

ZPRÁVA O ČINNOSTI NÁRODNÍ KNIHOVNY V PROJEKTU INTERPI ZA ROK 2013

INTERPI – Interoperabilita v paměťových institucích

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

Sestavila:
Marie Balíková



INTERPI

7. 11. 2012 | verze 1.0

Obsah

1	Analýza, ověřování, doplňování a úprava dat	2
1.1	Úprava autoritních údajů	2
1.2	Export autoritních údajů.....	3
1.3	Aktualizace klasifikačního systému MDT	3
2	Vymezení tříd a podtříd entit	6
2.1	Třída entit rod/rodina	6
2.2	Třída entit dílo/výtvor	6
2.3	Třída entit obecný pojem	7
3	Formulace funkčních požadavků na rozhraní pro doplňující skupinu entit.....	8
4	Zdůvodnění odkladu certifikace Metodiky	9

1 Analýza, ověřování, doplňování a úprava dat

V roce 2013 se NKČR zaměřila na analýzu, export a úpravu autoritních údajů jako základního předpokladu pro úspěšný vývoj znalostní databáze INTERPI. Údaje autoritní databáze byly podrobeny důkladné analýze z hlediska jejich budoucí integrace do tříd a podtříd entit znalostní báze INTERPI.

Byla definována kritéria pro export jednotlivých typů autorit v souladu s jejich budoucím zařazením. Údaje ze souboru personálních autorit jsou distribuovány do 3 tříd entit projektu INTERPI: personální údaje budou zařazeny do třídy entit osoba, kombinace údajů autor/dílo bude zařazena do třídy entit dílo/výtvor, do podtřídy autorská a umělecká díla a šlechtické rody a větve budou zařazeny do specifické třídy entit rod/rodina.

Autoritní údaje ze souboru korporativních autorit budou rozděleny do 3 tříd entit znalostní báze INTERPI, a to názvy vlastních korporací (např. firmy, školy, univerzity, úřady...) do třídy entit korporace, projekty, programy do třídy entit dílo/výtvor, stavby/budovy, např. hrady, zámky, kostely, kaple do třídy entit dílo/výtvor, do podtřídy stavby, trasy ... , akce typu konference, semináře, výstavy, workshopy ... do třídy entit událost.

Autoritní údaje ze souboru geografických autorit jsou/budou zařazeny do následujících tříd entit: vlastní geografické názvy se zařazují do třídy entit geografický objekt¹, stavby jako mosty, rozhledny... se zařazují do třídy entit dílo/výtvor do podtřídy stavby, trasy ..., liniové stavby, trasy do třídy entit dílo/výtvor do podtřídy stavby, trasy, atd.

Autoritní údaje ze souboru tematických autorit se řadí: termíny označující obecné pojmy spadají do třídy obecný pojem, názvy označující ozbrojené střety se řadí do třídy entit událost, termíny označující výrobky, produkty se řadí do třídy dílo/výtvor, termíny označující společenské a dětské hry do třídy entit dílo/výtvor. Označení fiktivních osob budou v databázi INTERPI zařazena do třídy entit osoba.

1.1 Úprava autoritních údajů

Záznamy autorit, u nichž došlo k zásadní změně v jejich zařazení, např. záznamy ozbrojených střetů, byly podrobeny důkladnější analýze z pohledu formy jejich zápisu. Na základě této analýzy bylo přistoupeno k jejich úpravě již v rámci stávajícího souboru autorit: forma jejich zápisu byla přizpůsobena formě zápisu obvyklému pro třídu entit, do které budou integrovány, byly opatřeny stručnou definující poznámkou a dalšími údaji tak, aby jejich zařazení do „nové“ třídy entit bylo pokud možno bezproblémové.

¹ Odkaz na výčet entit zařazených do třídy entit geografický objekt

Takto bylo upraveno 10 500 autoritních záznamů různých typů.

1.2 Export autoritních údajů

Podle definovaných kritérií byly vytvořeny exporthy stávajících autorit.

