a zrušení záznamu. Podpůrné funkce byly stanoveny jako: odeslání záznamu do báze národních autorit ČR, synchronizace záznamů s národními autoritami ČR, odeslání e-mailu (o změně, vytvoření, schválení nebo zrušení záznamu).

Každá funkce je ve funkčním konceptu podrobně rozpracována - základní funkce jsou reprezentovány základními obrazovkami webového rozhraní. Funkce pro supervizora a podpůrné funkce představují takové úkony, které jsou realizovány především na úrovni serveru a supervizor nebo zpracovatel je o jejich informován prostřednictvím informačních oken.

Kompletní funkční koncept včetně podrobného popisu editačních formulářů je v příloze F.4 této zprávy.

B.9.2 Funkční prototyp webového rozhraní

Funkční prototyp webového rozhraní je realizován jako aplikace typu klient/server a jako webové aplikace. Cílem bylo zajistit autonomní fungování aplikace bez omezení na hardwarové a softwarové vybavení uživatelů a bez nutnosti instalace softwaru, zabezpečení zálohování dat na straně uživatelů a pod. Data jsou uchovávaný v databázi na serveru a uživatel k nim přistupuje prostřednictvím klienta - tzn. webového rozhraní (aplikace). Pro komunikaci webového rozhraní se serverem je využíván zabezpečený protokol HTTPS a princip webových služeb. Aplikace splňuje bezpečnostní požadavky a požadavky na ochranu dat.

Webové rozhraní též splňuje základní požadavky pro tvorbu webových aplikací - výhodou je např. oddělení funkční a grafické podoby rozhraní. Grafická podoba je z části realizována prostřednictvím externí definice kaskádových stylů, co umožňuje efektivně měnit vzhled rozhraní.

Webové rozhraní je realizováno technologií ExtJS. Ukázky webového rozhraní jsou na Obr. 6 - Obr. 8.

Funkční prototyp webového rozhraní bude představen spolupracovníkům na semináři projektu (19. - 20. 11. 2012) a bude zpřístupněn pro testování širší skupinou spolupracovníků od 1. 1. 2013.

