

ZPRÁVA O ČINNOSTI NÁRODNÍ KNIHOVNY V PROJEKTU INTERPI ZA ROK 2015

INTERPI – Interoperabilita v paměťových institucích

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

Sestavila:
Marie Balíková



INTERPI

13. 11. 2015

Obsah

1	Analýza, ověřování, doplňování a úprava dat	2
1.1	Role souboru národních autorit.....	2
1.2	Soubory národních autorit v projektu INTERPI.....	3
1.3	Závěr	3
2	Implementace pravidel RDA z pohledu INTERPI	4
2.1	Úvod	4
2.2	Rozšíření autoritních záznamů o potřebné informace.....	4
2.3	Připojení údajů o korporativnosti	5
2.4	Fiktivní, legendární, mytologické postavy	5
3	Úprava autoritních termínů.....	7
3.1	Úprava souboru tematických autorit	7
3.2	Úprava souboru geografických autorit	7
3.3	Aktualizace klasifikačního systému MDT	8
4	Propagace a informování o projektu	10
4.1	IFLA.....	10
4.2	Lisabon.....	10
4.3	Tvorba monografií	10

1 Analýza, ověřování, doplňování a úprava dat

V roce 2015 NKČR pokračovala v pracích spojených s analýzou, exportem a úpravou autoritních údajů jako základního předpokladu pro úspěšný vývoj znalostní databáze INTERPI. Byla vypracována analytická studie o možnosti aplikace souboru národních autorit při tvorbě znalostní báze INTERPI.

1.1 Role souboru národních autorit

Soubor národních jmenných autorit obsahuje soubor personálních autorit (639 472 termínů), korporativních autorit (140 002), unifikovaných názvů (21 301), tematických autorit (36 954), geografických autorit (27 170) a autorit označujících formu/žánr (1 783).

Soubory národních autorit byly původně vytvořeny pro podporu sdílené katalogizace a pro jednotné zpřístupnění fondů knihoven. Jmenné autority se vytvářely podle pravidel AACR2, věcné autority se vytvářely podle zásad LCSH. Později se při jejich tvorbě uplatňovala postkoordinace, byly autorizovány izolované lexikální řetězce předmětových hesel, nikoli celé řetězce předmětových hesel. Hlavním cílem autoritních záznamů byla autorizace preferovaných forem jmen a termínů, zaznamenání jejich variantních forem a poznámkového aparátu. Pro zápis jejich struktury se používá formát MARC.

Silné stránky tradičních autoritních souborů

V tradičních autoritních záznamech se definují selekční prvky, přičemž poznámkový aparát je uplatněn pouze v minimální potřebné míře. Používají se při sdílené katalogizaci v knihovnách i v jiných kooperujících institucích. Většina knihoven v Česku spolupracuje v kooperačním systému Kooperativní tvorba souborů národních autorit. Muzea vytvářejí ve spolupráci s Národní knihovnou Muzejní autority, na jejichž základě pak vzniká Registr sbírek výtvarného umění. Všechny tyto projekty jsou založeny na národních autoritách, přičemž Muzejní autority a Registr se omezují na tvorbu personálních autorit.

Slabé stránky tradičních autoritních souborů

K slabým stránkám tradičních autoritních souborů patří především to, že jsou primárně určeny pro intelektuální vnímání, nikoli pro strojové zpracování. Dále jsou jejich údaje jen částečně strukturovány, jsou tudíž charakterizovány nízkou granularitou. AACR2 a formát MARC totiž nepodporují dostatečnou granularitu údajů, která je potřebná pro Sémantický Web a princip propojených dat. „MARCový záznam nebyl vytvořen pro zápis datových segmentů/elementů, byl vytvořen pro ukládání a zobrazení textu katalogizačního lístku.” (Coyle, 2012)

Vedle toho se entity téhož typu zařazují v různých souborech autorit v přímé závislosti na poli/tagu formátu MARC. Některé entity jsou zařazeny do dvou různých souborů, např. jména akcí (konferencí, seminářů, workshopů) jsou řazeny v souboru korporativních autorit, názvy

bítev, válek do souboru tematických autorit. Tyto entity by měly být řazeny na základě atributů a vztahů, nikoli podle pole formátu MARC.