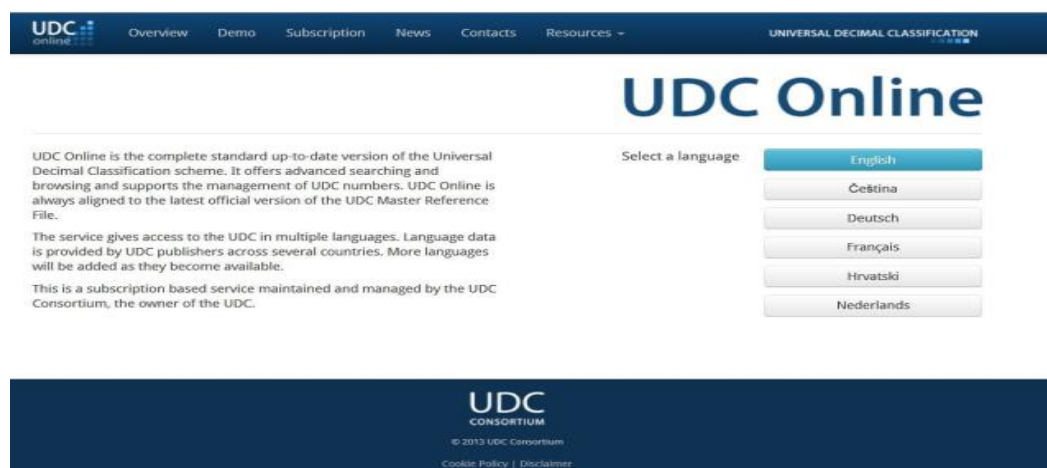
Bylo exportováno:

- 562016 personálních autorit označujících osoby,
- 13299 autorit označujících kombinace jména autora a název jeho díla,
- 128103 korporativních záhlaví,
- 3296 kombinací korporativních záhlaví a názvů díla,
- 25210 geografických autorit.

1.3 Aktualizace klasifikačního systému MDT

V tomto roce (2013) řešení projektu jsme pokračovali 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.

V roce 2013 pokračovaly práce na vývoji vícejazyčné překladové databáze MDT - UDC online.



UDC Online

UDC Online is the complete standard up-to-date version of the Universal Decimal Classification scheme. It offers advanced searching and browsing and supports the management of UDC numbers. UDC Online is always aligned to the latest official version of the UDC Master Reference File.

The service gives access to the UDC in multiple languages. Language data is provided by UDC publishers across several countries. More languages will be added as they become available.

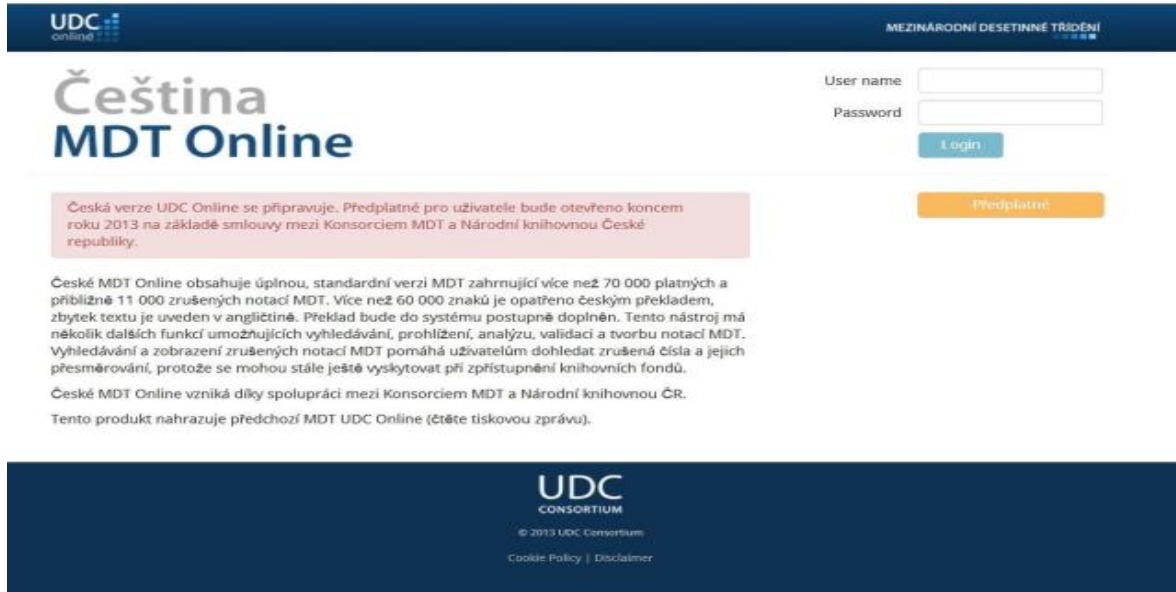
This is a subscription based service maintained and managed by the UDC Consortium, the owner of the UDC.

