

## K otázkám metadatového popisu systémů organizace znalostí

Towards issues of descriptive metadata for knowledge organization systems

*PhDr. Eva Bratková, Ph.D., PhDr. Helena Kučerová / Ústav informačních studií a knihovnictví FF UK v Praze (Institute of Information Studies and Librarianship, Faculty of Arts, Charles University in Prague)*

### Resumé:

Předkládaná studie navazuje na studii zveřejněnou v roce 2014, jež byla věnována vymezení termínu „systém organizace znalostí“ a typologii těchto systémů. Nová studie je zaměřena na problematiku jejich metadatového popisu. Východiskem studie se stala reprezentativní množina záznamů systémů organizace znalostí zpracovaných a uložených v prototypu znalostní báze, která je připravovaná v rámci projektu DF13P01OVV013 programu NAKI „Znalostní báze pro obor organizace informací a znalostí“. Sběr dat reprezentujících tyto systémy poukázal na nezbytnost popisovat je jako entity velmi proměnlivého charakteru od doby jejich zrodu, přes jejich průběžné proměny až po jejich současné reprezentace ve volně dostupných strukturách propojených otevřených dat daných technologiemi sémantického webu. Studie představuje výsledek návrhu popisu těchto systémů ve znalostní bázi, který vychází z obecných principů konceptuálního modelu FRBR a také z návrhu metadatových prvků pro popis systémů organizace znalostí v rámci Aplikačního profilu NKOS, který připravuje Pracovní skupina DCMI NKOS (Dublin Core Metadata Initiative Networked Knowledge Organization Systems Task Group). Navržené metadatové schéma je porovnáno s existujícími metadatovými schématy, používanými pro podobné účely. Studie zahrnuje množinu vybraných příkladů popisu hlavních typů těchto systémů (klasifikační schémata, tezaury, předmětové hesláře).

**Klíčová slova:** systémy organizace znalostí, metadatový popis, metadatové prvky, Aplikační profil NKOS, konceptuální model FRBR, znalostní báze

### Summary:

Submitted study follows the previous one published in 2014 which was devoted to „knowledge organization system“ term definition and to typology issues of such systems. The new study is aimed at problems of their descriptive metadata. The study is based on representative set of knowledge organization system records processed and stored in the prototype of knowledge base, which is prepared within the project DF13P01OVV013 „Knowledge base for the subject area of knowledge organization“ as part of the NAKI Programme. The data collection representing these systems pointed out a necessity to describe these systems as entities of very changeable character, from the beginning of their creation, via their current transformations, up to their contemporary representations in free available structures of linked open data given by semantic web technologies. The study presents the result of these systems description in knowledge base proposal which is based on general principles of FRBR conceptual model, and on NKOS Application Profile – metadata elements for knowledge organization systems description being prepared by Dublin Core Metadata Initiative Networked Knowledge Organization Systems (DCMI NKOS) Task Group. Designed metadata schema is compared with existing metadata schemas used for similar purposes. The study comprises selected examples set of these systems description main types (classification schemes, thesauri, subject headings schemes).

**Keywords:** knowledge organization systems, metadata description, metadata elements, NKOS Application Profile, FRBR conceptual model, knowledge base

## Úvod

Předložená studie je věnována problematice **metadatového popisu systémů organizace znalostí**. Navazuje na předcházející studii z roku 2014<sup>1</sup>, věnovanou otázkám jejich typologie. Určení typu systému organizace znalostí (čes. zkr. SOZ) je jednou z důležitých hodnot jeho metadatového popisu. Nová studie je založena na reprezentativní množině záznamů SOZ zpracovaných v prototypu znalostní báze, která je připravována v rámci projektu programu NAKI „Znalostní báze pro obor organizace informací a znalostí“ (DF13P01OVV013). Jedním z cílů výzkumu bylo navrhnout koncepci a obsah popisu SOZ odpovídající cílům a funkcím znalostní báze. Sběr metadat reprezentujících SOZ poukázal na nezbytnost popisovat je jako entity velmi proměnlivého charakteru od doby jejich zrodu, přes jejich průběžné proměny až po jejich současné reprezentace ve volně dostupných strukturách propojených otevřených dat. Studie nejdříve představuje výsledky analýzy stávajícího metadatového popisu SOZ ve vybraných databázích a poté předkládá návrh jejich popisu ve znalostní bázi. Vychází z obecných principů konceptuálních modelů FRBR a FRSAD a také z návrhu metadatových prvků pro popis SOZ v rámci Aplikačního profilu NKOS (AP NKOS), který připravuje Pracovní skupina DCMI NKOS (DCMI Networked Knowledge Organization Systems Task Group). Navržené metadatové schéma přihlíží též k dalším existujícím metadatovým schémátům, používaným pro podobné účely (BARTOC, DCAT, DC-terms, FRBR/FRSAD, OMV, Schema.org, TaxoBank).

Text studie je členěn do tří částí. První je věnována výsledkům analýzy a vyhodnocení popisů SOZ ve stávající praxi bibliografických databází a také známých online registrech SOZ. Druhá část představuje návrh standardu pro popis SOZ skupiny DCMI NKOS. Závěry obou částí jsou východisky třetí části, která představuje návrh popisu těchto systémů v projektované znalostní bázi pro obor organizace informací a znalostí. Návrh popisu je doplněn vybranými příklady popisu zvolených typů SOZ (klasifikační schéma, tezaurus, předmětový heslář) ze znalostní báze s připojenými grafickými reprezentacemi.

## 1 Popis systémů organizace znalostí ve stávající praxi

Tato část studie předkládá výsledek analýzy metadatových popisů SOZ, které jsou dnes uplatňovány v různorodých online dostupných databázích. Vyhodnocen bude popis těchto systémů ve vybraných bibliografických a katalogizačních databázích, jež je popisují na úrovni publikací vedle jiných typů publikací, zejména ale pak ve specializovaných registrech nebo databankách, jež popisují výlučně SOZ na základě různých cílů. Vyhodnocováno je zejména množství a povaha uplatňovaných prvků (údajů).

### 1.1 Popis systémů organizace znalostí v bibliografických nebo katalogizačních databázích

Systémy organizace znalostí (klasifikační schémata, taxonomie, tezaury, ontologie aj.) jsou, zejména ve svých tradičních formách, běžně popisovány v **bibliografických**

---

<sup>1</sup> BRATKOVÁ, Eva, KUČEROVÁ, Helena. Systémy organizace znalostí a jejich typologie. In: *Knihovna*. 2014, **25**(2), 5-29. ISSN 1801-3252 (Print). ISSN 1802-8772 (Online). Dostupný také z: <http://oldknihovna.nkp.cz/pdf/1402/142005.pdf>. Anglická verze dostupná z: <http://oldknihovna.nkp.cz/pdf/1402sup/142001.pdf>.

**databázích** národně registrujícího charakteru a také **databázích lokálních a souborných katalogů**. SOZ nemusejí být – z různých důvodů – v uvedených typech databází zaznamenány v úplnosti (například SOZ produkované komerčními producenty jako šedá literatura), popřípadě jsou zaznamenány torzovitě, anebo nejsou zaznamenány vůbec. Zaznamenávány jsou již také SOZ v online formě, dostupné volně nebo licenčně. SOZ zpřístupňované nově ve strukturách propojených otevřených dat jsou v katalogích reprezentovány zatím málo nebo vůbec ne. Zjišťování záznamů SOZ v průběhu jednoho roku v uvedených typech databází potvrdilo předpoklad, že katalogizační záznamy nejsou vhodnou reprezentací těchto systémů pro projektovanou znalostní bázi. Z podobných důvodů zřejmě nejsou využívány ani v systémech zapojených do prostředí sémantického webu. Nelze jich využít také z důvodu menšího počtu popisných údajů včetně nerelevantních struktur pro jejich současnou komunikaci a užití.

Nicméně, bibliografické a katalogizační databáze posloužily v rámci výzkumné úlohy jako zdroje pro zjišťování existence SOZ včetně přebírání jejich identifikačních údajů. Získané údaje musely ale být následně kontrolovány, sjednocovány a doplňovány o další údaje, specifické pouze pro SOZ, a to s ohledem na cíle výzkumu. Prozatím bylo do znalostní báze uloženo přes 230 záznamů, jež reprezentují necelých 100 významných SOZ<sup>2</sup>. Předmětem zjišťování byly jak SOZ, které vycházely v různých intervalech na papírových nosičích, tak SOZ, které jsou zveřejňovány na nosičích elektronických, počínaje jejich vydáváním na CD-ROM, přes zpřístupňování v režimu online bází, anebo jejich nejnovější volně zpřístupňování formou vysoce strukturovaných propojených otevřených dat.

Jako náročný úkol se ukázalo zejména zjišťování existence SOZ na tradičních papírových nebo pevných elektronických nosičích. V tomto případě řešitelský tým pro excerpní práce využil v maximální možné míře zejména databáze souborného katalogu **WorldCat** v kombinaci s databázemi národních bibliografií, popřípadě také národních souborných katalogů<sup>3</sup>. V případě, že ve zvolených databázích záznamy SOZ nebyly nalezeny, anebo byly popsány minimálně nebo rozporuplně, bylo nutné zjišťovat informace v **dalších zdrojích**. Konzultována byla, existuje-li, zejména literatura o jednotlivých SOZ, nebo přímo SOZ v tradičních, spíše ale v digitalizovaných formách (trvale uchovávaných například v systémech HathiTrust Digital Library, Internet Archive aj.). Pro kontrolu byly sledovány také záznamy SOZ dostupné od roku 2012 v systému **Google Books**, byť jde o záznamy převzaté z databáze WorldCat prostřednictvím jejich volného vystavení v nové struktuře „Schema.org“.

Jak bylo naznačeno výše v textu, **katalogizační záznamy** uložené primárně zpravidla ve formátech typu MARC nemají z hlediska cílů výzkumného projektu určitě dostatečné množství údajů (viz obr. 1). Postrádají některé specifické údaje (například údaj o *typu SOZ*, *vnitřní struktuře SOZ*, *počtu lexikálních jednotek* aj.). Nemají v sobě uveden explicitní údaj o tom, že je popisován co do druhu materiálu SOZ, nebylo by proto možné přímo je automaticky vyhledat. Záznamy nejsou, až na výjimky, opatřeny *abstrakty* nebo *anotacemi*. U SOZ s dlouhým vývojem nejsou zaznamenávány potřebné *vztahy mezi SOZ* – je obtížné „vystopovat“ jednotlivá za sebou jdoucí vydání či verze systému, zejména v případě změn jejich názvů nebo jiných vlastností. U většiny záznamů chybí

<sup>2</sup> Ukládání dalších záznamů SOZ pokračuje i v roce 2015.

<sup>3</sup> Například identifikace jednotlivých tištěných vydání a verzí dánského Desetinného třídění (Decimalklassedeling) byla relativně úspěšně vyřešena až prostřednictvím dánského národního souborného katalogu veřejných knihoven „bibliotek.dk“ v kombinaci s katalogem Královské knihovny.

jednotlicí údaj typu *Unifikovaný název*, popřípadě údaj typu *Autor–Název*, jež by pomohly jejich efektivnímu online vyhledání. Prozatím chybí, až na malé výjimky, jedinečné a trvalé *identifikátory děl nebo vyjádření děl*<sup>4</sup>. Předností naopak bývá relativně kvalitní uvádění údajů o *alternativních názvech SOZ, údajů o vydání nebo verzi SOZ* nebo údajů o *jazyku a zemi* jejich publikování. Katalogizační záznamy SOZ a jejich jednotlivé údaje mají v tradičních formátech MARC (včetně formátů MARCXML nebo MODS) statický ráz, a nejsou tudíž vhodné k dynamickému propojování v prostředí sémantického webu.

```
>040  GEBAY $e rakwb $b ger $c GEBAY $d OCLCQ
>015  05,N36,0101 $2 dnb
>016 7  976026287 $2 DE-101
>020  3598116519
>020  9783598116513
>080
>084  AN 93550 $2 rvk
>049  CQKA
>245 00 Dewey-Dezimalklassifikation und Register : $b DDC 22 / $c begr. von Melvil
      Dewey. Hrsg. von Joan S. Mitchell. Hrsg. von Der Deutschen Bibliothek.
>246 3  DDC 22
>250  Dt. Ausg.
>260  München : $b Saur.
>515  Erschienen: 1 - 4.
>630 07 Dewey-Dezimalklassifikation. $2 swd
>650 07 Inhaltserschließung. $2 swd
>650 07 Bibliothek. $2 swd
>650 07 Klassifikation. $2 swd
>650 07 Übersetzung. $2 swd
>651  7 Deutsch. $2 swd
>655  7 Richtlinie. $2 swd
>700  1 Dewey, Melvil.
>700  1 Mitchell, Joan S.
>710  2 Deutsche Bibliothek (Frankfurt, Main; Leipzig)
>730  0 Dewey decimal classification and relative index. $l Deutsch.
>856 41 $3 Inhaltsverzeichnis $u http://bvbr.bib-bvb.de:8991/F?...
>029 0  GEBAY $b 7716793
```

Obr. 1 Záznam (OCLC 643133344) německého překladu 22. vydání DDT v rádkovém formátu MARC Text Area (bez kódovaných polí 00X) z databáze WorldCat; záznam obsahuje v poli 730 unifikovaný název úrovně vyjádření (OCLC, 2015)

Katalogizační záznamy uložené a komunikované ve formátech typu MARC zahrnují, a to i v případě že jsou konvertovány do některého z formátů propojených otevřených dat, prozatím pouze určitou základní množinu popisných údajů o SOZ, jež byly využity

<sup>4</sup> V databázi VIAF se začaly objevovat již také první trvalé identifikátory (VIAF ID) děl (works) některých SOZ, k dispozici je například identifikátor (URI) pro MDT (<http://viaf.org/viaf/184301709/>) nebo pro LCC (<http://viaf.org/viaf/203733980/>), nebo také identifikátory vyjádření děl (expressions), například identifikátor ruského překladu MDT (<http://viaf.org/viaf/185511525/>) nebo identifikátor českého překladu vybraných znaků MDT (<http://viaf.org/viaf/186352368/>).

v rámci výzkumného projektu. Patří k nim: *názvové údaje*, které jsou deskriptivního nebo formalizovaného charakteru, *údaje o tvůrcích* v autoritních tvarech, *údaje o vydání nebo verzi SOZ* (ty jsou v případě SOZ velmi užitečné), *nakladatelské údaje*, *trvalé identifikátory* typu URI, ISBN a ISSN, počty stran v případě papírových nosičů, *věcný popis* ve formě znaků zpravidla některé z univerzálních klasifikací (DDT, MDT, LCC aj.) nebo termínů z řízených slovníků, *údaje o edicích*, někdy i bohatší *údaje poznámkového charakteru* a také *lokalizace* jednotek v knihovnách nebo jiných informačních institucích.