Autoritní záznamy také obsahují málo informací o entitě, kterou zpřístupňují: soustřeďují se pouze na preferovanou formu jména a na její varianty zapsané jako odkazy viz. V autoritních záznamech však není možné uvést informaci o jejich vlastnostech strukturovaně, nýbrž pouze jako volný text v poli pro poznámky.

Navíc se podle zásad pravidel AACR2 i RDA nepoužívají označení typu korporace, např. *n. p.*, *s r. o.* apod. Nicméně jsou tato označení nesmírně důležitá pro archivní popis.

1.2 Soubory národních autorit v projektu INTERPI

Poté, co byly národní autority podrobeny důkladné analýze, bylo rozhodnuto, že mohou být v systému INTERPI využity při dodržení následujících principů:

- Standardizované přístupové body budou využity jako preferovaná a variantní označení entit.
- Všechny informace, které uvádějí entitu do kontextu, budou v záznamu o entitě využity a budou zapsány ve strukturované podobě.
- Entity se budou řadit do tříd na základě vlastností a vztahů, do nichž vstupují, nikoliv na základě pole formátu MARC.
- Bude vytvořen proprietární formát datové struktury, v němž budou zohledněny potřeby odborných komunit paměťových institucí.
- Konceptuální model INTERPI bude založen na entitách, nikoli na autoritách.

1.3 Závěr

Kvalitně koncipované soubory autoritních termínů jsou přínosné i v současném informačním prostředí. Lze je po nezbytných úpravách využít i v rámci sémantického webu. V tomto případě ale musejí projít důkladnou transformací v intencích konceptuálních modelů FRAD a FRASD zvláště v oblasti struktury dat a v oblasti explicitního vyjádření vztahů a vazeb, do kterých zpřístupňované autoritní entity vstupují.

Ukazuje se, že rozvoj ICT technologií výrazným způsobem ovlivňuje všechny postupy v knihovnách, v meziknihovní spolupráci, a nově i v kooperaci mezi paměťovými institucemi všech zaměření, jež směřuje zejména k integraci nástrojů zpřístupnění informací a kulturního dědictví, za něž odpovídají. Právě integrovaný přístup ke kulturnímu dědictví v širokém smyslu znamená pro obecné uživatele i profesionály zásadní proměnu a krok dopředu při vyhledávání informací.

2 Implementace pravidel RDA z pohledu INTERPI

2.1 Úvod

Z pohledu INTERPI představují pravidla Resource Description and Access (RDA) nový standard pro bibliografický popis informačních zdrojů, převážně textových. Za pramen popisu se považuje celý informační zdroj, jsou však nadále stanoveny preferované prameny popisu, např. titulní stránka. Bibliografické údaje se přepisují do záznamu přesně tak, jak se vyskytují v informačním zdroji. Pravidla jsou tedy určena pro knihovnickou obec, protože archivářská a muzejní komunita velmi často pracují se zdroji, které tyto základní bibliografické údaje postrádají a musejí být jimi - často na základě rozsáhlých rešerší - vytvořeny, dodány.

Pravidla RDA jsou pravidla obsahová (content standard), vycházejí z konceptuálních modelů a jejich primárním cílem je jednoznačná identifikace entit a vyjádření vlastností a vztahů, do kterých vstupují. Z tohoto důvodu představují v projektu základní standard pro tvorbu obsahu záznamů o entitě. Avšak vzhledem k rozdílnému pojetí některých entit v jednotlivých komunitách paměťových institucí, např. třídy entit dílo, osoba (anonymní tvůrce), korporace (vyjádření korporativnosti) byla pravidla RDA v těchto případech rozšířena o některá ustanovení pocházející z jiných pravidel, např. Cataloging Cultural Objects (CCO).