Select a language

- English
- Čeština
- Deutsch
- Français
- Hrvatski
- Nederlands

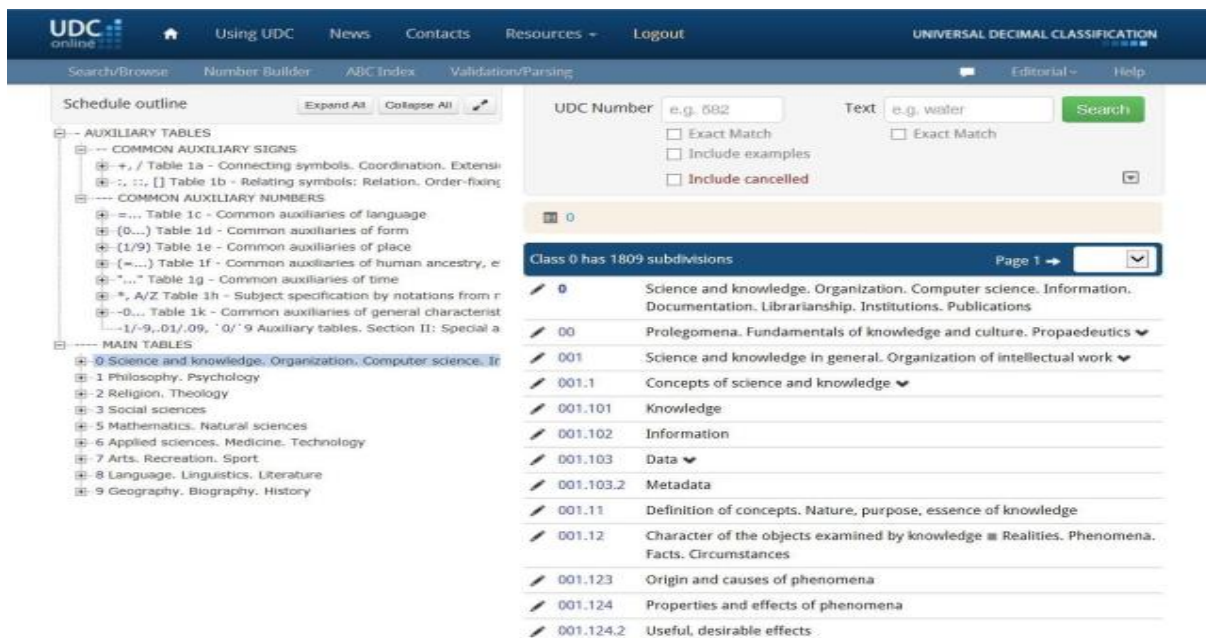
UDC CONSORTIUM
© 2013 UDC Consortium
Cookie Policy | Disclaimer

Obr. UDC_1



Obr. UDC_2

V souvislosti s tím pokračují práce i na překladu a aktualizaci systému MDT, který nyní obsahuje 70626 znaků. Česká verze obsahuje nyní 64 899 přeložených znaků, zbývajících 5727 je přeloženo, po posouzení budou připojeny do vícejazyčné překladové databáze.



Obr. UDC_3

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,

kteří vytváří prostředí pro jejich realizaci, např. vývoj, 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.

2 Vymezení tříd a podtříd entit

Základní schéma tříd a podtříd entit základní i doplňující skupiny entit navržené v loňském roce projektu bylo podrobena důkladné analýze z pohledu jednotlivých odborných komunit.

2.1 Třída entit rod/rodina

Znovu byla zkoumána třída entit rod/rodina a navržen postup pro jejich zpracování. Byly navrženy 3 postupy:

1. Stávající řešení v souboru jmenných autorit, kdy se použije jako základní jméno rodu a všechny větve rodu se odkáží na tento základ
2. Preferovaná podoba rodu rozšířená o rozlišující GEO doplněk označující rezidenci
3. Preferovaná podoba rodu rozšířená o doplněk typu titulární rozlišení - knížecí linie a hraběcí linie, hraběcí linie a linie rytířská (svobodných pánů)

Tato řešení vyhovují pro soubor národních jmenných autorit, pro projekt INTERPI jsou - vzhledem k požadavkům archivářské komunity - přijatelná jenom s velkými výhradami.

Z tohoto důvodu byla tato skupina entit vyčleněna do speciální třídy entit rod/rodina a protože se v podstatě jedná o korporaci², bylo pro zápis preferované formy větve rodu použito schéma pro zápis korporativního záhlaví s podřízenou složkou

Schwarzenbergové. Orlická větev

2.2 Třída entit dílo/výtvar

Třída entit dílo/výtvar byla vyčleněna na základě společných vlastností tohoto typu entit, především na základě nejdůležitější charakteristiky, a to že se jedná o umělý výtvar, intencionálně lidskou bytostí vytvořené dílo. Proto do této kategorie spadají i stavby jako intencionálně člověkem vytvořený artefakt.