## 1.2 Popisná metadata SOZ ve schématu Schema.org

V kontextu informací uvedených v části 1.1 je možné také analyzovat a vyhodnotit popis SOZ uvedený v jednom z nejnovějších metadatových schémat – **Schema.org** (se jmenným prostorem <http://schema.org/>, viz xmlns:schema v deklaraci formátu rdf na obr. 2), a to na základě zkušeností, které učinila síť OCLC se záznamy svého souborného katalogu **WorldCat**. Schéma je v současnosti využíváno globálně pro potřeby volně zpřístupňovaných propojených otevřených dat (Linked Open Data, LOD), která mohou popisovat **cokoliv** (v terminologii schématu jde o **popisování „věcí“**, angl. „**things**“), tedy i **informační zdroje** včetně SOZ. Stručná prezentace bude uvedena na základě volně dostupných propojených dat o SOZ poskytovaných databází WorldCat, která jsou uložena jako mikrodata (Microdata Section) ve struktuře RDFa na webových stránkách zobrazovaných katalogizačních záznamů v kódu HTML. Síť OCLC byla jednou z prvních, která katalogizační záznamy tímto způsobem zpřístupnila a zvolila si k tomu právě Schema.org<sup>5</sup>. Docílila tím velkého efektu – informace z katalogu se rychle dostávají ke koncovým uživatelům, zejména k těm, kteří je přijímají z mobilních zařízení, a to prostřednictvím vyhledávače Google<sup>6</sup>, který data převedl také do systému Google Books.

Před prezentací uplatněných prvků metadatového schématu je nezbytné poznamenat, že nejde co do koncepce o „klasický“ (statický) popis zdrojů, kdy jsou vlastnostem zdrojů v rámci daných popisných prvků přidělovány příslušné hodnoty. V koncepci LOD jde o propojování přidělených hodnot s hodnotami ve volně dostupných řízených slovnících, anebo přímé propojování s hodnotami těchto slovníků pomocí identifikátoru URI. Schema.org umožňuje dynamicky generovat popis určitého zdroje pomocí kombinace různých prvků sdružených v rámci definovaných základních objektových kategorií (tvůrčí dílo, událost, akce, organizace, fyzická osoba, místo, produkt, nematná věc aj.).

Systémy organizace znalostí, prezentované v databázi WorldCat jako LOD, zahrnují v rámci úvodního označení typu popisovaného zdroje v prvku `<schema:about>` jednoznačnou identifikaci záznamu (URI), který SOZ popisuje – viz obr. 2 (první výskyt prvku „rdf:Description“). Využito je jedinečného čísla záznamu v databázi WorldCat. Samotná databáze je identifikována pomocí prvku `<void:inDataset>` z jiného jmenného prostoru „VOID“ (<http://rdfs.org/ns/void#>).

Základní data popisovaného SOZ je možné uvádět jako popis některého z typů **Tvůrčích děl** (CreativeWork). K dispozici je například typ **Knihy** (<http://schema.org/Book>) – tento typ platí též pro SOZ v tištěné formě. Pomocí prvku `<schema:bookFormat>`

<sup>5</sup> FONS, Ted, PENKA, Jeff, WALLIS, Richard. OCLC's Linked Data Initiative: Using Schema.org to Make Library Data Relevant on the Web. In: *Information Standards Quarterly*. Spring/Summer 2012, **24**(2/3), s. 29. ISSN 1041-0031. Dostupné také z: [http://www.niso.org/apps/group\\_public/download.php/9408/IP\\_Fons-etal\\_OCLC\\_isqv24no2-3.pdf](http://www.niso.org/apps/group_public/download.php/9408/IP_Fons-etal_OCLC_isqv24no2-3.pdf).

<sup>6</sup> Metadatové schéma Schema.org připravily ve spolupráci společnosti Google, Bing, Yahoo a Yandex.

Ize ale specifikovat dílčí formu, například formu **Elektronická kniha** (<http://schema.org/EBook>), pod níž zatím spadá také **online dostupný web** – viz obr. 2. Jemnější kategorizace pro potřeby popisu SOZ zatím není k dispozici. Hlavní název SOZ je pak možné uvádět v prvku `<schema:name>` a další názvy v prvku `<schema:alternateName>`. Oba typy názvů je vhodné specifikovat navíc vnitřním atributem kódu jazyka názvu. Pokud je k dispozici unifikovaný název díla, bude uveden přímo v rámci popisu Díla (<http://schema.org/CreativeWork>) v prvku `<schema:name>`. Schema.org reflektuje v tomto směru koncepci modelu FRBR.

Údaje o vydání SOZ lze uvádět v prvku `<schema:bookEdition>`, jméno nakladatele v prvku `<schema:publisher>` a datum vydání v prvku `<schema:datePublished>`. Jde-li o SOZ jako pokračující zdroj (web apod.), lze využít i prvku `<schema:startDate>`, a to v rámci popisu objektu **Údátost publikování** (<http://schema.org/PublicationEvent>), včetně místa této události v prvku `<schema:location>` a jejího organizátora, tedy vydavatele v prvku `<schema:organizer>`. Nakladatele (stejně tak korporativní tvůrce SOZ) lze ale navíc také popsat jménem v prvku `<schema:name>` v rámci popisu objektu **Organizace** (<http://schema.org/Organization>). Stejně tak místo vydání lze popsat jménem v rámci popisu **Místa** (<http://schema.org/Place>). Uvádění místa vydání pomocí kódu lze realizovat na základě již existujícího slovníku míst (<http://id.loc.gov/vocabulary/countries>) v rámci prvku `<dcterms:identifiers>` ze jmenného prostoru „dcterms“ (<http://purl.org/dc/terms/>). Jazyk zdroje lze uvést kódem v prvku `<schema:inLanguage>` v propojení na zvolený číselník (například <http://id.loc.gov/vocabulary/iso639-1>). Trvalé identifikátory ISBN lze uvádět v prvku `<schema:ISBN>` v rámci popisu objektové kategorie **Model produktu** (<http://schema.org/ProductModel>).

Jména tvůrců SOZ typu fyzická osoba je možné popisovat jejich jmény v rámci popisu objektu **Fyzická osoba** (<http://schema.org/Person>), uplatnit lze nejenom již zmíněný prvek `<schema:name>`, ale také detailnější prvky `<schema:familyName>`, `<schema:givenName>`, `<schema:birthDate>` nebo `<schema:deathDate>`. Jména fyzických osob a také korporací jsou již ve velkém počtu odkazována do databáze systému VIAF zahrnující jedinečné identifikátory VIAF ID a popř. také již mezinárodní identifikátory ISNI.

**Věcný popis** SOZ (též jiných materiálů) lze realizovat především na základě propojování k hodnotám velkého počtu systémů organizace znalostí (řízených slovníků aj.), a to v rámci objektové kategorie **Nehmotná (věc)** (<http://schema.org/Intangible>). Ukázky lze vidět na obr. 2. Záznamy z databáze WorldCat mohou být propojovány například přímo na třídíky Deweyho desetinného třídění na serveru <http://dewey.info> (viz <http://dewey.info/class/025.431/>), nebo na řízené předmětové termíny zpřístupněné OCLC v novém systému FAST (například předmětový termín „Classification, Dewey decimal“ je již volně k dispozici na serveru propojených dat OCLC (<http://experimental.worldcat.org/fast/863693/>)<sup>7</sup>. Pro text abstraktu nebo anotace je ve schématu připraven prvek `<schema:abstract>` – ukázku lze vidět na obr. 2.

<sup>7</sup> Záznam Deweyho desetinného třídění ve formě databáze WebDewey, prezentovaný na obr. 2, je na úrovni věcného popisu řešen zcela pragmaticky: má přidělen (a je propojen na) znak „025.431“ (Dewey Decimal Classification) a také na řízené heslo systému FAST „Classification, Dewey decimal“. Je zjevné, že popisované klasifikační schéma nepojednává „samo o sobě“, a tudíž se nejedná o vyjádření věcného obsahu, nýbrž o formální pojmenování SOZ.



```

<rdf:RDF
  xmlns:wdrs="http://www.w3.org/2007/05/powder-s#"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:library="http://purl.org/library/"
  xmlns:genont="http://www.w3.org/2006/gen/ont#"
  xmlns:dcterms="http://purl.org/dc/terms/"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#"
  xmlns:void="http://rdfs.org/ns/void#"
  xmlns:schema="http://schema.org/"
  xmlns:bgn="http://bibliograph.net" >
  <rdf:Description rdf:about="http://www.worldcat.org/title/-/oclc/49510336">
    <rdf:type rdf:resource="http://www.w3.org/2006/gen/ont#ContentTypeGenericResource"/>
    <rdf:type rdf:resource="http://www.w3.org/2006/gen/ont#InformationResource"/>
    <schema:about rdf:resource="http://www.worldcat.org/oclc/49510336"/>
    <void:inDataset rdf:resource="http://purl.oclc.org/dataset/WorldCat"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://id.loc.gov/vocabulary/countries/ohu">
    <rdf:type rdf:resource="http://schema.org/Place"/>
    <dcterms:identifier>ohu</dcterms:identifier>
  </rdf:Description>
  <rdf:Description rdf:about="http://experiment.worldcat.org/entity/work/data/38483699#Place/dublin_ohio">
    <rdf:type rdf:resource="http://schema.org/Place"/>
    <schema:name>Dublin, Ohio</schema:name>
  </rdf:Description>
  <rdf:Description rdf:about="http://experiment.worldcat.org/entity/work/data/38483699#Agent/oclc_forest_press">
    <rdf:type rdf:resource="http://bibliograph.net/Agent"/>
    <schema:name>OCLC Forest Press</schema:name>
  </rdf:Description>
  <rdf:Description rdf:about="http://id.worldcat.org/fast/863707">
    <rdf:type rdf:resource="http://schema.org/Intangible"/>
    <schema:name xml:lang="en">Classification, Library of Congress.</schema:name>
  </rdf:Description>
  <rdf:Description rdf:about="http://viaf.org/viaf/148269885">
    <rdf:type rdf:resource="http://schema.org/Organization"/>
    <schema:name>Forest Press.</schema:name>
  </rdf:Description>
  <rdf:Description rdf:about="http://id.worldcat.org/fast/863693">
    <rdf:type rdf:resource="http://schema.org/Intangible"/>
    <schema:name xml:lang="en">Classification, Dewey decimal.</schema:name>
  </rdf:Description>
  <rdf:Description rdf:about="http://www.worldcat.org/title-/oclc/49510336#PublicationEvent/dublin_ohio_oclc_forest_press_2002">
    <rdf:type rdf:resource="http://schema.org/PublicationEvent"/>
    <schema:location rdf:resource="http://experiment.worldcat.org/entity/work/data/38483699#Place/dublin_ohio"/>
    <schema:organizer rdf:resource="http://experiment.worldcat.org/entity/work/data/38483699#Agent/oclc_forest_press"/>
    <schema:startDate>2002</schema:startDate>
  </rdf:Description>
  <rdf:Description rdf:about="http://id.worldcat.org/fast/1136458">
    <rdf:type rdf:resource="http://schema.org/Intangible"/>
    <schema:name xml:lang="en">Subject headings.</schema:name>
  </rdf:Description>
  <rdf:Description rdf:about="http://dewey.info/class/025.431"/>
    <rdf:type rdf:resource="http://schema.org/Intangible"/>
  </rdf:Description>
  <rdf:Description rdf:about="http://www.worldcat.org/oclc/49510336">
    <wdrs:describedby rdf:resource="http://www.worldcat.org/title/-/oclc/49510336"/>
    <schema:bookFormat rdf:resource="http://schema.org/EBook"/>
    <schema:about rdf:resource="http://id.worldcat.org/fast/863693"/>
    <schema:publication rdf:resource="http://www.worldcat.org/title-/oclc/49510336#PublicationEvent/dublin_ohio_oclc_forest_press_2002"/>
    <library:placeOfPublication rdf:resource="http://experiment.worldcat.org/entity/work/data/38483699#Place/dublin_ohio"/>
    <schema:description xml:lang="en">An easy-to-use, browser-based interface that allows you to search the DDC (and related terminology) efficiently and navigate intuitively. Includes thousands of Relative Index terms and built numbers not available in the print DDC, Library of Congress Subject Headings (LCSH) that have been intellectually mapped to Dewey headings by DDC editors, selected LCSH mappings from the new OCLC Forest Press publication, People, Places & Things,
  </schema:description>
  </rdf:Description>

```

```

LCSH that have been statistically mapped to Dewey numbers from records in WorldCat (the OCLC Online Union Catalog),
links from mapped LCSH to the LCSH authority records, quarterly updates, incorporating the latest changes to the
Classification and new LCSH mappings, index terms, and built numbers, and an annotation capability, which allows you to
add your own notes into WebDewey to reflect local classification practices.</schema:description>
<schema:alternateName xml:lang="en">Web Dewey</schema:alternateName>
<schema:about rdf:resource="http://id.worldcat.org/fast/1136458"/>
<rdf:value rdf:parseType="Literal">Unknown value: rs</rdf:value>
<schema:inLanguage>en</schema:inLanguage>
<rdf:type rdf:resource="http://schema.org/Book"/>
<schema:contributor rdf:resource="http://viaf.org/viaf/148269885"/>
<rdf:type rdf:resource="http://schema.org/MediaObject"/>
<schema:exampleOfWork rdf:resource="http://worldcat.org/entity/work/id/38483699"/>
<library:oclcnum rdf:datatype="http://www.w3.org/2001/XMLSchema#integer">49510336</library:oclcnum>
<schema:about rdf:resource="http://dewey.info/class/025.431"/>
<schema:about rdf:resource="http://id.worldcat.org/fast/863707"/>
<library:placeOfPublication rdf:resource="http://id.loc.gov/vocabulary/countries/ohu"/>
<rdf:value rdf:parseType="Literal">Unknown value: map</rdf:value>
<schema:datePublished>2002</schema:datePublished>
<rdf:value rdf:parseType="Literal">Unknown value: dcl</rdf:value>
<schema:publisher rdf:resource="http://experiment.worldcat.org/entity/work/data/38483699#Agent/oclc_forest_press"/>
<schema:name xml:lang="en">WebDewey</schema:name>
<schema:datePublished>2002/9999</schema:datePublished>
</rdf:Description>
</rdf:RDF>

```

Obr. 2 (na str. 11–12) Záznam online databáze WebDewey

(permalink <http://www.worldcat.org/oclc/49510336>) z databáze WorldCat ve formě propojených otevřených dat (rdf.xml)

(OCLC, 2015, žlutě jsou autorkami vyznačeny vybrané prvky schématu Schema.org, modře jejich hodnoty)

Metadatové schéma **Schema.org** prozatím neumožňuje popis specifických vlastností SOZ (typ SOZ, vnitřní vztahy v SOZ, struktura SOZ aj.). Totéž platí i pro jiné specifické typy zdrojů. Schéma se ale dále rozvíjí, pracuje se na dalších podrobnostech iniciovaných různými komunitami, které mají zájem popisovat „věci“. Konkrétně pro oblast popisu informačních zdrojů vyvíjí velkou aktivitu odborná skupina **SchemaBibEx** (<https://www.w3.org/community/schemabibex/>), kterou vede pracovník OCLC Richard Wallis. Náměty lze vznášet prostřednictvím e-mailové nebo chatové komunikace.