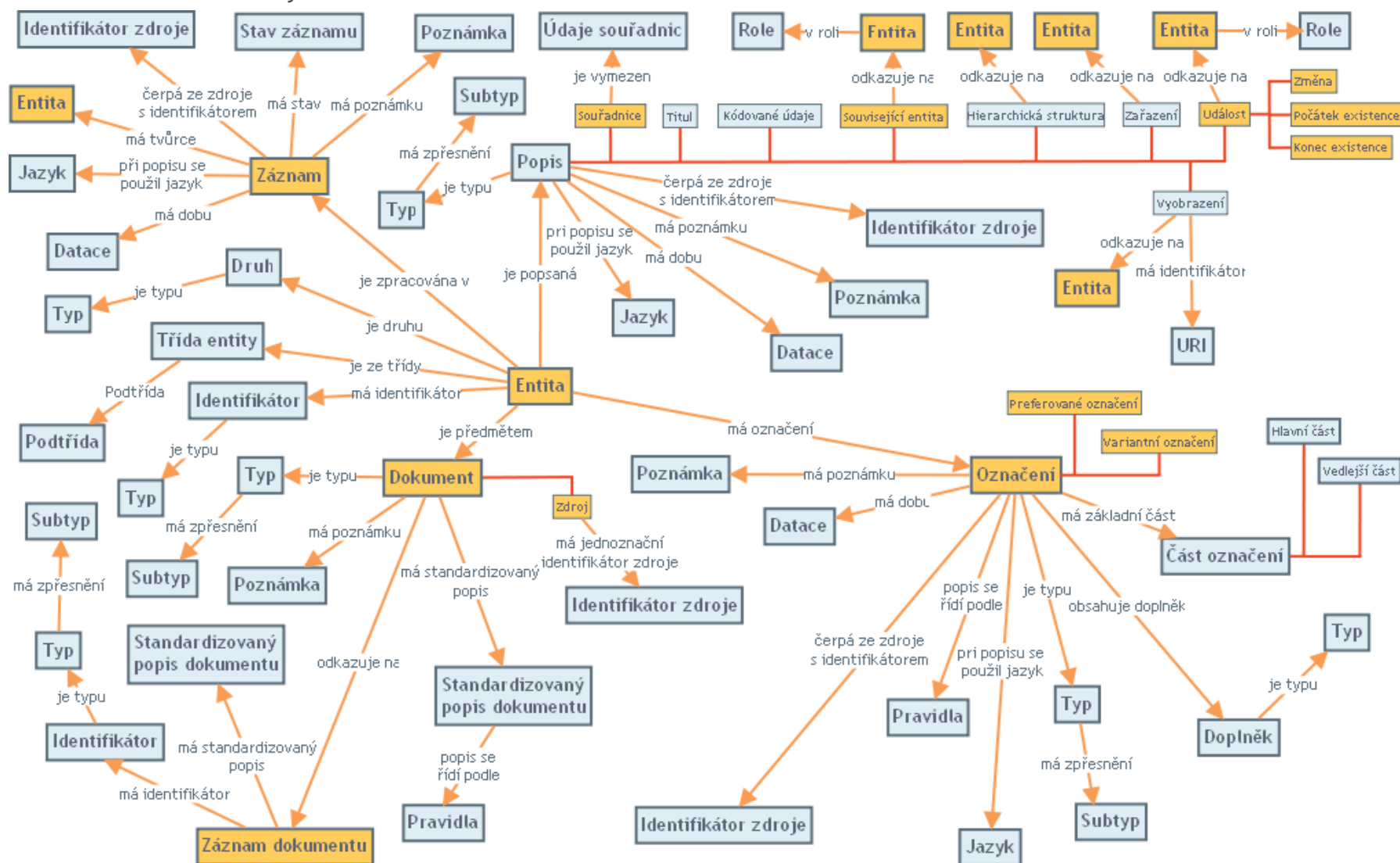
B.10 Propagace a informování o projektu

Hlavními činnostmi v oblasti zpřístupňování výsledků projektu je v roce 2012 organizace semináře (uskuteční se 19. - 20. 2012) a prezentace projektu na konferenci Archivy, knihovny a muzea v digitálním světě.

Seminář ke druhému roku projektu bude zaměřen na představení základní metodiky v podobě konceptu a prototypu webového rozhraní. Kromě toho budou účastníci obeznámeni se závěry terminologické analýzy (rozdělením entit do tříd a podtříd).

Kromě těchto aktivit bylo zadáno vypracování koncepce a grafického návrhu samostatné webové stránky - grafické návrhy budou k dispozici do 14. 12. 2012. Předpokládá se, že se zpřístupněním prototypu webového rozhraní pro testování bude vhodné začít realizaci samostatné stránky projektu, které bude sloužit jako přístupový bod k softwarovým řešením a také na zveřejňování informací o průběhu řešení projektu.

Obr. 5: Schéma datové struktury



Obr. 6: Prototyp webového rozhraní: editační formulář pro korporaci - identifikační údaje

INTERPI (interpi)

Vyhledat: Nový Menu

Seznam záznamů: **interpi_us_auth*new1**

Uložit Zrušit Schválit Odeslat do NA Zavřít **supervizor**

Třída entity INTERPI: Podtřída entity INTERPI: ID entity INTERPI: Alternativní ID entity INTERPI:

Stav záznamu: Autor záznamu: Datum vytvoření záznamu: Datum poslední změny záznamu:

- Preferovaná forma jména

Jméno: Doplněk: Rozpis iniciál:

Typ preferované formy jména: Datace jména od-do:

☐ korporace: jméno úřední
☐ korporace: jméno preferované entitou
☐ korporace: jméno, pod nímž je entita nejvíce známá
☐ korporace: jméno nejdéle užívané
☐ korporace: jediná známá podoba jména
☐ korporativní fyzická osoba: jméno úřední
☐ korporativní fyzická osoba: jméno preferované entitou
☐ korporativní fyzická osoba: jméno, pod nímž je entita nejvíce známá
☐ korporativní fyzická osoba: jméno nejdéle užívané
☐ korporativní fyzická osoba: jméno rodné
☐ korporativní fyzická osoba: jméno získané
☐ korporativní fyzická osoba: pseudonym
☐ korporativní fyzická osoba: jediná známá podoba jména

Použitá pravidla pro jméno:

Zánik/konec:

Datum od-do:

- Stručná charakteristika

Variantní/paralelní forma jména Jiné ID Poznámky Prameny Zařazení Obrázky Historie a jiné Vztahy a události

- Variantní forma jména

Jméno: Doplněk: Rozpis iniciál:

Obr. 7: Prototyp webového rozhraní: editační formulář pro korporaci - záložky

INTERPI (interpi)

Vyhledat: Nový Menu

Vznik/začátek: Zánik/konec:

- Sídlo korporace

ID entity: Geografický objekt: Datum od-do:

- Stručná charakteristika

Variantní/paralelní forma jména Jiné ID Poznámky Prameny **Zařazení** Obrázky Historie a jiné Vztahy a události

- Geografická působnost

ID entity: Geografický objekt: Datum od-do:

- Obor působnosti

ID entity: Pojem: Datum od-do:

- Určení jazykové oblasti

Určení jazykové oblasti:

- Kódované údaje

Zdroj kódu: Kód číselníku: Datace od-do:

Poznámka:

Obr. 8: Prototyp webového rozhraní: editační formulář pro korporaci - pomocný formulář

The screenshot displays the INTERPI (interpi) web application interface. At the top, there's a search bar with 'Korporace' selected and a 'Nový' button. Below this is a list of records with columns for 'Seznam záznamů', 'interpi_us_auth*new1', and 'interpi_us_auth*new2'. The main form area contains several sections: 'Uložit' (Save), 'Zrušit' (Cancel), 'Schválit' (Approve), 'Odeslat do NA' (Send to NA), 'Zavřít' (Close), and 'supervizor'. The 'Třída entity INTERPI' is set to 'Korporace'. The 'Podtřída entity INTERPI' is 'new'. The 'ID entity INTERPI' is 'new'. The 'Alternativní ID entity INTERPI' is empty. The 'Stav záznamu' is 'Vznik/začátek'. The 'Autor záznamu' is 'supervizor'. The 'Preferovaná forma jména' section includes fields for 'Jméno', 'Doplněk', 'Typ preferované formy jména', 'Jazyk jména', and 'Poznámka'. The 'Vznik/začátek' section includes a 'Sídlo korporace' section with 'ID entity' and 'Geografická působnost' fields. The 'Stručná charakteristika' section is empty. A modal window titled 'Vznik/začátek' is open, showing fields for 'Typ události', 'Subtyp události', 'Poznámka', 'Datování' (Date), 'Související osoba' (Related person), and 'Místo' (Location). The modal window has 'OK' and 'Zrušit' buttons. At the bottom, there are tabs for 'Variantní/paralelní forma jména', 'Jiné ID', 'Poznámky', 'Prameny', 'Zařazení', 'Obrázky', 'Historie a jiné', and 'Vztahy a události'. The 'Zařazení' tab is selected. The 'Geografická působnost' section is also visible at the bottom.

C Návrhová část

Obsahuje cíle projektu INTERPI pro rok 2013 - tedy pro druhou etapu projektu. Pro druhou etapu byly v projektové dokumentaci stanoveny tyto cíle:

1. vývoj softwaru pro základní skupinu entit,
2. specifikace vlastností doplňující skupiny entit z hlediska sémantické interoperability v paměťových institucích (kombinace autor/dílo, dílo, materiální objekty, pojmy, časové souvislosti).

Vzhledem k úpravě cílů projektu pro rok 2012, které jsou vysvětleny v části B je potřebné upravit také cíle pro rok 2013 tak, aby aktivity projektu na sebe vhodně navazovaly:

1. vývoj softwaru pro základní skupinu entit - v této oblasti se zaměříme na úpravu prototypu softwaru, jeho testování a doplnění podle potřeb tak, aby výsledek (tj. výsledná verze softwaru) vyhovoval požadavkům pro zpracování a prezentaci základní skupiny entit (tj. entit třídy osoba/bytost, rod, geografický objekt a korporace) - software bude uplatnitelný jako výsledek do RIV pro rok 2013;
2. analýza, ověřování, doplňování a úprava dat;
3. specifikace vlastností doplňující skupiny entit - půjde o činnosti v rozsahu:
 - analýza potřeb paměťových institucí v oblasti doplňující skupiny entit,
 - vymezení tříd a podtříd entit (aktualizace letošních závěrů podle potřeb paměťových institucí),
 - formulace funkčních požadavků na doplňující skupinu entit,
 - formulace funkčních požadavků na software pro doplňující skupinu entit a doplnění stávajícího funkčního konceptu o podrobnosti týkající se doplňující skupiny entit;
4. příprava metodiky pro základní skupinu entit - metodika bude uplatnitelná jako výsledek do RIV pro rok 2013;
5. propagace projektu - v oblasti propagace projektu a informování spolupracovníků projektu o dosažených výsledcích bude v roce 2013 realizován workshop a zpřístupněná samostatná stránka projektu.