² Korporace: Organizace nebo skupina osob, která se označuje konkrétním jménem a která vystupuje nebo může vystupovat jako entita. Rovněž zahrnuje individuální vystupování osob v rámci jejich postavení v korporaci.

Abychom zohlednili všechny charakteristiky všech lidských výtvorů zařazených do této kategorie, bylo nutné vybrat typické příklady. Ukázka těchto příkladů je uvedena v příloze F 7.

2.3 Třída entit obecný pojem

Třída obecný pojem představuje „tvrdý oříšek“ ve všech znalostních bázích. Bývá velmi obtížné stanovit přesné hranice mezi individuálním a obecným pojmem (např. formalizované pojmenované entity typu bitvy, války. Termíny označující pojmy jsou vázány na přirozený jazyk - problematika nejednoznačnosti termínů - homonymie, synonymie a především jsou vázány na národní jazyky, jejichž systémy se někdy výrazně odlišují.

3 Formulace funkčních požadavků na rozhraní pro doplňující skupinu entit

V tomto roce řešení projektu byly zpracovány funkční požadavky na rozhraní pro doplňující skupinu entit - tj. třídy entit: událost, dílo/výtvar a obecný pojem. Funkční požadavky vycházely z požadavků na rozhraní pro základní skupinu entit a z návrhu Metodiky tvorby znalostního modelu.

V rozhraní pro doplňující skupinu entit je kladen důraz na respektování rozdílných přístupů ve tvorbě označení pro tyto entity, tj. rozhraní musí umožňovat zápis označení podle různých pravidel. Pro třídu entit obecný pojem musí být zachována možnost zapsat deskriptor (označení) existujícího tezauru. V této třídě je také důležité poskytnout nástroje podporující proces začleňování vysoce specializovaných odborných termínů do databáze INTERPI.

4 Zdůvodnění odkladu certifikace Metodiky

Komplexní metodika tvorby znalostního modelu INTERPI se skládá ze dvou částí:

- základní metodiky, která obsahuje popis struktury znalostní báze INTERPI
- jednotlivých metodik pro zpracování jednotlivých tříd entit

Základní metodika je dokončena: v rámci projektu INTERPI bylo dosaženo dohody a definováno 7 tříd entit, které se dále dělí na podtřídy (viz dále)

Při tvorbě jednotlivých metodik pro třídy a podtřídy entit jsme narazili na problematiku pomalého vývoje mezinárodních a v návaznosti i národních standardů a pravidel a jejich národních interpretací.

Znalostní báze INTERPI je budována na objektovém principu a na bázi konceptuálních modelů. Jedná se o poměrně nové postupy a metody, jejichž vývoj není ustálený a navíc probíhá (či donedávna probíhal) odděleně v rámci jednotlivých odborných komunit; náznaky společných postupů knihovníků, muzejníků a archivářů jsou patrné až v posledních letech a jsou předmětem dosud neukončených projektů, např. vývoj technického standardu EAC.

Typickým příkladem postupného vývoje jsou konceptuální modely tvořené zpočátku pro zpracování bibliografického univerza, určené tedy primárně pro knihovnickou obec. Konceptuální model FRBR (Functional Requirements for Bibliographic Records - Funkční požadavky na bibliografické záznamy) publikovaný v roce 1997 definuje 10 základních entit (...), důležitá entita rod/rodina zde však zastoupena není. K její integraci dochází až v rámci konceptuálního modelu FRAD (Functional Requirements for Authority Data - Funkční požadavky na autoritní údaje) publikovaného v roce 2009 (poslední revize v roce 2013), který se zabývá tvorbou autoritních forem personálních jmen, jmen rodů/rodin, korporativních názvů a definováním vazeb mezi těmito entitami. Neřeší však problematiku označení a vztahů geografických entit, které jsou spolu s problematikou „pojmu“, „objektu“ a „události“ řešeny v konceptuálním modelu FRSAD (Functional Requirements for Subject Authority Data - Funkční požadavky na předmětové autoritní údaje) publikovaném v roce 2010. Všechny uvedené modely jsou považovány za entitně relační, které definují základní entity, jejich atributy a vztahy mezi těmito entitami. Zabývají se převážně tvorbou selekčních prvků/přístupových bodů a obsahují z pohledu muzejníků a archivářů málo kontextuálních informací. Jak již bylo uvedeno, jsou určeny primárně pro knihovnickou obec a nelze je bez úprav používat např. při popisu archiválií, muzejních exponátů a kolekcí.