### 1.3 Popisná metadata SOZ v registru BARTOC

Tato část prezentuje výsledek analýzy popisu SOZ v registračním systému **BARTOC**<sup>8</sup>. Jeho typologie systémů byla diskutována v již ve výše citované studii<sup>9</sup>. Registr je v provozu od roku 2013 a v srpnu 2014 bylo v jeho databázi 667 záznamů. V dubnu 2015 byl počet zaregistrovaných systémů téměř dvojnásobný (1 137 záznamů).

V bázi registru **BARTOC** jsou SOZ popisovány z hlediska úrovně popisu pouze **jedním záznamem**. Popisnou jednotkou je **systém jako celek** bez ohledu na nosič nebo formát, s důrazem na jeho současnou, zejména elektronickou formu a online přístup.

<sup>8</sup> Universitätsbibliothek Basel. *BARTOC.org: BAseL Register of Thesauri, Ontologies & Classifications* [online]. Projektleiter Andreas LEDL. Basel: Universitätsbibliothek Basel, 2013- [cit. 2015-02-25]. Dostupné volně ze serveru Basilejské univerzity: <http://www.bartoc.org/>.

<sup>9</sup> BRATKOVÁ, Ref. č. 1, s. 19-20.



Historický vývoj v podobě zaznamenávání průběžných tištěných produktů a produktů na CD-ROM není realizován, prozatím je pouze v některých záznamech uveden stručný vývoj v krátkém abstraktu. Popis SOZ není založen na modelu FRBR. Tato koncepce je dána hlavním cílem registru – podat uživatelům rychle přehlednou informaci o existenci příslušného SOZ. Celkový počet metadatových prvků proto není velký a většina jejich hodnot je trvale uložena ve vlastním strukturovaném slovníku – **taxonomii termínů**<sup>10</sup>. Předností registru je, že záznamy jsou vedle online zpřístupňování k dispozici volně také ve více formátech jako otevřená propojená data. V rámci online komunikace s registrem BARTOC lze využívat až 20 jazyků (jazykové ekvivalenty se týkají dokonce i řady hodnot popisných údajů).

Registr BARTOC prozatím nezpřístupnil ve své veřejné dokumentaci **specifikaci metadat**, která je využívána pro popis zahrnutých systémů. Následující komentář proto vychází pouze z prezentační formy zobrazovaných metadatových záznamů v rámci online vyhledávacího systému s přihlédnutím k metadatům lokalizovaným ve volně dostupných záznamech ve struktuře rdf.xml. Příklad popisu známého tezauru AGROVOC je na obr. 3.

Každý **metadatový záznam** registru BARTOC zahrnuje vždy jedinečný a trvalý *Identifikátor samotného záznamu* (URI), který hraje důležitou roli při jakémkoli propojování. Při náhledu záznamu přes internetový prohlížeč je URI zobrazován v adrese (například tezaurus AGROVOC má URI <http://www.bartoc.org/node/305>). Je-li zobrazován záznam ve strukturách propojených dat, je k základnímu URI připojována příslušná extenze (<http://www.bartoc.org/node/305.rdf>, <http://www.bartoc.org/node/305.xml> aj.). Součástí záznamu jsou také údaje o *Datu vytvoření* (Date created) a *Datu aktualizace* (Date modified) záznamu. Tato technická metadata nejsou zobrazována v průběhu online vyhledávání, jsou ale součástí záznamů propojených dat, a to prostřednictvím metadat definovaných specifikací *Dublin Core* s atributem typu dat, například:

```
<dc:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">
2013-09-10T15:34:42+02:00</dc:created>
<dc:modified rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">
2014-08-12T14:27:30+02:00</dc:modified>
```

V každém záznamu je uveden jedinečný *Název* (Title) SOZ jako celku, který má zjevnou povahu „jednotného“ či „unifikovaného“ názvu. Jeho podobu v jazyce popisovaného systému (například „Dewey Decimal Classification“, „Gemeinsame Normdatei Online“, „Kielitieteen ontologia“) stanovuje podle vlastního rozhodnutí producent registru. Některé názvy zahrnují také geografický doplněk v kulaté závorce, výjimečně název zahrnuje také zkratku bez oddělení interpunkcí (např. „JITA Classification System of Library and Information Science“). V záznamu jsou i *Alternativní názvy* (Alternative Title) SOZ, pod kterými by je mohl uživatel hledat. Může jít o zkratky (například „DDC“, „EuroVoc“ apod.), jindy producent zvolil jako alternativu úplný oficiální název (například „The Life Science Thesaurus“ pro známý komerční tezaurus „Emtree“, který je v roli unifikovaného názvu).

<sup>10</sup> Slovník je uložen v adresáři: <http://www.bartoc.org/taxonomy/term/>.

# AGROVOC

View

Voting results

## Abstract:

AGROVOC is the world's most comprehensive multilingual agricultural vocabulary. Organized as a concept scheme, AGROVOC contains close to 40,000 concepts. Downloaded over a thousand times a year by dozens of countries it is in daily institutional use to index and search documents, web pages and digital objects.\*

**Author:** Food and Agriculture Organization of the United Nations (FAO)

**VIAF:** Food and Agriculture Organization of the United Nations (FAO) <sup>Ⓔ</sup>

**Link:** <http://aims.fao.org/vest-registry/vocabularies/agrovoc-multilingual-agricultural-thesaurus> <sup>Ⓔ</sup>

**Topic:** agronomy

**DDC:** 630

**Access:** Free

**Format:** Floppy Disc Online Printed RDF SKOS

**Type:** Thesaurus

## Language:

Arabic Chinese Czech English French German Hindi Hungarian Italian Japanese Korean  
Lao Malay Persian Polish Portuguese Slovak Spanish Telugu Thai Turkish

**Geo:** POINT (12.4885194 41.882769)

## Rate



EuroVoc URI:

<http://eurovoc.europa.eu/685> <sup>Ⓔ</sup>

ISO 639-2 URI:

<http://id.loc.gov/vocabulary/iso639-2/ara> <sup>Ⓔ</sup>

Obr. 3 Záznam tezauru AGROVOC v prezentačním formátu v registru BARTOC.org (ve zvolené anglické verzi jazykové komunikace, 2015)

V prvku *Autor* (Author) jsou uváděna ponejvíce jména společností nebo organizací, které jsou za SOZ zodpovědné. Například v záznamu DDT (<http://www.bartoc.org/node/241>) je jako autor uvedeno centrum „Online Computer Library Center (OCLC)“, jméno původního autora M. Deweyho je pouze stručně zmíněno v krátkém abstraktu. Existuje-li, pak je jméno autora propojeno pomocí hyperlinku v rámci prvku *VIAF* se záznamy z databáze VIAF prostřednictvím URI, jehož součástí je jedinečný identifikátor VIAF (VIAF ID), například hyperlink <http://viaf.org/viaf/156508705> identifikuje výše zmíněné centrum OCLC.

Prvek *Vazba* (Link) je určen k propojení záznamu s webovou stránkou, která SOZ popisuje, anebo ze které je SOZ online dostupný. Vazba na webovou stránku portálu AIMS, ze které je dostupný thesaurus AGROVOC, je zobrazena na obr. 3. V záznamu tezauru EuroVoc je zase vazba mířící na jeho hlavní domovskou stránku (<http://eurovoc.europa.eu/>).

K velmi cenným informacím v krátkém záznamu registru patří **věcné údaje**. Zahrnut je *Abstrakt* (Abstract) v anglickém jazyce. Má informativní ráz a obsahuje různorodé informace o oblasti použití SOZ, o jeho vývoji, ale někdy také počtu jeho jednotek v současnosti. Pro údaj *Téma* (Topic) jsou vhodně využívány hodnoty známého více-

jazyčného tezauru EuroVoc . Identifikátory URI deskriptorů tezauru EuroVoc jsou zobrazovány za základním záznamem. Velkou předností registru je organizování záznamů pomocí schématu DDT v prvku *DDC* (užívané je až na třetí hierarchickou úroveň). Třídníky jsou opět součástí vlastní taxonomie termínů (například hodnota „630“ pro zemědělství je uvedena na URI <http://www.bartoc.org/taxonomy/term/10622>). Propojování do záznamů k volně online dostupné databázi DDT (<http://dewey.info>) prozatím není realizováno.

K formálním údajům záznamu patří prvek *Přístup* (Access) k SOZ, který je pojat v právním slova smyslu. Užíváno je jednoduchých hodnot, které jsou opět součástí vlastní taxonomie termínů, například „Volný“ („Free“, <http://www.bartoc.org/taxonomy/term/1377>) nebo „Registrovaný“ („Registered“, <http://bartoc.org/taxonomy/term/17>). Prvek *Formát* (Format) SOZ zahrnuje až trojí typ hodnot (jsou také součástí vlastní taxonomie): jednak formu zpřístupnění (tištěný, online, CD-ROM aj.), označení formátu (RDF, XML aj.) nebo také strukturu (například OWL, SKOS aj.). Posledním a podstatným prvkem je prvek *Typ* (Type) SOZ, který užívá hodnot vlastní definované typologie BARTOC. Počet typů se zvedl z pěti (v srpnu 2014) na současných 6 (typ „Glosář“ přibyl nedávno). Typy jsou také součástí vlastní taxonomie (typ „Classification“ má URI <http://www.bartoc.org/taxonomy/term/3>). Vlastní taxonomie termínů ještě zahrnuje hodnoty pro prvek *Jazyk* (Language) SOZ (URI <http://www.bartoc.org/taxonomy/term/10747> patří například francouzštině). Pokud existuje, je záznam registru BARTOC propojen navíc také pomocí hyperlinku s heslem anglické verze Wikipedie (v zásadě už ale jde o odkaz na literaturu o systému).

## 1.4 Popisná metadata SOZ v registrační databance TaxoBank

Jinou koncepcí popisných metadat SOZ co do jejich rozsahu a obsahu reprezentuje registrační databanka **TaxoBank**<sup>11</sup>, která je v provozu od roku 2009. Je spravována odborným týmem „taxonomistů“ soukromé americké společnosti „Access Innovations“, specializující se na služby v oblasti softwarových aplikací se zaměřením na lingvistické aspekty dat ukládaných v databázích. Záznamy registrační databanky obsahují **velmi podrobné a kvalitně strukturované údaje o „řizených slovnících“**, tedy systémech organizace znalostí<sup>12</sup>. Databanka je doplňována nepravidelně a náměty k registraci SOZ mohou zasílat také běžní uživatelé internetu, záznam ale vždy zpracovává odborný pracovník. Zaregistrovaných je v současnosti přibližně **250** systémů organizace znalostí.

Rovněž databanka TaxoBank nezpřístupnila ve své veřejné dokumentaci **specifikaci metadat**. Proto je i následující komentář založen na záznamech zobrazovaných v prezentačním formátu v rámci online vyhledávacího systému. Příklad kratšího popisu známého tezauru AGROVOC je na obr. 4. Delší popis představuje například záznam významného amerického zemědělského tezauru NAL<sup>13</sup>.

<sup>11</sup> Access Innovations. *TaxoBank Terminology Registry: TaxoBank* [online]. Albuquerque (New Mexico, USA): Access Innovations, 2009- [cit. 2015-02-25]. Dostupné z: <http://www.taxobank.org/>.

<sup>12</sup> Rozbor typologie KOS a užívané terminologie, viz: BRATKOVÁ, Ref. č. 1, s. 18-19.

<sup>13</sup> Podrobný záznam tezauru NAL je na URL: <http://www.taxobank.org/content/nal-agricultural-thesaurus>.

<http://www.taxobank.org/content/agrovoc-thesaurus>

AGROVOC Thesaurus

Thu, 2009-11-19 08:42 — [aii\\_admin](#)

AGROVOC is a multilingual, structured and controlled vocabulary designed to cover the terminology of all subject fields in agriculture, forestry, fisheries, food and related domains (e.g. environment). The AGROVOC Thesaurus was developed by FAO and the Commission of the European Communities in the early 1980s. Since then it has been updated continuously by FAO and local institutions in member countries.

#### ▼ General information

Vocabulary type: thes

Vocabulary Sample URL: <http://aims.fao.org/agrovoc/page?c=12332>

Was vocabulary created as a course project: 0

#### ▼ Vocabulary characteristics

Type of display: alph, html, diag, perm, other

Relationship types: eq\_pri\_eq, hier\_bn, rel\_t, ont

Characteristics Comments:

„AGROVOC Linked Open Data (LOD) is a project to turn the AGROVOC thesaurus into a multilingual, terminological backbone for agricultural digital goods. Hosted by research partner [MIMOS](#) it provides web-accessible, structured data records on agricultural concepts and even more importantly, links those concepts to other online thesauri.“

#### ▼ Terms and Conditions

Availability Comments: available also via webservice

Import/download instructions :

See <http://aims.fao.org/website/Download/sub>

#### ▼ Provider

Vocabulary provider name: FAO

Provider URL: <http://www.fao.org/>

Provider contact details:

„National organizations and institutes are welcome to help enrich and maintain AGROVOC's many languages by joining our growing community. Please contact us at <http://aims.fao.org/contact>“

Provider section comments:

„The Food and Agriculture Organization of the United Nations leads international efforts to defeat hunger. Serving both developed and developing countries, FAO acts as a neutral forum where all nations meet as equals to negotiate agreements and debate policy. FAO is also a source of knowledge and information. We help developing countries and countries in transition modernize and improve agriculture, forestry and fisheries practices and ensure good nutrition for all. Since our founding in 1945, we have focused special attention on developing rural areas, home to 70 percent of the world's poor and hungry people.“

Obr. 4 Záznam tezauru AGROVOC v prezentačním formátu v databance TaxoBank (2015)

**Metadatový záznam** databanky *TaxoBank* je identifikován *Identifikátorem záznamu* (URL). Hodnota je při náhledu záznamu přes internetový prohlížeč zobrazována v adrese. Příklad URL záznamu, který popisuje tezaurus AGROVOC, je zobrazen na obr. 4 (v prvním řádku).