2.2 Rozšíření autoritních záznamů o potřebné informace

V důsledku implementace pravidel RDA v knihovnickém prostředí dochází i ke změnám v záznamech autorit: důraz na důsledné vyjádření charakteristik jednotlivých entit a vzájemných vztahů si vyžádal rozšíření repertoáru polí ve výměnném formátu MARC 21. Nově lze v záznamech autorit zaznamenat informace jako související místo, oblast působení, organizace spojená s osobou (dříve afiliace), profese, pohlaví, jazyk a lze také uvádět rozpis iniciál daného pojmenování. Došlo tak k přímé integraci většiny požadavků komunit paměťových institucí, které se ve své praxi nespokojily jenom se standardizací selekčních prvků, naopak trvaly na důsledném uvádění všech potřebných informací v autoritních záznamech ve strukturované podobě. Repertoár nově připojených polí je zveřejněn na adrese http://www.nkp.cz/o-knihovne/odborne-cinnosti/zpracovani-fondu/katalogizacni-politika/vecne-udaje_-rda

Obr. 1: Ukázka autoritního záznamu obsahujícího nově připojená pole pro zápis strukturovaných informací

LDR	-----nz--a22-----n--4500
FMT	JA
001	jo2015891526
003	CZ-PrNK
005	20151103115752.0
008	151103 n aznnaabn-----n-aja-----
040	a ABA001 b cze e rda
1001	a Queen, Sarah A. q (Sarah Ann) 7 jo2015891526
370	c Spojené státy americké
372	a sinologie a čínská filozofie a konfucianismus a taoismus
373	a Connecticut College s 1992
374	a sinoložky a historičky
375	a žena
377	a eng
670	a Liu An: The essential Huainanzi. (translated, edited, and selected by John S. Major, Sarah A. Queen ... et al.), [2012] a www(Connecticut College), cit. 3. 11. 2015 b biografické údaje a LC (Names), cit. 3. 11. 2015 b autoritní forma
6780	a Americká historička a sinoložka zaměřená na čínské kulturní dějiny, filosofii a náboženství.

2.3 Připojení údajů o korporativnosti

Typ korporace, tzv. korporativnost představuje významný prvek při stanovení standardizované autoritní formy pojmenování v případě korporací; změní-li se korporativnost (n.p. -> a.s. apod.), je nutno vytvořit nový samostatný záznam entity. Typ korporace se uvádí jako součást názvu korporace. Tím se záznam entity korporace výrazně odlišuje od autoritního záznamu korporativního záhlaví.

2.4 Fiktivní, legendární, mytologické postavy

Fiktivní postava je osoba, identita nebo entita, jejíž existence je smyšlená a založená na fikci. Fiktivní postava může nabývat nejrůznějších podob, od živých lidí, rostlin či zvířat až po různé nadpřirozené či pohádkové bytosti¹.

Za fiktivní postavy se považují:

- Vlastní fiktivní postavy, např. *Jára da Cimrman*, *Jane Marple*
- Legendární postavy, např. *Romulus*, *Přemysl Oráč*
- Mytologické postavy, např. *Sisyfos*, *Medúza*
- Bohy, božstva, např. *Merkur*, *Zeus*, *Kršna*, *Lakšmí*
- Reálné postavy, které nejsou lidského původu, např. *Buddy (pes)*, *Bo (pes)*, *Socks (kočka)*,
- Fiktivní postavy, které nejsou lidského původu, např. *Šemík*,

- Bájná stvoření, např. *Hroznová koza*

V současné době jsou fiktivní entity součástí souboru věcných tematických autorit. Ve znalostním modelu INTERPI tvoří podtřídu třídy entit osoba, viz: „ Za entitu z třídy osoba/bytost se považuje skutečná, fiktivní nebo předpokládaná osoba.“ Při konverzi budou autoritní záznamy fiktivních entit převedeny do třídy entit osoba. Před konverzí bylo nutné tyto záznamy obohatit o informace vyjadřující jejich základní vlastnosti a vztahy. Bylo upraveno 490 záznamů.