D Použití účelové podpory v členění dle rozpočtu projektu

E Resumé

Druhý rok 1. etapy projektu navazuje v základním výzkumu na výsledky roku 2012 a je zaměřen především na syntetizaci poznatků o potřebách paměťových institucí v oblasti základní skupiny entit s přesahem na další entity. Základní výzkum spočíval v ověřování a doplňování hlavních zdrojů termínů, které budou tvořit obsah projektu INTERPI - jde o bázi národních autorit, rejstříky archivů a další terminologické a faktografické specializované zdroje.

V souvislosti se získáváním, propojením a prezentací informací z různých zdrojů se základní výzkum zaměřil na technologie sémantického webu a nástroje pro propojení různých terminologických systému. V této oblasti se pokračovalo ve zkoumání možností MDT.

V tomto roce se ukázalo jako důležité věnovat zvýšenou pozornost vývoji softwarového řešení i přesto, že je jeho vývoj hlavní náplní následující etapy. Na základě požadavků odborné komunity jsme přistoupili k urychlené formulaci funkčních požadavků a vytvoření funkčního konceptu softwarového řešení (webového rozhraní) v rozšířeném rozsahu (tj. pro více tříd entit: osoby/bytosti, korporace, geografický objekt, pojem, rod/rodina, událost, dílo) a také k vývoji funkčního prototypu.

V oblasti propagace projektu se pozornost soustředila na přípravu semináře, který se bude konat 19. - 20. 11. 2012. Kromě toho bude do konce roku 2012 vypracován koncept samostatné webové stránky projektu včetně grafického návrhu. Projekt byl (a po odevzdání průběžné zprávy bude) prezentován v roce 2012 na odborných akcích.

Za nejdůležitější výsledky projektu pro rok 2012, které vedly k nejvýznamnějšímu posunu znalostí a lze je považovat za nejcennější přínos řešitelů, jsou funkční koncept webového rozhraní (viz příloha F.4), funkční prototyp webového rozhraní, aktualizace datové struktury a vymezení tříd a podtříd entit. Tyto výsledky jsou jedinečné i v tom, že jsou projevem konkrétní kooperace subjektů různých typů (paměťových institucí, soukromého sektoru), při které jsou zohledněny a respektovány oprávněné specifické požadavky a rozdílné úhly pohledu všech zúčastněných subjektů.

Uplatněný výsledek - druh výsledku W podle RIV:

INTERPI. Seminář projektu INTERPI. 19. 11. 2012 - 20. 11. 2012. Brno. Pozn.: výsledek bude uplatněn po uskutečnění semináře.

INTERPI. Workshop spolupracovníků projektu INTERPI. 28. 11. 2011 - 29. 11. 2011. Brno. Pozn.: výsledek bude uplatněn zpětně pro rok 2011.

F Přílohy

- F.1 Použití účelové podpory v členění dle rozpočtu projektu - komentář
- F.2 Dodatky, rozhodnutí
- F.3 Cestovní zprávy
- F.4 Funkční koncept webového rozhraní pro zpracování entit
- F.5 Zpráva o činnosti Národní knihovny v projektu INTERPI za rok 2012
- F.6 Zpráva o činnosti Národního archivu v projektu INTERPI za rok 2012
- F.7 Přehled požadavků paměťových institucí na zpracování entit pro projekt INTERPI - zpráva za rok 2012
- F.8 Nástroje sémantického webu a věcné authority - aplikace v projektu INTERPI

F.1 Použití účelové podpory v členění dle rozpočtu projektu - komentář

F.2 Dodatky, rozhodnutí

F.3 Cestovní zprávy

F.4 Funkční koncept webového rozhraní pro zpracování entit

F.5 Zpráva o činnosti Národní knihovny v projektu INTERPI za rok 2012

F.6 Zpráva o činnosti Národního archivu v projektu INTERPI za rok 2012

F.7 Přehled požadavků paměťových institucí na zpracování entit pro projekt INTERPI - zpráva za rok 2012

F.8 Nástroje sémantického webu a věcné authority - aplikace v projektu INTERPI