Hlavní blok **popisných informací** zahrnuje jedinečný *Název SOZ*, který i v tomto případě stanovují dle svých interních pravidel zpracovatelé systému. U některých SOZ se registry *TaxoBank* a BARTOC shodují (například „Dewey Decimal Classification“), u jiných SOZ jsou odlišnosti při stanovování unifikovaných názvů veliké (srv. například názvy „NAL Agricultural Thesaurus“ a „United States Department of Agriculture Thesaurus“). Zobrazované záznamy obsahují také několik málo systémových údajů, a to *Datum a čas zpracování záznamu* (například: Wed, 2009-11-11 11:08) a *Zpracovatele záznamu*. Může jít o úplné jméno osoby nebo její šifru (například: Barbara Gilles; aii\_admin aj.). Tyto údaje jsou ještě doplněny stručným *Abstraktem* (viz obr. 4), též někdy ve formě citátu z dokumentace daného SOZ. Abstrakt může zahrnovat různé informace o vzniku SOZ a jeho zakladateli, současném majiteli nebo provozovateli, charakteru SOZ a určení uživatelské komunity, o formách jeho zpřístupňování apod.

Blok **obecných informací** (General information) může zahrnovat větší množství informací. Zahrnuta jsou i v případě tohoto registru všechna známá *Alternativní jména nebo zkratky slovníku* (Vocabulary alternative name or acronym). V záznamu již výše zmíněného Zemědělského tezauru NAL zpracovatelé uvedli hodnoty „NAL Thesaurus“, „Agricultural Thesaurus“ a „NALT“. Podstatným povinným údajem je prvek *Typ slovníku* (Vocabulary type). Hodnoty jsou dány interním číselníkem a jsou uvedeny zkráceným termínem, například: *clsssys* (klasifikační schéma), *thes* (tezaurus), *subjh* (předmětový heslář), *taxon* (taxonomie), *concmp* (pojmová mapa), *contrvoc* (řízený slovník), *gaz* (zeměpisný slovník), *glos* (glosář) a *ont* (ontologie). Dále může být, je-li známo, uvedeno jméno *Autora nebo editora* (Author or Editor) SOZ včetně jeho přesné role, například: „Lori Finch, Thesaurus Coordinator“. Databanka zaznamenává další, pro SOZ příznačné a velmi cenné údaje, pokud jsou známy. Jsou jimi *Aktuální verze/vydání* (Current version/edition) SOZ, například: 2009 Edition, *Datum aktuální verze* (Current version date) SOZ, například: Wed, 2008-12-31, *Frekvence aktualizace* (Update frequency) SOZ, například: Annually, *Dostupné formáty* (Available formats), například: XML, SKOS, PDF, MARC, TXT aj. (i v případě této registrace se vyskytují společně jak konkrétní metadatové formáty, tak označení jejich struktur). Dvě indikativní hodnoty „0“ nebo „1“ jsou zaznamenávány ve specifickém prvku s označením formou otázky *Byl slovník vytvořen jako projekt kursu* (Was vocabulary created as a course project). Je-li k dispozici, pak je v prvku *URL slovníku* (Vocabulary URL) uvedena u volně dostupných SOZ adresa, na které je systém dostupný, například: <http://agclass.nal.usda.gov/>. Zejména v případě komerčně dostupných SOZ je velmi zajímavým prvkem prvek *URL vzorku slovníku* (Vocabulary Sample URL), například vzorek dat z Tezauru CAB (<http://www.cabi.org/cabthesaurus/mtwdk.exe?yi=sample>).

U některých záznamů jsou již k dispozici specifické a velmi užitečné informace, a to v rámci bloku **Rozsah a využití** (Scope and Usage). Jsou jimi *Jazyky* (Languages), ve kterých jsou uvedeny termíny SOZ (hodnoty jsou uvedeny v angličtině, například „English“, „Spanish“ aj.), dále věcné prvky *Hlavní pokrývané předměty* (Major subjects covered) a *Vedlejší pokrývané oblasti* (Minor subjects covered). Hodnoty jsou uvedeny v angličtině a lze předpokládat, že jsou přebírány z vlastního řízeného slovníku producenta (například: Agricultural subjects, biological sciences // Rural and agricultural sociology; physical sciences). Formou kratšího nebo delšího textu může být zaznamenán účel popisovaného SOZ v prvku *Purpose* (Účel) a popřípadě také označení uživatelských komunit, které mohou SOZ využívat v prvku *Využito kým* (Used By). Cennou informací je také vazba na *Příbuzné slovníky* (Related vocabularies), například v záznamu již

zmíněného Zemědělského tezauru NAL je v tomto prvku uvedena informace o existenci souvisejícího *Glosáře NAL* (The NAL Glossary) včetně podrobností o jeho obsahu a počtu zahrnutých termínů. K velmi specifickým informacím, které mohou kvalifikovaně zajišťovat jenom odborníci, patří údaje zaznamenané v prvcích *Překrývání s příbuznými slovníky* (Overlap with related vocabularies) a *Mapování do jiných slovníků* (Mappings to other vocabularies). Příslušné údaje jsou formulovány formou textu.

Rada cenných informací, zejména číselné a kódované povahy, se zaznamenává v prvcích bloku **Charakteristiky slovníku** (Vocabulary characteristics). Je-li informace známa, pak lze v prvku *Popis celkové struktury* (Description of overall structure) uvést formou URL dokument, ve kterém je popsána celková struktura SOZ. Prvek *Typ termínů* (Type of terms) může obsahovat detailní informace o typu termínů obsažených v SOZ, například (hodnoty jsou uvedeny v angličtině): „Agricultural and scientific concepts“, „Concepts in the domain of medicine and related domains, expressed in professional medical and scientific terminology“ aj. Pro odborníky užitečný prvek *Typ zobrazení* (Type of display) může obsahovat podrobné informace o zobrazování zahrnutých termínů. V databance TaxoBank jsou užívány pevně stanovené hodnoty z vlastního číselníku ve formě zkrácených výrazů, například: hier = hierarchický, alph = abecední, perm = permutovaný, html, other = jiný. Podobně, v prvku *Typy vztahu* (Relationship types) mohou být zaznamenány podrobné informace o vztazích mezi termíny (opět formou zkrácených hodnot), například: eq\_pri\_eq = ekvivalence, eq\_lang = jazyková ekvivalence, hier\_bn = rododruhová hierarchie, hier\_inst = třída instance, rel\_t = asociace, othr = ostatní). Jsou-li k dispozici, jsou v prvcích *Počet tříd* (Number of classes) uvedeny v angličtině údaje o počtech tříd daného SOZ (například: 16 top terms in hierarchical display, 109 hierarchies of descriptors aj.), celkový *Počet termínů* (Number of terms), (například: English: 73,194, Spanish: 69,118), *Počet preferovaných termínů* (Number of preferred terms), (například: English: 44,857; Spanish: 44,857), *Počet nepreferovaných termínů* (Number of Non-Preferred terms), (například: English: 28,337; Spanish: 24,261), nebo také *Hloubka hierarchie* (Depth of Hierarchy) s číselným označením úrovní dělení v hierarchii (například: 11 levels). Záznam může obsahovat také prvek *Komentář k charakteristikám* (Characteristics Comments) s poznámkami textového charakteru, například o projektu převodu termínů do struktury propojených otevřených dat.

Různorodé údaje o možnostech využívání systémů organizace znalostí, zejména pak komerční povahy, jsou zaznamenávány v rámci prvků bloku **Termíny a podmínky** (Terms and Conditions). Formou jednoduchých výrazů v angličtině z vlastního číselníku lze v prvku *Dostupnost* (Availability) zaznamenat způsob dostupnosti SOZ, například „free“ aj. Podrobnější poznámky lze zaznamenat formou textu v prvku *Komentář k dostupnosti* (Availability Comments). V rámci prvku *Volby licence* (Licensing Options) je možné uvádět opět formou textu v angličtině podrobnosti k podmínkám získání licence pro využívání SOZ. Doplnkově mohou být v prvku *Instrukce k importu/stahování* (Import/download instructions) podány například adresy URL, ze kterých je stahování dat možné.

Navazující poslední blok **Provozovatel** (Provider) zaznamenává v prvku *Jméno provozovatele slovníku* (Vocabulary provider name) jméno společnosti, která v současnosti spravuje a zpřístupňuje popisovaný SOZ. Jména se vyskytují prozatím v nejednotných formách, někdy je uvedeno úplné jméno včetně zkratky, jindy jenom zkratka, například: „Online Computer Library Center (OCLC)“, „FAO“ apod. V rámci prvku *URL provozovatele* (Provider URL) může být připojena adresa URL dané společnosti a v prvku *Detaily kontaktů na provozovatele* (Provider contact details) jiné užitečné informace v textovém tvaru. Poslední prvek *Sekce komentářů k provozovateli* (Provider section comments) může rovněž formou textu obsahovat další komentáře k činnosti provozovatele.



## 2 Popisná metadata aplikačního profilu AP NKOS

Rozbor současné praxe metadatových popisů SOZ v předcházející části ukázal, že nejvíce prvků zahrnuje popis v databázi TaxoBank – je velmi rozsáhlý a v některých momentech přesahuje i do větších textových popisů. Zařazení některých specifických prvků je inspirativní (například typ vztahů v SOZ, typ zobrazování termínů SOZ nebo mapování SOZ do jiných SOZ). Tato část studie představí jiný metadatový popis, který zahrnuje také řadu specifických popisných prvků, navíc je však vystaven na **konceptuálním modelu FRBR**, tj. nabízí hierarchicky koncipovaný popis SOZ na více úrovních – jednoho **díla**, jeho možná vícečetná **vyjádření** a také **provedení** (publikace). Představení tohoto popisu je uvedeno v této části proto, že jde o významné metadatové schéma připravované mezinárodním týmem expertů a že jde prozatím jenom o návrh.

Metadatový popis SOZ je připravován v rámci komplexního **Aplikačního profilu Dublin Core** pro systémy organizace znalostí (Dublin Core Application Profile, DC-AP NKOS<sup>14</sup>)<sup>15</sup> a také Slovníků NKOS<sup>16</sup>, které od roku 2010 připravuje Pracovní skupina DCMI/NKOS (Dublin Core Metadata Initiative / Networked Knowledge Organization Systems Task Group)<sup>17</sup>. Aplikační profil je postaven v maximální míře na **Metadatových termínech DCMI** (DCMI Metadata Terms, <http://purl.org/dc/terms/>) včetně základního souboru **DCMES** (<http://purl.org/dc/elements/1.1/>). Počítá se i s využitím jmenného prostoru **adms** (<http://www.w3.org/ns/adms#>), **dcat** (<http://www.w3.org/ns/dcat#>), **frbrer** (<http://iflstandards.info/ns/fr/frbr/frbrer/>) a **wdrs** (<http://www.w3.org/2007/05/powder-s#>).

Nově navrhovaný standard specifikoval dále uvedené prvky v rámci tří popisných úrovní. Pro popis systému organizace znalostí na úrovni entity „**Díla**“ je jako podstatný prvek navrhován jeho *Název* <dc:title>. Bližší specifikace prozatím nebyla stanovena, název by ale měl být jedinečný (například unifikovaný název) a měl by být v původním jazyce. Databáze VIAF již obsahuje některé unifikované názvy (v rámci budoucích klastrů) s jedinečným identifikátorem VIAF ID, například „UMLS“ (<http://viaf.org/viaf/176304810>), „Gemeinsame Normdatei“ (<http://viaf.org/viaf/215300162>), „Library of Congress Classification“ (<http://viaf.org/viaf/203733980>), „Library of Congress Subject Headings“ (<http://viaf.org/viaf/211744653>) aj.

Proponovaný prvek *Identifikátor* <dc:identifier> by měl jednoznačným způsobem identifikovat dané dílo. Jako optimální se jeví užití identifikátorů systému VIAF pro dílo (VIAF ID), prozatím však je však na úrovni odborných děl k dispozici těchto identifikátorů málo (VIAF ID viz záznam na obr. 5). Jinou možností by byla identifikace pomocí Mezinárodního identifikátoru textových děl (ISTC)<sup>18</sup>, jak ale ukazuje databáze ISTC, identifikátory textových děl typu SOZ zatím nejsou přiděleny. Prvek *Popis* <dc:description> je připraven pro různorodé typy textových sdělení (abstrakt, obsah, grafická reprezentace nebo libovolný text popisující SOZ jako dílo). Významným formálním prvkem, navrženým Pracovní skupinou DCMI NKOS, je prvek *Typ* <dc:type> SOZ. Hodnoty pro tento prvek jsou již sta-

<sup>14</sup> Jmenný prostor pro NKOS AP bude připraven na adrese: <http://purl.org/nkos/>.

<sup>15</sup> Dublin Core Metadata Initiative. NKOS Task Group. NKOS AP Elements. In: *DCMI NKOS Task Group* [online]. 2013, updates 2014-04-03 final, polished 2014-08-03 [cit. 2015-02-26]. Dostupné z: [http://wiki.dublincore.org/index.php/NKOS\\_AP\\_Elements](http://wiki.dublincore.org/index.php/NKOS_AP_Elements).

<sup>16</sup> Dublin Core Metadata Initiative. NKOS Task Group. NKOS Vocabularies. In: *DCMI NKOS Task Group* [online]. 2013, updated 2013-12-16 [cit. 2015-02-26]. Dostupné z: [http://wiki.dublincore.org/index.php/NKOS\\_Vocabularies](http://wiki.dublincore.org/index.php/NKOS_Vocabularies).

<sup>17</sup> Prezentace skupiny a výsledky činnosti jsou dostupné na jejím webu: <http://nkos.slis.kent.edu/>.