Obr. 2: Ukázka

Ident. číslo	ph473687
Záhlaví	● Avalókitéšvara (buddhistické božstvo)
Odkaz. forma	Kuan Jin (buddhistické božstvo) Kuan-jin (buddhistické božstvo) Kuan Yin (buddhistické božstvo) Guanyin (buddhistické božstvo) Kuan-š'-jin (buddhistické božstvo) Kannon (buddhistické božstvo) Kwanim (buddhistické božstvo) Kwanseum (buddhistické božstvo) Quan Âm (buddhistické božstvo) Čánrázig (buddhistické božstvo) Guan Im (buddhistické božstvo)
Nadřazený termín	asijské bohyně asijské bohové buddhové bódhisattvové buddhistická božstva
Asoc. termín	asijská náboženství buddhismus taoismus čínská mytologie
MDT	● 24-14-36
Angl. ekvivalent	Avalokitéšvara (Buddhist deity)
Skupina Konspektu	● 22/24 - Náboženství Dálného východu. Indická náboženství. Hinduismus. Buddhismus [5]

¹ http://cs.wikipedia.org/wiki/Fiktivn%C3%AD_postava

3 Úprava autoritních termínů

3.1 Úprava souboru tematických autorit

V roce 2015 jsme pokračovali v zásadních úpravách v souboru tematických autorit pro potřeby projektu INTERPI: do souboru tematických autorit byly připojeny jako variantní termín (odkaz typu viz) víceslovné obecné termíny v invertovaném slovosledu. Cílem této zásadní úpravy je napomoci orientaci uživatele při vstupu do systému INTERPI pomocí přístupových rejstříků a také usnadnit postup při harmonizaci/mapování řízených slovníků. Bylo upraveno 800 tematických termínů.

3.2 Úprava souboru geografických autorit

V souboru geografických autorit jsme navázali na úpravy realizované v roce 2014, kdy bylo dořešeno propojení této databáze s Databází sídelních lokalit Čech, Moravy a Slezska cz_retro pomocí pole 052, kde se uvádí kód cz_retro. V záznamech geografických autorit se postupně doplňuje pole 052 v případech, ve kterých nedošlo k automatickému propojení na bázi údajů v poli 951 z důvodů rozdílnosti údajů. Bylo upraveno 600 záznamů.

V únoru 2015 předána data CZ_RETRO firmě Cosmotron k testovacímu provozu v rámci výsledného řešení INTERPI. Součástí toho byla i ideová a grafická představa samostatné webové aplikace CZ_RETRO.

V následujícím období pokračovaly kontroly a revize, které vedly k upřesnění polohy řady drobných lokalit. Současně byly do systému doplněny stovky nových lokalit, zejména těch administrativně neevidovaných a přitom z hlediska kulturního dědictví významných. Další důležitou součástí právě bylo doplnění všech zaniklých vsí na Moravě a ve Slezsku, aby mohlo dojít k propojení s údaji zpracovávanými k nim v Národním archivu.

Ke dni 9. 9. 2015 obsahuje CZ_RETRO:

	subtyp	význam	lokality
LOKALIZOVANÉ			39 600
<i>body lokalit</i>			39 600
<i>polygony historických osad (pouze Česká republika)</i>			22 112
<i>polygony čtvrtí (pouze Česká republika)</i>			3 594
<i>polygony dílů (pouze Česká republika)</i>			27 346
	1	historická osada	21 459
	10	zaniklá historická osada	1 229
	11	dávno zaniklá historická osada	530
	2	místní část	4 524
	22	zaniklá místní část	2 259
	3	hlavní čtvrť	633
	33	zaniklá hlavní čtvrť	13
	333	odlišně vymezená hlavní čtvrť	12

	4	čtvrť (ostatní)	2 887
	44	zaniklá čtvrť (ostatní)	95
	444	odlišně vymezená čtvrť (ostatní)	75
	5	hradiště	667
	55	zaniklé hradiště (odtěžením) či hradiště domnělé	21
	6	nesídelní a ostatní lokalita	176
	7	díl polygonu historické osady či čtvrtě	3 632
	77	díl polygonu zaniklé historické osady či čtvrtě	263
	777	území předané jinému státu	81
	8	celek historických osad	82
	0	umělá lokalita	962
NELOKALIZOVANÉ			22 683
	119	dávno zaniklá historická osada dosud nelokalizovaná	2 118
	29	místní část dosud nelokalizovaná	20 038
	229	zaniklá místní část dosud nelokalizovaná	292
	701	feudální díl vsi (Berní rula/Lánové rejstříky)	111
	702	ves vyjma svobodné rychty či statku (Berní rula/Lánové rejstříky)	39
	703	svobodná rychta či statek v rámci vsi (Berní rula/Lánové rejstříky)	51
	977	nelokalizovatelný díl či chyba evidence	34
CELKEM			62 283
	<i>z toho</i>	<i>Česká republika</i>	<i>60 836</i>
<i>Polsko</i>			<i>1 406</i>
<i>Slovensko</i>			<i>22</i>
<i>Rakousko</i>			<i>12</i>
<i>Německo</i>			<i>7</i>