Na tento neuspokojivý stav reaguje archivní a muzejnická obec publikováním „svých“ standardů: archiváři publikují v roce 2004 standard pro tvorbu autoritních záznamů korporativních a personálních záhlaví ISAAR CPF (geografické autority jsou opět opominuty),

protože „Archivní autoritní záznamy ... musejí podporovat mnohem širší soubor požadavků, než jaký existuje u záznamů knihovnických autorit. Tyto dodatečné požadavky jsou spojeny s významem dokumentování informací o původcích a kontextu vytváření archiválií v systémech archivního popisu. Jako takové jdou záznamy archivních autorit mnohem dále a obvykle obsahují mnohem více informací než záznamy knihovnických autorit³.“

V roce 2006 publikuje muzejnická obec standard CIDOC CRM (CIDOC Conceptual Reference Model) schválen jako ISO 21127:2006, který je založen na objektovém principu - definuje třídy entit a jejich vlastnosti; v přímé návaznosti na tuto normu vzniká a je v letech 2006 a 2008 publikována objektově orientovaná verze modelu FRBR, tedy FRBRoo.

Podobně je tomu i při tvorbě pravidel týkající se zápisu formy označení/pojmenování jednotlivých typů entit. Pravidla AACR (Anglo-American Cataloguing Rules) vznikla v roce 1967 v přímé návaznosti na výměnný formát MARC. V roce 1978 byla publikována revidovaná verze AACR2R, která se prakticky používá (u nás) dodnes. O vývoji nových pravidel se jedná od roku 1997, v roce 2004 bylo rozhodnuto, že pravidla AACR2R nebudou aktualizována a budou se vyvíjet pravidla nová - RDA (Resource Description and Access), která budou odpovídat konceptuálním modelům a potřebám webu, v současné době sémantického webu. Bylo přislíbeno, že v roce 2009/2010 bude publikována 1. verze těchto pravidel. Pravidla byla publikována a dána k dispozici až v květnu 2013. Navíc tato verze stále neobsahuje řešení pro zápisu entit „událost“, „pojem“ a „objekt“. Jde o složitou problematiku, jejíž řešení není jednoduché, protože se částečně odvíjí od existence jednotlivých národních jazyků. Tuto skutečnost jsme předvíдали a připravili pro projekt INTERPI dílčí řešení v intencích pravidel RDA, např. zápis označení událostí typu bitvy, války apod.

Pravidla pro tvorbu a zápis řízených selekčních prvků - AACR2R i nově implementovaná pravidla RDA jsou důležitá nejenom pro knihovnickou obec; představují důležitý standard i pro muzejnickou obec: pravidla CCO (Cataloguing of Cultural Objects) přímo vycházejí z těchto pravidel. Nepřímo na ně navazují i standardy určené pro archivní komunitu, i když tato se na ně neodvolává přímo. Ve standardu ISAAR CPF se uvádí „Tento standard byl vytvořen pro sdílení archivních autoritních záznamů na podporu přípravy konzistentních, vhodných a samo vypovídajících popisů korporací, osob a rodů, které vytvářejí dokumenty. Byl vytvářen tak, aby mohl být použit jako doplněk existujících národních standardů nebo jako základ pro vývoj národních standardů“ 4. K nejrozšířenějším národním standardům určených pro tvorbu popisných údajů patří právě pravidla AACR2R a nově implementovaná pravidla RDA. I nově vznikající národní pravidla, jako např. „Základní pravidla pro zpracování archiválií“⁵ v části „Tvorba přístupových rejstříků a přístupových bodů“ v podstatě vycházejí z principů pravidel AACR2R či RDA

³ http://asta.nacr.cz/data/fil_1831.pdf

⁴ http://asta.nacr.cz/data/fil_1831.pdf

⁵ <http://www.mvcr.cz/clanek/metodiky.aspx>

Nyní je potřeba ještě připravit národní interpretace pravidel RDA. Byla založena pracovní skupina pro implementaci pravidel RDA v Česku, jejímž je řešitel INTERPI členem, byl připraven harmonogram přechodu na aplikaci těchto pravidel: bylo rozhodnuto, že v Česku začneme používat pravidla RDA v květnu v roce 2015.

Podstatou metodiky jednotlivých tříd entit je definování preferované a variantních forem daných entit. Tyto formy jména však nelze stanovit bez rámcových pravidel, kterými jsou pravidla RDA.

Z těchto objektivních důvodů nemohla být dokončena tvorba dílčích metodik jednotlivých tříd entit. Tato část metodiky bude dokončena v roce 2014 a metodika bude v tomto roce předložena k certifikaci.