<sup>18</sup> Databázi identifikátorů ISTC vede v databázi Mezinárodní agentura ISTC: <http://www.istc-international.org/>.

noveny v rámci aplikačního profilu ve specifickém strukturovaném slovníku<sup>19</sup>. Prvek *Tvůrce* <dc:creator/dc:creator> je určen pro jméno osoby (fyzické nebo korporativní) primárně odpovědné za vytvoření zdroje (díla). Formální prvek *Práva* <dc:rights> je připraven pro textové informace stanovující různé typy vlastnických práv souvisejících se zdrojem. Věcný popis díla je možné vyjádřit v rámci prvku *Předmět* <dc:subject>. Počítá se s využíváním jak volně tvořených termínů, tak zejména s již celou škálou existujících řízených slovníků (například klasifikačními schémata nebo tezaury), které jsou připraveny k propojování v rámci sémantického webu. U popisu díla se počítá také s prvkem *Vytvoření* <dc:created>, který by mohl být primárně i ve významu data skutečného vytvoření. V případě, že by tato historická informace nebyla k dispozici, lze ji nahradit i datem vyjádření nebo provedení (vydání) díla. Doporučeným způsobem je standardizovaný zápis data podle normy ISO 860, resp. profilu W3CDTF. K formálním prvkům ještě náleží specifický prvek *Uživatelské určení* <dc:audience> (stanovení komunity využívající určitý SOZ) a také zcela nový prvek *Užit kým* <nkos:usedBy>, který bude definován ve jmenném prostoru (<http://purl.org/nkos/>). Je určen pro jméno agenta (programu, aplikace), který využívá popisovaný SOZ. Poslední skupina navržených prvků je věnována **vztahům** SOZ k jiným SOZ (na úrovni díla). Navrženy byly prvky: *Vztah* <dc:relation>, určený k propojení na související SOZ prostřednictvím jednoznačného údaje (optimálně identifikátoru), *Je částí* <dc:isPartOf>, určený k propojení se SOZ, jehož je popisovaný zdroj součástí (opět pomocí jednoznačného údaje), *Je založen na* <nkos:isBasedOn>, určený k propojení na zdroj, který je nějakým způsobem ve vztahu s popisovaným zdrojem. Specifický vztah je reprezentován prvkem *Podporující dokumentace* <wdrs:describedBy>, jehož hodnota může zahrnovat citace zdrojů, které popisovaný zdroj samy popisují, též v předmětném slova smyslu.

## DÍLO

```
<dc:title>Universal Decimal Classification</dc:title>
<dc:identifier>http://viaf.org/viaf/184301709 </dc:identifier>
<dc:description>„Univerzální desetinné třídění“ je jedním z největších
klasifikačních schémat univerzálního charakteru. Jde o schéma
hierarchického a také fasetového typu založeného na desetinném principu.
Hlavní schéma zahrnuje deset základních tříd označených arabskými
číslíci (0 – 9, třída 4 je neobsazena), které pokrývají všechna odvětví
a obory lidského poznání. Pomocné tabulky zahrnují fasety pojmů (místa,
jazyky, data a formy dokumentů). Třídění je využíváno zejména v oblasti
knihovnictví. Je publikováno ve třech základních úrovních: úplné, střední či
standardní a zkrácené. Publikována bývají také specializovaná tematicky
zaměřená vydání. </dc:description>
<dc:type>http://purl.org/nkos/nkostype/classificationScheme </dc:type>
<dc:creator>UDC Consortium</dc:creator>
<dc:rights>Copyright UDC Consortium</dc:rights>
<dc:subject>mezinárodní desetinné třídění</dc:subject>
<dc:created>1902</dc:created>
<nkos:isBasedOn>Dewey Decimal Classification</nkos:isBasedOn>
<nkos:isBasedOn>http://viaf.org/viaf/198122328 </nkos:isBasedOn>
<wdrs:describedBy>McILWAINE, Ia C. Universal decimal classification
(UDC). In: Ed. M. J. BATES a M. N. MAACK. Encyclopedia of library and
information sciences. 3rd ed. Boca Raton (Florida): CRC Press, © 2010,
s. 5432-5439. ISBN 978-0-8493-9712-7 (soubor, Print).
ISBN 978-0-8493-9711-0 (Online) Dostupný také komerčně online z DK
Taylor & Francis (DOI): http://www.tandfonline.com/doi/full/10.1081/
E-ELIS3-120043532. </wdrs:describedBy>
```

<sup>19</sup> Dublin Core Metadata Initiative. NKOS Task Group. NKOS Vocabularies. 2., KOS Types Vocabulary. In: *DCMI NKOS Task Group* [online]. 2013, updated 2013-12-16 [cit. 2015-02-26]. Dostupné z: [http://wiki.dublincore.org/index.php/NKOS\\_Vocabularies#KOS\\_Types\\_Vocabulary](http://wiki.dublincore.org/index.php/NKOS_Vocabularies#KOS_Types_Vocabulary).

## VYJÁDRĚNÍ

```

<dct:title>Universal decimal classification. Online. Česky</dct:title>
<dct:identifier>neuvvedeno</dct:identifier>
<dcat:contactPoint>http://cz.udc-hub.com/cs/contacts.php
</dcat:contactPoint>
<dct:description>Překlad z anglické, standardní verze do české připraven
Národní knihovnou ČR. Více než 70 000 znaků je opatřeno českým
překladem, zbytek textu je uveden prozatím v angličtině. Překlad bude do
systému postupně doplněn. </dct:description>
<dct:creator>UDC Consortium</dct:creator>
<dct:creator>Národní knihovna České republiky</dct:creator>
<dct:language>http://id.loc.gov/vocabulary/iso639-2/cze </language>
<dct:language>http://id.loc.gov/vocabulary/iso639-2/eng </language>
<nkos:sizeNote>70 626 platných a 11 000 zrušených notací</sizeNote>
<dct:created>2015-01-05</dct:created>
<dct:modified>2015-02-05</dct:modified>
<dct:rights>Copyright UDC Consortium</dct:rights>
<dct:rights>Copyright Národní knihovna České republiky </dct:rights>
<nkos:updateFrequency scheme="http://dublincore.org/groups/collections/
frequency/" >freq:annual</nkos:updateFrequency>
<dct:audience>Určeno knihovníkům, badatelům a studentům v oborech
knihovnictví a informatika; je volně přístupná v souladu se Všeobecnými
smluvními podmínkami.</dct:audience>
<dct:relation>http://www.udc-hub.com/en/login.php<dct:relation>
<dct:relation>http://nl.udc-hub.com/nl/login.php<dct:relation>
<frbrer:isRealizationOf>http://viaf.org/viaf/184301709 </frbrer:isRealizationOf>

```

## PROVEDENÍ

```

<dct:title>České MDT Online</dct:title>
<dct:identifier>http://cz.udc-hub.com/cs/login.php </dct:identifier>
<dcat:contactPoint>http://cz.udc-hub.com/cs/contacts.php
</dcat:contactPoint>
<dct:description>Nové online zpřístupnění české verze „UDC Online“. Jde o
standardní vydání MDT. Obsahuje úplnou verzi MDT, která zahrnuje 70 626
platných a 11 000 zrušených notací MDT. Online nástroj na serveru Konsorcia
UDC má několik funkcí umožňujících vyhledávání, prohlížení, analýzu,
validaci a tvorbu notací MDT. </dct:description>
<dct:creator>UDC Consortium</dct:creator>
<dct:creator>Národní knihovna České republiky</dct:creator>
<dct:publisher>UDC Consortium</dct:publisher>
<dct:format>online</dct:format>
<dct:format scheme="IMT">text/html</dct:format>
<dct:issued>2015-02-05</dct:issued>
<dct:relation>http://www.udc-hub.com/en/login.php<dct:relation>
<dct:relation>http://nl.udc-hub.com/nl/login.php<dct:relation>
<nkos:serviceOffered>Online vyhledávání a tvorba a validace notací</
nkos:serviceOffered>
<frbrer:isEmbodimentOf>Universal decimal classification. Online. Česky
</frbrer:isRealizationOf>
<dct:isPartOf>http://www.udc-hub.com/ </dct:isPartOf>
<adms:sample>http://cz.udc-hub.com/cs/demo.php </adms:sample>

```

Obr. 5 Trojúrovňový záznam nejnovějšího českého online vydání MDT v navrhovaném metadatovém schématu AP-NKOS (hypotetický záznam)

Prvek *Název*<dct:title> je navržen také na úrovni entity „**Vyjádření**“. I jeho bližší specifikace prozatím není stanovena, v krátké definici je ale zmiňována možnost doplňkového údaje uvedeného v rámci názvu, zejména pak údaje o vydání nebo jazyce SOZ.

Lze předpokládat, že by název vyjádření měl být také jedinečný. V databázi VIAF se i v tomto případě již objevila řada takových formalizovaných názvů, zatím z iniciativy Kongresové knihovny a Německé národní knihovny, například „Universal decimal classification. Selections. Czech“ (<http://viaf.org/viaf/186352368>) nebo „EUROVOC“ (<http://viaf.org/viaf/215952972>). Ukázka záznamu na obr. 5 nese unifikovaný název pro MDT v online české verzi, který je hypoteticky připraven podle příkladů názvů vyjádření z báze VIAF. Ve stejném příkladu je v prvku *Identifikátor* <dct:identifier> prozatím hodnota „neuvedeno“. Lze očekávat, že se časem tato hodnota v systému VIAF objeví. Formální prvek *Kontakt* <dcat:contactPoint> je v návrhu jištěn doménou **DCAT**. Hodnotou prvku může být odkaz na místo, kde mohou být poskytnuty další informace o vyjádření díla (o překladu, formě vydání apod.). I na úrovni vyjádření je k dispozici prvek *Popis* <dct:description>. Je připraven pro textová sdělení (abstrakt, obsah aj.), popisující proces vyjádření díla (překlada, vydání aj.). Prvek *Tvůrce* <dct:creator/dc:creator> je i v rovině vyjádření přítomen a měl by zahrnovat jména osob (fyzické nebo korporativní) odpovědné za vyjádření díla (překlada, realizátor online verze apod.). Podstatným formálním prvkem úrovně vyjádření u textových děl, a tudíž i SOZ, je prvek *Jazyk* <dct:language>. Jde o jazyk textu a doporučeno je využít číselníků jazyků. Současná praxe řady systémů využívá volně vystavené slovníky jazyků na serveru Kongresové knihovny. Záznam na obr. 5 zahrnuje v hodnotě prvku URI jazyka (trojznakový kód) z citovaného slovníku. Ryze specifickým a cenným údajem v záznamech SOZ je jeho velikost (počtu termínů). Navrhovaná specifikace AP NKOS proto zařadila prvek *Poznámka o velikosti* <nkos:sizeNote>. U popisu vyjádření se počítá samozřejmě s prvkem *Vytvořen* <dct:created> – ponese, bude-li známo, datum dokončení realizace vyjádření (překlada aj.). Nově je ale zařazen prvek *Modifikován* <dct:modified>. Doporučeným způsobem je u obou datací standardizovaný zápis data podle normy ISO 8601, resp. profilu W3CDTF. Zařazen je i na této úrovni popisu prvek *Práva* <dct:rights> pro textové informace stanovující různé typy vlastnických práv souvisejících s vyjádřením díla (práva na překlad apod.). Úroveň vyjádření může nést rovněž formální prvek *Uživatelské určení* <dct:audience> (stanovení komunity, která může využívat například překlad SOZ – viz obr. 5) a také prvek *Užitým* <nkos:usedBy> (viz informace výše již uvedené). Nově je definován na této úrovni (a opět bylo nutno použít vlastní jmenný prostor nkos:) specifický, cenný prvek *Frekvence aktualizace SOZ* <nkos:updateFrequency>. Hodnoty je doporučeno přebírat ze Slovníku frekvencí DCMI (<http://dublincore.org/groups/collections/frequency/>). Skupina navržených prvků pro **vztahy** SOZ k jiným SOZ (na úrovni vyjádření) zahrnuje většinu stejných prvků jako úroveň díla: prvek *Vztah* <dct:relation> je určený k propojení na související vyjádření SOZ prostřednictvím jednoznačného údaje (na obr. 5 jsou vazby na jiné jazykové verze online vydání dané klasifikace), nový prvek ze jmenného prostoru frbr: *Je realizací* <frbr:isRealizationOf>, jenž je určen k jednoznačnému odkazu na realizované dílo (optimálně URI díla), dále prvek *Je částí* <dct:isPartOf>, určený k propojení s vyjádřením SOZ, jehož je popisovaný zdroj součástí. Specifický vztah je opět reprezentován prvkem *Podporující dokumentace* <wdrs:describedBy>, jehož hodnota může zahrnovat citace zdrojů, které popisované vyjádření popisují, též v předmětném slova smyslu. Novým formálním prvkem (z domény adms:) je prvek *Vzorek* (adms:sample), který by měl obsahovat odkaz na web s demonstračními verzemi SOZ, s příklady záznamů těchto SOZ, zejména v případě, že jde o komerčně dostupný zdroj, anebo zdroj dostupný po registraci (viz obr. č. 5).

Poslední úroveň popisu SOZ „**Provedení**“, tedy konkrétních produktů, zajišťuje poslední skupina prvků. Prvek *Název* <dct:title> by měl obsahovat deskriptivní název publikovaného SOZ. Ukázka záznamu na obr. 5 nese název online produktu v češtině. V případě existence více názvů (alternativních názvů) navrhovaná specifikace nevyužila prvku Dublin Core „dct:alternative“ a prozatím není jasně specifikováno, kam by se

měly alternativní názvy zapisovat. Prvek *Identifikátor* <dct:identifier> může nést známé identifikátory tištěných i elektronických produktů (ISBN, ISSN, DOI, URI, URN aj.). Formální prvek *Kontakt* <dcat:contactPoint> je definován stejným způsobem jako u úrovně vyjádření. K dispozici je také prvek *Popis* <dct:description> pro potřebu textového zápisu s dalšími podrobnostmi o charakteru daného produktu. Prvek *Tvůrce* <dct:creator/dc:creator> je přítomen i u této úrovně popisu se stejným významem. Nově zařazeným prvkem je důležitý prvek *Vydavatel* <dct:publisher/dc:publisher>, který je určen pro jméno korporativní nebo fyzické osoby, která SOZ nebo jeho konkrétní verzi/vydání zveřejňuje. Specifickým údajem pro SOZ je údaj o formátu a fyzickém nosiči. Navržen byl proto prvek *Formát* <dct:format>, u online dostupných produktů je doporučeno užívat hodnot ze slovníku MIME (Internet Media Types, IMT, <http://www.iana.org/assignments/media-types/media-types.xhtml>). U publikací je definován nový prvek *Datum (vydání)* <dct:issued>. Zařazen je i na této úrovni popisu prvek *Práva* <dct:rights> týkající se práv ke zveřejněným SOZ. Novým prvkem je prvek z vlastního jmenného prostoru *Nabízené služby* <nkos:serviceOffered> s uvedením služeb, jichž mohou uživatelé využít (stahování, anotování, dotazování aj.). Skupina prvků pro **vztahy** SOZ k jiným SOZ (na úrovni provedení) zahrnuje také většinu stejných prvků jako úroveň díla a vyjádření: prvek *Vztah* <dct:relation> je určen k propojení na související publikace SOZ, prvek *Je částí* <dct:isPartOf>, určený k propojení s publikací SOZ, jehož je popisovaný zdroj součástí. Novým prvkem opět ze jmenného prostoru frbr: je prvek *Je ztělesněním* <frbr:isEmbodimentOf>, jenž je určen k jednoznačnému odkazu na vyjádření díla (optimálně URI vyjádření). Specifický vztah je reprezentován prvkem *Podporující dokumentace* <wdrs:describedBy> a prvkem *Vzorek* (adms:sample).