V současné době jsou databázová i gisová data CZ_RETRO připravena k plné integraci do systému INTERPI s tím, že je nezbytné jimi nahradit pracovní data poskytnutá na počátku roku 2015.

3.3 Aktualizace klasifikačního systému MDT

V roce 2015 jsme pokračovali také s aktualizací systému MDT a v teoretickém průzkumu o aplikaci klasifikačního systému MDT jako potenciálního propojovacího jazyka při řešení terminologické různorodosti termínů aplikovaných v jednotlivých paměťových institucích a jako nástroje pro tvorbu ontologií. Byl dokončen překlad systému MDT, který nyní obsahuje 70626 znaků.

Databáze České MDT online je pod registrací dostupná na adrese <http://cz.udc-hub.com/cs/login.php>

Obr. 3: Ukázka aplikace České MDT online

The screenshot shows the UDC online application interface. At the top, there is a navigation bar with the UDC logo and menu items: Aplikace MDT, Novinky, Kontakty, Zdroje, and Odhlásit. On the right, it says 'MEZINÁRODNÍ DESETINNÉ TRÍDĚNÍ' and features the NK logo. Below the navigation bar, there are tabs for 'Vyhledat', 'Tvorba notací MDT', 'Index', and 'Validace/Parsování'. The main content area is divided into a left sidebar and a right main panel. The sidebar, titled 'Přehled', shows a tree structure of MDT notations under 'POMOCNÉ TABULKY' and 'HLAVNÍ TABULKY'. The right panel has a search input 'Notace MDT' with the value '124' and a 'Vyhledat' button. Below the search input are checkboxes for 'Přesná shoda', 'Zahrnuje příklady', and 'Zahrnuje zrušené notace'. A blue banner indicates 'Třída 124 Má 6 pododdílů'. Below this, a list of results is shown, with the first item '122/129 Speciální metafyzika' highlighted in yellow. Other items include '124 Teleologie', '124.1 Řád. Chaos', '124.2 Smysl. Smysluplnost', and '124.3 Účel. Účelnost'.

V současné době je registrováno 525 uživatelů.

4 Propagace a informování o projektu

Propagace a informování o projektu na mezinárodním poli byl jedním z primárních a stěžejních úkolů NKČR.

4.1 Prezentace projektu na kongresu IFLA v Kapském Městě.

Příspěvek je dostupný na adrese <http://library.ifla.org/1135/1/148-balikova-en.pdf>.

Projekt INTERPI je mezinárodní komunitou vysoce hodnocen; vyzdvihuje se náš kritický přístup k pravidlům a postupům aplikovaným v jednotlivých odborných komunitách, včetně samotného knihovnického standardu RDA a kooperace mezi paměťovými institucemi. Oceňována je realizace propojení paměťových institucí a interoperabilita na sémantické úrovni.

4.2 Prezentace projektu na semináři o MDT v Lisabonu

téma konference: Classification & Authority Control: Expanding Resource Discovery

Příspěvek Subject authority control supported by classification: the case of National Library of the Czech Republic je dostupný na adrese:

http://seminar.udcc.org/2015/images/Balikova_InternationalUDCSeminar2015.pdf.

Také zde byla hodnocena schopnost definovat a jasně pojmenovat problémy spojené se sémantickou interoperabilitou v oblasti paměťových institucí.

4.3 Tvorba monografií

V projektové dokumentaci projektu jsme se zavázali k publikování dvou kolektivních monografií. Vzhledem k zájmu vyvolanému u zahraničních kolegů jsme se rozhodli publikovat jednu z těchto monografií v anglickém jazyce.

Jsou to:

- *Interoperabilita v paměťových institucích*
- *Facilitating access to cultural heritage content in Czechia*