### 3 Návrh metadatového popisu SOZ ve znalostní bázi organizace znalostí

#### 3.1 Schéma popisu ve znalostní bázi

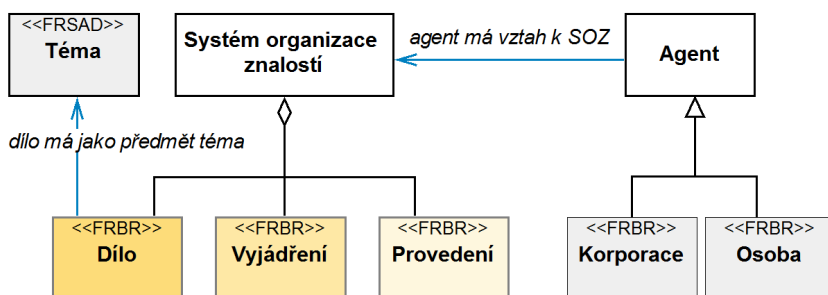
Cílem projektu NAKI je shromáždit a systemizovat současné poznatky oboru organizace znalostí ve formě znalostní báze, která umožní ukládání, prohlížení, vyhledávání stávajících a odvození nových znalostí. Znalostní báze bude zpřístupněna ve formátu propojených otevřených dat a bude sloužit pro další výzkum a jako nástroj výuky a vzdělávání především v oboru informační vědy a knihovnictví. Ontologickou strukturu znalostní báze představují dva typy znalostních jednotek: 1) výroky („čisté znalosti“ – věty formalizované jako logické predikáty, instanciované formou textových dat) a 2) popisná metadata o relevantních dokumentových i nedokumentových zdrojích (osoby, instituce, události, nástroje, aktivity a procesy). Úvodní fáze akvizice znalostí poskytla materiál pro prototyp znalostní báze s cca 2 300 jednotkami deklarativních znalostí: 150 výroků ve formátu RDF, výkladový slovník o rozsahu 900 termínů, 1 000 záznamů odborné literatury, 230 záznamů systémů organizace znalostí. Popis SOZ tvoří významnou množinu znalostních entit v druhém modulu znalostní báze, představovaném popisnými metadaty.

Pro návrh popisných metadat SOZ byla zvolena rámcová metodika, doporučená Pravidly pro aplikační profily Dublin Core<sup>20</sup>. V souladu s ní byly nejprve vymezeny

<sup>20</sup> COYLE, Karen, BAKER, Thomas. *Guidelines for Dublin Core Application Profiles* [online]. Dublin (Ohio): DCMI, 2009-05-18 [cit. 2015-02-26]. Dostupné z: <http://dublincore.org/documents/profile-guidelines/>.

**funkční požadavky na metadata** ve znalostní bázi, jejichž východiskem se staly obecné požadavky na znalostní bázi jako celek: 1) umožnit shromáždění a systemizaci aktuálních poznatků o SOZ s ohledem na lokální kulturní a jazyková specifika českého prostředí, 2) umožnit zpřístupnění stávajících a odvozování nových znalostí ve formátu propojených otevřených dat, 3) poskytnout zdrojový materiál pro zpracování původní české monografie, 4) poskytnout zdroje pro aktualizaci české terminologie, 5) umožnit využití při výuce a vzdělávání v oboru informační vědy a knihovnictví. Z těchto požadavků vyplynulo, že metadata ve znalostní bázi mají obsahovat co nejúplnější popis včetně zachycení historie a vzájemných vztahů SOZ.

Druhým krokem byla tvorba **doménového modelu**, který má za úkol vymezit entity (objekty, věci), jež mají metadata popisovat. Jeho schematické znázornění je uvedeno na obr. 6. Klíčovou entitou je samozřejmě **Systém organizace znalostí**, další důležitou entitou je **Agent**, který v našem pojetí zastupuje osobu či instituci, jež se SOZ souvisí, ať už jako tvůrce, provozovatel nebo jako uživatel. V návrhu znalostní báze je uplatňována zásada maximálního opětovného využití již ověřených konceptů a modelů. Již na úrovni doménového modelu jsou proto využity a uživatelsky přizpůsobeny entity modelů FRBR a FRSAD *Dílo*, *Vyjádření*, *Provedení*, *Osoba*, *Korporace* a *Thema*<sup>21</sup> a jejich vztahů. Trojice entit *Dílo*, *Vyjádření* a *Provedení* představuje vzájemně propojený a logicky strukturovaný celek údajů o SOZ. Poslední významnou komponentou doménového modelu je entita **Téma**, jež je přímou implementací entity *Thema* z modelu FRSAD a zahrnuje údaje o věcném obsahu SOZ.



Obr. 6 Doménový model, založený na modelu FRBR a FRSAD

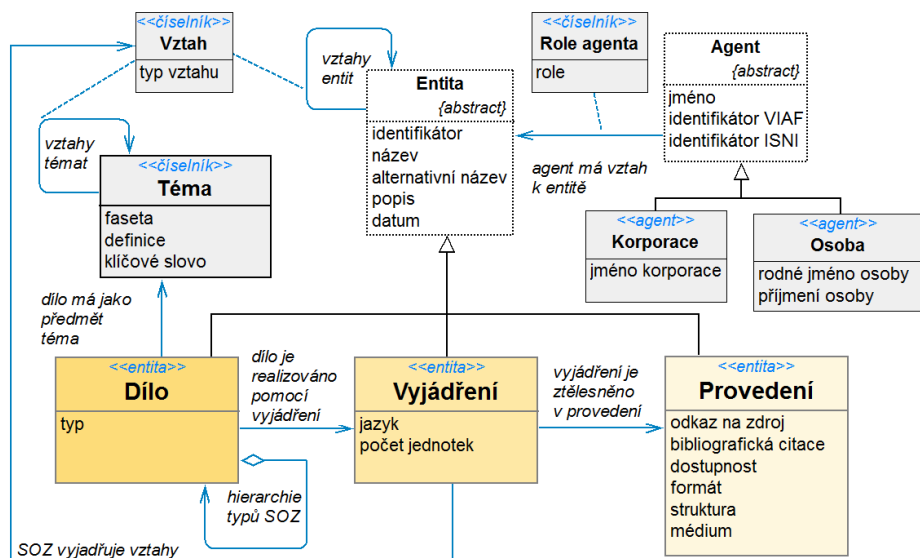
Struktura metadataového popisu na obr. 6 a 7 je znázorněna pomocí diagramu tříd v jazyce UML (Unified Modeling Language), používaného v souladu s ISO 24156, jež upravuje použití notace UML v terminologické práci<sup>22</sup>. Datové prvky jsou seskupeny do tříd, propojených sémantickými vztahy. **Asociace** znázorňuje libovolný sémantický vztah prostřednictvím čáry zakončené u asymetrické asociace šipkou. **Agregace** znázorňuje partitivní hierarchický vztah („je součástí“) šipkou zakončenou kosočtvercem. **Generalizace** zobrazuje rododruhovou neboli generickou hierarchii s dědičností („je typ“) pomocí šipky zakončené trojúhelníkem, který míří z podřazené třídy („podtyp“) na nadřazenou třídu („nadtyp“).

<sup>21</sup> Doslovné převzetí entity *Thema* z modelu FRSAD, který ještě nebyl přeložen do češtiny.

<sup>22</sup> ISO 24156-1:2014. *Graphic notations for concept modelling in terminology work and its relationship with UML -- Part 1: Guidelines for using UML notation in terminology work*. 1st ed. Geneva: International Organization for Standardization, 2014. 24 s.



Po zkonstruování doménového modelu následovala etapa definování vlastností pro příslušné objekty a specifikace jejich vztahů, jejímž výsledkem je **model struktury** metadatového popisu, graficky znázorněný na obr. 7. Opět bylo naší snahou využít již definovaných prvků popisu a standardizovaných slovníků, přičemž hlavním zdrojem se stal aplikační profil AP NKOS, popsany v předchozí části. Z jeho 21 prvků je ve znalostní bázi přímo aplikováno 11 prvků, 6 prvků je namapováno s různým stupněm ekvivalence a 4 prvky zatím aplikovány nejsou (*frekvence aktualizace, práva, uživatelské určení, nabízené služby*). Kromě prvků popisu AP NKOS jsou využity i jeho dva slovníky hodnot (NKOS Vocabularies<sup>23</sup>): *Slovník pro vztahy SOZ* (KOS Relation-Type Vocabulary) byl převzat v úplnosti a *Slovník pro typy SOZ* (KOS Type Vocabulary<sup>24</sup>) je mapován do vlastní typologie systémů organizace znalostí<sup>25</sup>. Naše vlastní typologie je doplněna klasifikací typů SOZ, převzatou z *Klasifikačního systému literatury oboru organizace znalostí*<sup>26</sup>, konkrétně z části 0 – Form Divisions, třídy 01, 03 a 04. Další slovníky hodnot byly přejaty pro prvek *jazyk* (v současné době je aplikován třípísmenný kód a jména podle ISO 639-2), *formát* (jsou používány kódy formátů podle standardu MIME), *role* (využit je slovník Kongresové knihovny<sup>27</sup>). Identifikátory jsou přejímány z následujících jmenných



Obr. 7 Struktura metadatového popisu SOZ ve znalostní bázi

<sup>23</sup> Dublin Core Metadata Initiative. NKOS Task Group. NKOS Vocabularies, Ref. č. 15, [http://wiki.dublincore.org/index.php/NKOS\\_Vocabularies](http://wiki.dublincore.org/index.php/NKOS_Vocabularies).

<sup>24</sup> Dublin Core Metadata Initiative. NKOS Task Group. NKOS Vocabularies. 2., KOS Types Vocabulary, Ref. č. 18, [http://wiki.dublincore.org/index.php/NKOS\\_Vocabularies#KOS\\_Types\\_Vocabulary](http://wiki.dublincore.org/index.php/NKOS_Vocabularies#KOS_Types_Vocabulary).

<sup>25</sup> BRATKOVÁ, Ref. č. 1, s. 22. -- Na rozdíl od typologie AP NKOS, která je lineárním seznamem, uplatňuje naše typologie jednostupňovou hierarchii.

<sup>26</sup> Classification System for Knowledge Organization Literature [online]. ISKO 2011-2012, last update 2012-04-13. Dostupné z: <http://www.isko.org/scheme.php>.

<sup>27</sup> Library of Congress. MARC Code List for Relators [online]. Washington (D.C.): The Library of Congress, modified 2011-04-26 [cit. 2015-02-26]. Library of Congress Linked Data Service. Authorities and Vocabularies. Dostupné z (URI): <http://id.loc.gov/vocabulary/relators>.

prostorů: VIAF ID pro *osoby*, *korporace* a zatím ojedinele pro *díla* a *vyjádření*, ISNI pro *osoby* a *korporace*, ISBN, ISSN, DOI, URL, URN a URI pro *provedení*.

V modelu struktury metadatového popisu na obr. 7 se vyskytují 3 typy tříd: «entita», «agent» a «číselník». Třídy typu «entita» odpovídají entitám z 1. skupiny modelu FRBR – *Dílo*, *Vyjádření* a *Provedení*. Sémantika jejich vztahů je rovněž převzata z modelu FRBR – *dílo je realizováno* pomocí *vyjádření* a *vyjádření je ztělesněno* v *provedení*. Třídy typu «agent» odpovídají entitám z 2. skupiny modelu FRBR, odpovědným za obsah, výrobu, rozšiřování nebo správu entit 1. skupiny. Třídy typu «číselník» jsou autoritními seznamy povolených hodnot, jež se vztahují k vlastnostem tříd typu «entita» a «agent». Významnou roli mezi číselníky hraje třída *Téma*, která je ve smyslu modelu FRBRAD zobecněním entit 3. skupiny z modelu FRBR, týkajících se věcného obsahu díla. Mezi jednotlivými tématy mohou být definovány vztahy (typicky hierarchie nebo asociace), což v modelu znázorňuje rekurzivní asociace *vztahy témat*. Asociativní třídy *Vztah* a *Role agenta* zachycují situaci, kdy vlastnosti *typ vztahu* a *role* nejsou vlastnostmi jednotlivých tříd, ale vlastnostmi vztahů mezi nimi.

Kromě tříd reprezentujících datové prvky se v modelu vyskytují ještě dvě tzv. abstraktní třídy – *Entita* a *Agent*. Obě jsou rodičovskými třídami v generické hierarchii, jež nemají vlastní instance, ale veškeré jejich vlastnosti a vztahy dědí hierarchicky podřazené dětské třídy *Dílo*, *Vyjádření* a *Provedení*, *Osoba* a *Korporace*. Konstrukce generalizace v UML umožňuje na jednom místě (v rodičovské třídě) definovat vlastnosti a vztahy, jež se ve stejné nebo podobné míře vyskytují u více objektů, a tak omezit redundanci v modelu. Tedy například *identifikátor*, resp. tvrzení „má identifikátor“ z třídy *Entita* se v implementační rovině týká jak *díla*, tak *vyjádření* a *provedení*, přičemž v každé z těchto tříd má věcně specifickou sémantiku. Některé zděděné vlastnosti mohou být v dětských třídách modifikovány. Vlastnost *popis* je v dětských třídách specifikována jak svým označením, tak svým obsahem – jsou stanoveny specifické požadavky pro popis SOZ na úrovni *díla* (*popis díla*), na úrovni *vyjádření* (*popis vyjádření*) a na úrovni *provedení* (*popis provedení*). Pro abstraktní třídu *Agent* platí totéž. Veškeré její vlastnosti a vztahy dědí třídy *Osoba* a *Korporace*, v nichž dochází ke specifikaci vlastnosti *jméno* na *jméno korporace*, *rodné jméno osoby* a *příjmení osoby*.

Jak *dílo*, tak *vyjádření* a *provedení* mohou být spojeny vzájemnými vztahy určitého typu (např. *Dílo A je založeno na Díle B*, *Vyjádření A je mapováno do Vyjádření B*, *Dílo A popisuje Dílo B*). Tuto skutečnost zachycuje rekurzivní asociativní vztah *vztahy entit*.

Již bylo uvedeno, že dědičnost se týká nejen atributů rodičovských tříd, ale i jejich vztahů. Konkrétně vztah *agent má vztah k entitě* se může v implementační rovině uplatnit jako vztah *osoby* a *díla*, vztah *korporace* a *provedení* apod., přičemž konkrétní role daného agenta v daném vztahu je určena hodnotou prvku *role* z číselníku rolí (např. autor, vydavatel, editor, poskytovatel).

### 3.2 Definice a mapování datových prvků pro popis SOZ ve znalostní bázi

V následujícím přehledu jsou uvedeny prvky popisu, odvozené z atributů tříd, definovaných v modelu struktury (viz obr. 7). Struktura metadatového popisu na obr. 7 je v zájmu přehlednosti mírně zjednodušena. Pro účely zápisu údajů do znalostní báze jsou některé datové prvky ze strukturního modelu ještě vnitřně strukturovány: *jazyk* (kód, jméno), *dostupnost* (popis, typ), *typ* (vlastní číselník, Klasifikační systém literatury oboru organizace znalostí), *typ vztahu* (vztahy entit, vztahy pojmů/termínů). Každý prvek popisu je doplněn definicí svého významu a příklady možných hodnot.

## ENTITA

Abstraktní třída, nadtyp pro třídy *Dílo*, *Vyjádření*, *Provedení*. Zobecňuje vlastnosti společné pro všechny dětské třídy.

### identifikátor

Nadtyp pro identifikátor díla/vyjádření/provedení. Každá entita má identifikátor.  
*Příklady: URI; URN; ARK; VIAF ID; ISBN; ISSN; DOI; URL*

### název

Nadtyp pro název díla/vyjádření/provedení. Každá entita má 1 název.  
*Příklady: Dewey decimal classification; EUROVOC*

### alternativní název

Nadtyp pro alternativní název díla/vyjádření/provedení. Každá entita může mít více alternativních názvů (především zkratky, případně názvy v jiných / rovnocenných jazycích, nebo český překlad).  
*Příklady: DDC; Deweyho desetinné třídění*

### popis

Nadtyp pro popis díla/vyjádření/provedení. Každá entita může mít popis.  
*Příklady: abstrakt; anotace; strukturovaný obsah*

### datum

Nadtyp pro datum vzniku díla/vyjádření nebo publikování provedení.  
*Příklady: 2013-01-12; 1876-00-00; 2015*

## VZTAH

Třída slouží pro specifikaci vzájemných vztahů entit (vztah *Entita – Entita*, *Dílo – Dílo*, *Vyjádření – Vyjádření*, *Provedení – Provedení*), sémantických vztahů pojmů (vztah *Téma – Téma*) a typů vztahů vyjadřovaných v SOZ (vztah *Vyjádření – Vztah*).

### typ vztahu

Hodnota z vlastního číselníku vztahů. Číselník zahrnuje typy vzájemného vztahu entit a sémantických a syntaktických vztahů mezi jednotkami, jež daný SOZ umožňuje reprezentovat (vzájemné vztahy SOZ jako celků i jejich strukturních částí).  
*Příklady: vztahy pojmů/termínů: generická hierarchie; asociace; vztahy entit: sumarizace; adaptace; doplněk; založeno na; celek-část*

## TÉMA

### klíčové slovo

Klíčová slova z vlastního řízeného výkladového slovníku vymezující doménu / obor / téma, jež SOZ pokrývá, případně doménu (obsah, obor, tematiku) organizovaných jednotek.  
*Příklady: organizace znalostí (vědní obor); zemědělství; lékařství*

### definice

Slovní vyjádření obsahu a rozsahu pojmu.  
*Příklad: Obor organizace znalostí zkoumá proces organizace znalostí a jeho kontext, tj. zdroje, jež jsou v procesu organizace transformovány, používané metody a nástroje a produkty, jež jsou tímto procesem vytvářeny, včetně zúčastněných aktérů – osob, institucí, technologií*

### faseta

Faseta používaná k organizaci výkladového slovníku.  
*Příklad: proces*

## DÍLO

### popis díla

Podtyp popisu entity. Slovní charakteristika obsahu díla v přirozeném jazyce. Zahrnuje popis pojmové základny SOZ, historii, údaje o způsobu aktualizace, informace o tvůrcích, účel, uživatelské určení.

### typ

Hodnota z vlastního číselníku typů systémů organizace znalostí.

*Příklad: ontologie; tezaurus*

### typ (KOL)

Znak Klasifikačního systému literatury oboru organizace znalostí (KOL). V případě popisu SOZ se používají znaky z části 0 – Form Divisions, třídy 01, 03 a 04.

*Příklad: 042 Universal Decimal Classification*

## VYJÁDŘENÍ

### popis vyjádření

Podtyp popisu entity. Slovní charakteristika obsahu vyjádření v přirozeném jazyce. Zahrnuje informace o vydání/verzi, o struktuře celého systému, o struktuře hesel/záznamů (deskriptorů, třídniců), o jazykové formě termínů, popisuje systém notace, možnosti zobrazování, prohlížení a vyhledávání.

### počet jednotek

Počet jednotek (lexémů) SOZ.

*Příklady: 15 000 znaků; 20 953 deskriptorů*

### jazyk

Označení jazyka podle ISO 639-2.

### jméno jazyka

Jméno jazyka podle ISO 639-2 (česká verze).

*Příklady: němčina; arabština; nizozemština*

### kód jazyka

Kód jazyka podle ISO 639-2.

*Příklady: ger; eng*

## PROVEDENÍ

### bibliografická citace

Bibliografická citace provedení podle ISO 690.

*Příklad: DEWEY, Melvil, devised. Abridged Dewey decimal classification and relative index. Ed. 15. Ed. by Joan S. MITCHELL, Editor in Chief, Julianne BEALL, Rebecca GREEN, Giles MARTIN, Michael PANZER, Assistant Editors. Dublin (Ohio): OCLC, 2012. lxxvii, 1228 s. ISBN 978-0-910608-81-7. ISBN 0-910608-81-4.*

### popis provedení

Podtyp popisu entity. Slovní charakteristika obsahu provedení v přirozeném jazyce. Zahrnuje informace o formátu daného provedení, způsobu zobrazení a případné další fyzické charakteristiky zdroje (např. způsob přístupu ke zdroji).

### odkaz na zdroj

Odkaz na místo, z něž je popisovaný SOZ dostupný (např. URI nebo URL).

*Příklad: <http://id.loc.gov/authorities/classification>*

### dostupnost

Charakteristika přístupu ke zdroji.

### dostupnost-popis

Slovní charakteristika přístupu ke zdroji (volný text).

*Příklady: Ve: NK ČR (ABA001) – prezenčně; NK ČR Knih. Inst. (ABA003) -- sign. Od 20.817, Od 20.818, Od 549/B1, Od 549/B2; Knihovna Jinočnice (ABD(107).*

#### **dostupnost-tp**

Hodnota z vlastního číselníku typů dostupnosti.

*Příklad: open access – volná díla*

#### **formát**

Hodnota z vlastního číselníku formátů (podmnožina slovníku MIME (IMT)).

*Příklady: application/rdf+xml; text/html*

#### **médium**

Hodnota z vlastního číselníku médií.

*Příklady: CD ROM; online*

#### **struktura**

Hodnota z vlastního číselníku struktur.

*Příklad: SKOS*

### **AGENT**

Abstraktní třída, nadtyp pro třídy *Osoba* a *Korporace*. Třída zobecňující entity 2. skupiny z modelu FRBR (*Osoba, Korporace*), odpovědné za intelektuální nebo umělecký obsah, fyzickou výrobu a rozšiřování nebo správu entit v první skupině.

#### **jméno**

Nadtyp pro rodné jméno osoby, příjmení osoby, jméno korporace.

#### **identifikátor VIAF**

Trvalý identifikátor (ID VIAF) agenta v databázi VIAF (Virtual International Authority File). Zdroj: <http://viaf.org>.

#### **identifikátor ISNI**

Trvalý identifikátor (ISNI) agenta v databázi ISNI (International Standard Name Identifier). Zdroj: <http://isni.org>.

### **OSOBA**

#### **rodné jméno osoby**

Podtyp jména agenta.

#### **příjmení osoby**

Podtyp jména agenta.

### **KORPORACE**

#### **jméno korporace**

Podtyp jména agenta.

### **ROLE AGENTA**

#### **role**

Role agenta ve vztahu k entitě. Hodnota z vlastního číselníku rolí. Mapováno do slovníku MARC Code List for Relators (<http://id.loc.gov/vocabulary/relators>)

*Příklad: autor (díla); překladatel (vyjádření); vydavatel (provedení)*

Zatím poslední realizovanou etapou návrhu prvků metadatového popisu SOZ je jejich **sémantické mapování** do věcně relevantních jmenných prostorů. Z výše uvedeného funkčního požadavku na zpřístupnění obsahu znalostní báze ve formátu propojených otevřených dat vyplynula nutnost navrhnout prvky popisu SOZ tak, aby pokud možno

splňovaly parametry tzv. pětihvězdičkového schématu<sup>28</sup> otevřenosti dat, který stanoví, že data mají být pomocí odkazů propojena na jiná související data. Pro tento účel bylo zvoleno osm jmenných prostorů – v předchozích částech studie popsaná schémata registrů BARTOC a TaxoBank, schémata AP NKOS a Schema.org. Kromě těchto systémů byly do mapování zařazeny ještě obecně použitelné Metadatové termíny DCMI (DC-terms) a modely FRBR a FRSAD a specializovaná schémata DCAT a OMV.

**Data Catalog Vocabulary (DCAT)**<sup>29</sup> je doporučení Konsorcia W3C, schválené v lednu 2014. Obsahuje schéma a slovník pro metadatový popis v katalozích souborů dat (*dataset*). Příkladem je Portál otevřených dat EU (<http://open-data.europa.eu/en/data/>). Pro účely propojení dat ze znalostní báze je schéma DCAT relevantní z toho důvodu, že současné SOZ jako propojená data mají z hlediska formátu charakter souboru dat. Slovník má výrazně eklektickou povahu, přináší pouze 7 vlastních prvků popisu (*stránka, z níž je soubor dat dostupný, kontakt, URL pro přístup, URL pro stahování, velikost v bytech, typ média, klíčové slovo*). I když jeho tvůrci DCAT neoznačují jako aplikační profil Dublin Core, značnou měrou prvky DC přepoužívají, přičemž obecnou definici obvykle doplní specifickými poznámkami pro účely popisu souborů dat. Schéma DCAT lze považovat za určitou analogii AP NKOS. Zatímco primárním účelem AP NKOS je popis SOZ v jejich online registrech, DCAT je určen pro popis souborů dat v jejich online katalozích.

**OMV (Ontology Metadata Vocabulary)**<sup>30</sup> je výsledkem mezinárodního projektu, realizovaného v letech 2004–2009 Konsorciem OMV, jež tvoří: Univerzita Brémy a její Centrum počítačových technologií, Madridská polytechnická univerzita, Univerzita Karlsruhe a Stanfordské centrum biomedicínského informatického výzkumu. Jeho cílem je tvorba formálního metaontologického rámce pro popis aplikovaných ontologií z hlediska reprezentace znalostí. Výstupem projektu je veřejně zpřístupněný slovník metadat pro jednotný popis ontologií (*OMV Core Metadata Entities* a *OMV Extensions*). Obsahuje seznam atributů popisu (např. *název, popis, datum vytvoření*), seskupených do 13 tříd. Klíčovou třídou je *ontologie*, dalšími třídami jsou *osoby* a *organizace*, jež ji vyvíjejí, a třídy blíže charakterizující danou ontologii: *licenční model, paradigma reprezentace znalostí, typ ontologie, stupeň formalizace, řešená ontologická úloha, doména, použitá metodika, použitý nástroj, syntax, ontologický jazyk*. Pro některé atributy (např. *typ ontologie*) jsou k dispozici seznamy hodnot. Jak název napovídá, OMV je metadatové schéma specializované na ontologie, jež jsou v současné době považovány za nejprogresivnější typ SOZ. Jeho přínosem je definování prvků popisu, specifických pro ontologie, jež nelze najít v jiných, obecněji zaměřených schématech (např. *stupeň formalizace, uplatněné paradigma reprezentace znalostí, použitá metodika ontologického inženýrství, používaný ontologický jazyk*). Zajímavým aspektem OMV je také to, že soustředí pozornost především na konceptuální bázi ontologie a nezaměřuje se na popis ontologie ve formě dokumentu, naplněného instancemi.

<sup>28</sup> BERNERS-LEE, Tim. *Linked data – design issues* [online]. 2006-07-27, last change 2009-06-18 [cit. 2015-04-10]. Dostupné z: <http://www.w3.org/DesignIssues/LinkedData.html>.

<sup>29</sup> World Wide Web Consortium. *Data catalog vocabulary (DCAT)* [online]. W3C Recommendation 16 January 2014. Ed. Fadi MAALI, John ERICKSON. 2014-01-16 [cit. 2015-02-28]. Dostupné z: <http://www.w3.org/TR/vocab-dcat/>.

<sup>30</sup> PALMA, Raúl, HARTMANN, Jens, HAASE, Peter. *OMV: Ontology Metadata Vocabulary for the Semantic web* [online]. Version 2.4.1. S.l.: OMV Consortium, March 2009 [cit. 2015-02-28]. 76 s. OMV Report. Dostupné z: <http://freefr.di.sourceforge.net/project/omv2/OMV%20Documentation/OMV-Reportv2.4.1.pdf>.



Výsledky mapování prvků popisu SOZ obsahuje tabulka na obr. 8. Lze konstatovat, že každý z prvků popisu se podařilo namapovat alespoň do jednoho relevantního jmenného prostoru. 7 prvků, které tvoří jádro obecných identifikačně-popisných údajů, bylo možné namapovat do všech zvolených systémů. Kromě ověření sémantické propojitelnosti navrhovaných prvků popisu přineslo mapování i náměty na další potenciálně použitelné prvky z relevantních jmenných prostorů, jež se zatím v našem schématu nenacházejí a jejichž začlenění bude v budoucnu zvaženo: *účel, použití, uživatelské určení, práva, nabízené služby, frekvence aktualizace, způsob zobrazení, stupeň formalizace, ontologický jazyk, uplatněné paradigma reprezentace znalostí*.

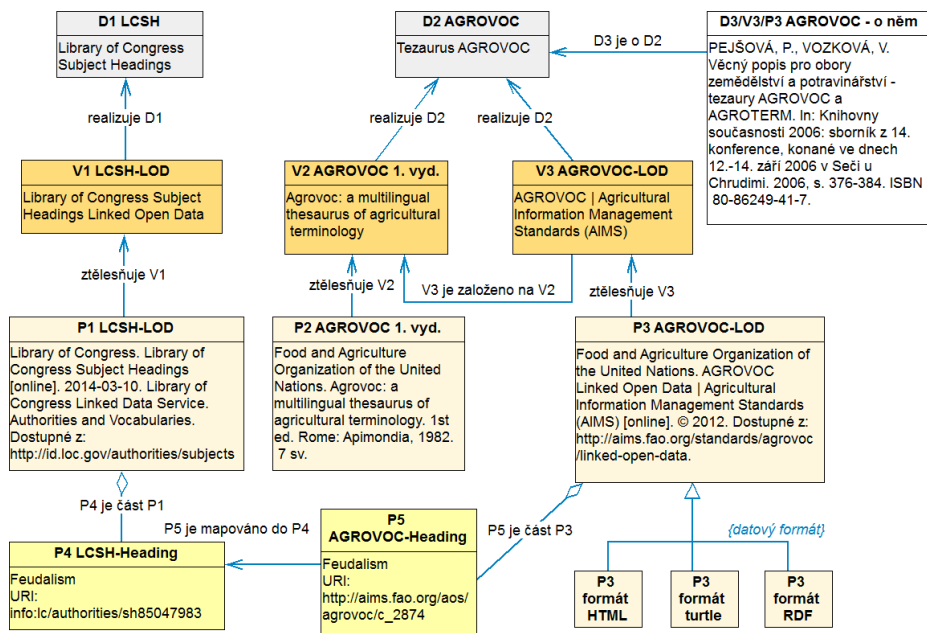
| prvek popisu<br>ve znalostní<br>bázi | jmenný<br>prostor |               |       |       |               |       |       |            |        |           |                            |                |                    |        |                |      |            |                       |          |        |           |
|--------------------------------------|-------------------|---------------|-------|-------|---------------|-------|-------|------------|--------|-----------|----------------------------|----------------|--------------------|--------|----------------|------|------------|-----------------------|----------|--------|-----------|
|                                      | datum             | identifikátor | jazyk | jméno | klíčové slovo | název | popis | dostupnost | formát | typ (SOZ) | odkaz na související zdroj | odkaz na zdroj | alternativní název | médium | počet jednotek | role | typ vztahu | bibliografická citace | definice | faseta | struktura |
|                                      | 8                 | 8             | 8     | 8     | 8             | 8     | 8     | 7          | 7      | 7         | 7                          | 6              | 5                  | 5      | 5              | 4    | 4          | 2                     | 2        | 1      | 1         |
| AP NKOS                              | ✓                 | ✓             | ✓     | ✓     | ✓             | ✓     | ✓     |            | ✓      | ✓         | ✓                          | ✓              |                    | ✓      | ✓              | ✓    | ✓          |                       |          |        |           |
| BARTOC                               | ✓                 | ✓             | ✓     | ✓     | ✓             | ✓     | ✓     | ✓          | ✓      | ✓         | ✓                          |                | ✓                  | ✓      |                |      |            |                       |          |        |           |
| DCAT                                 | ✓                 | ✓             | ✓     | ✓     | ✓             | ✓     | ✓     | ✓          | ✓      |           |                            | ✓              |                    | ✓      | ✓              | ✓    |            |                       |          |        |           |
| DCMI-terms                           | ✓                 | ✓             | ✓     | ✓     | ✓             | ✓     | ✓     | ✓          | ✓      | ✓         | ✓                          | ✓              | ✓                  | ✓      |                | ✓    | ✓          | ✓                     |          |        |           |
| FRBR/FRSAD                           | ✓                 | ✓             | ✓     | ✓     | ✓             | ✓     | ✓     | ✓          | ✓      | ✓         | ✓                          |                |                    | ✓      | ✓              |      |            |                       | ✓        |        | ✓         |
| OMV                                  | ✓                 | ✓             | ✓     | ✓     | ✓             | ✓     | ✓     | ✓          |        | ✓         | ✓                          | ✓              | ✓                  |        | ✓              | ✓    | ✓          |                       |          |        |           |
| Schema.org                           | ✓                 | ✓             | ✓     | ✓     | ✓             | ✓     | ✓     | ✓          | ✓      | ✓         | ✓                          | ✓              | ✓                  |        |                |      |            | ✓                     | ✓        |        |           |
| TaxoBank                             | ✓                 | ✓             | ✓     | ✓     | ✓             | ✓     | ✓     | ✓          | ✓      | ✓         | ✓                          | ✓              | ✓                  |        | ✓              |      | ✓          |                       |          |        | ✓         |

Obr. 8 Mapování prvků popisu SOZ do relevantních jmenných prostorů

### 3.3 Příklady uplatnění metadatového schématu v popisu SOZ ve znalostní bázi

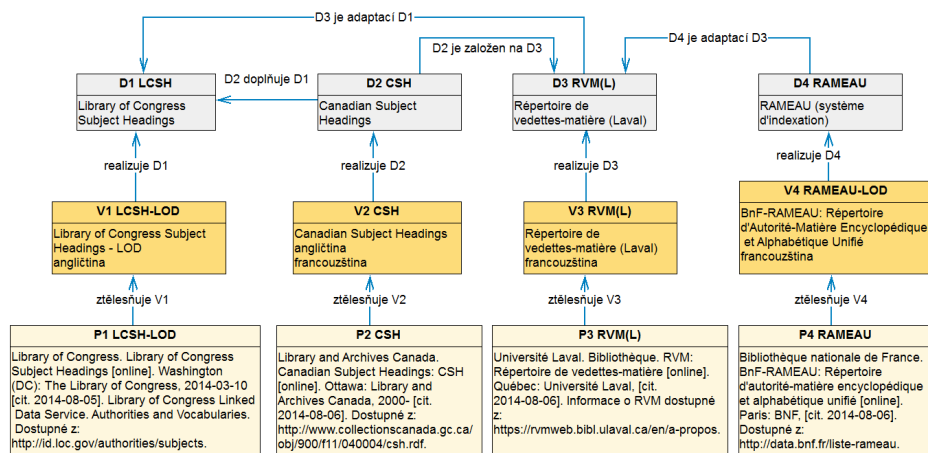
Základní strukturální prvky metadatového popisu jsou opět znázorněny pomocí diagramu tříd v jazyce UML. Klíčovou strukturální komponentou každého popisu je trojice využívaných entit z modelu FRBR – *Dílo*, *Vyjádření* a *Provedení*. Vztah *Dílo* – *Vyjádření* je označen asociací *realizuje*, vztah *Vyjádření* – *Provedení* je označen asociací *ztělesňuje*.

Příklad na obr. 9 graficky reprezentuje strukturu popisu dvou SOZ: Předmětového hesláře Kongresové knihovny (LCSH, VIAF ID díla: <http://viaf.org/viaf/211744653/>) a tezauru AGROVOC (zatím bez přidělení VIAF ID díla). Oba systémy jsou popsány na úrovni díla (D1 a D2), vyjádření (V1, V2 a V3) a provedení (P1, P2, P3, P4 a P5). U LCSH je znázorněna situace jednoho díla, jednoho vyjádření ve formě propojených otevřených dat a jednoho provedení. S využitím vztahu P4 je část P1 je provedení popsáno i na jemnější úrovni granularity. Provedení P1 je celek, tj. celý heslář, a provedení P4 LCSH-Heading, popisující heslo „Feudalismus“, je jeho část. Model FRBR by zřejmě umožňoval ještě alternativní řešení tohoto vztahu: jednotlivá hesla by se dala



Obr. 9 Schéma vazeb tezauru a předmětového hesláře ve znalostní bázi

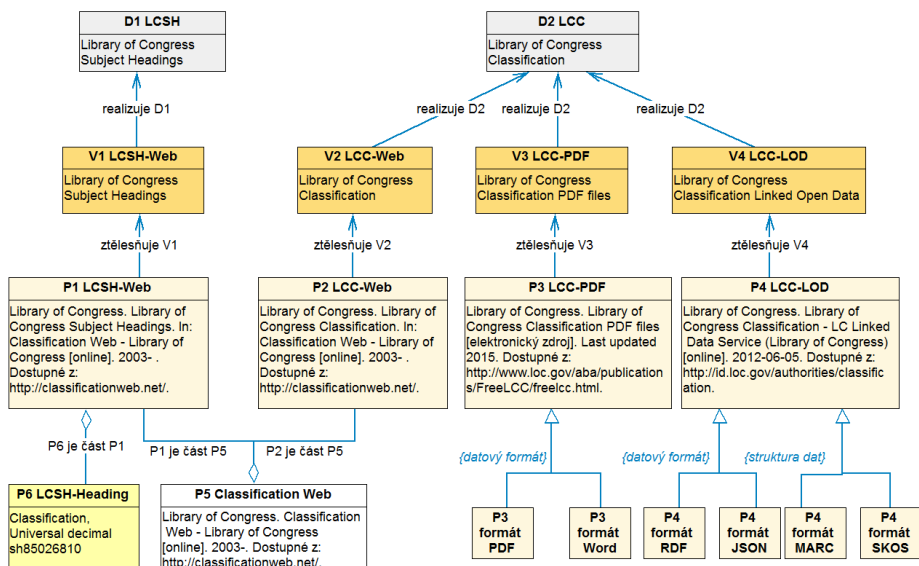
považovat za samostatná díla, která jsou částí jiného díla, tj. hesláře. Toto řešení by ale v případě rozsáhlých SOZ s desítkami až statisíci jednotek bylo dosti neekonomické a strukturovaný popis hesel na úrovni díla, vyjádření a provedení by vedl k značné redundanci. Obdobná linie dílo – vyjádření – provedení, opět až do úrovně jednotlivého záznamu (D2 – V3 – P3 – P5), je znázorněna v případě tezauru AGROVOC ve formě propojených otevřených dat. Provedení P3 AGROVOC-LOD je k dispozici ve třech datových formátech (HTML, turtle a RDF), což je znázorněno úspornou technikou generalizace, která umožňuje informaci o dostupných formátech vztáhnout nejen na celý tezaurus, ale i na jeho jednotlivé deskriptory. V tezauru AGROVOC ve formě LOD bylo jeho správcem provedeno výběrové mapování do hesláře LCSH (konkrétně jde o 1 075 deskriptorů ze skupiny Všeobecnosti), což znázorňuje asociace *je mapováno do* mezi provedeními (záznamy) P5 a P4. Popis tezauru AGROVOC v tomto příkladu zahrnuje kromě vyjádření ve formě LOD i popis historicky první verze tezauru (V2), která byla vydána v sedmisvazkové tištěné podobě v roce 1982 (P2). Vzájemný vztah těchto dvou vyjádření znázorňuje asociace *V3 je založeno na V2*. Třetím dílem (D3) v tomto příkladu je příspěvek ze sborníku, jehož tématem je tezaurus AGROVOC (toto dílo je v grafice na obr. 9 prezentováno ve zjednodušené formě jako jeden objekt, agregovaný s vyjádřením i provedením). Asociace *D3 je o D2* je ukázkou realizace konceptu z modelu FRBR, který stanoví, že tématem díla může být cokoli, tedy i jiné dílo. V příkladu si můžeme všimnout dvou způsobů vyjádření vzájemných vztahů SOZ. První způsob je přímým explicitním vyjádřením vztahů dílo – dílo (*D3 je o D2*), vyjádření – vyjádření (*V3 je založeno na V2*) nebo provedení – provedení (*P5 je mapováno do P4*). Druhý způsob vyjádření vztahů je nepřímý, vyplývající z inference: díla D1 LCSH a D2 AGROVOC mají vzájemný vztah, který vyplývá z toho, že jejich jednotlivá provedení nebo dokonce části jejich provedení jsou navzájem mapovány.



Obr. 10 Schéma vazeb několika předmětových heslářů ve znalostní bázi

Příklad na obr. 10 je ukázkou možností, jak lze pomocí konstrukčních prvků z modelu FRBR zachytit často velmi spletité vztahy systémů organizace znalostí. V diagramu jsou znázorněny čtyři samostatné systémy předmětových hesel – Předmětový heslář Kongresové knihovny (D1 LCSH), Kanadský předmětový heslář (Canadian Subject Headings – D2 CSH), Předmětový heslář Knihovny Lavalské univerzity v Quebecu (D3 RVM(L)) a francouzský heslář RAMEAU (D4) (<http://viaf.org/viaf/202974143>). Obsahové vztahy mezi těmito systémy jsou následující: heslář RVM(L) vznikl v roce 1946 překladem části LCSH z angličtiny do francouzštiny, přičemž tento překlad byl následně doplněn o další hesla odrážející kanadské realie. Na hesláři RVM(L) je založen heslář CSH, který je od 60. let 20. století k dispozici v anglické verzi, posléze doplněné do dvojjazyčné podoby (s anglickou a francouzskou verzí). Heslář CSH se používá v praxi kanadských knihoven současně s heslářem LCSH, který doplňuje o kanadské realie. Adaptací kanadského hesláře RVM(L) vznikl francouzský heslář RAMEAU. Reprezentací této skutečnosti jsou čtyři asociační vztahy: D3 je adaptací D1, D2 je založen na D3, D2 doplňuje D1 a D4 je adaptací D3. V tomto příkladu byla zvolena varianta znázornění vztahů mezi čtyřmi hesláři na nejvyšší úrovni obecnosti – vztahy jsou definovány na úrovni díla. Další možnou variantou popisu by mohl být záznam stejných vztahů (*adaptace, založeno na, doplněk*) na jemnější úrovni granularity, tj. mezi konkrétními vyjádřeními nebo dokonce mezi provedeními, eventuálně mezi vyjádřeními a dílem nebo mezi provedením a vyjádřením.

Příklad na obr. 11 ukazuje strukturu popisu dvou produktů Kongresové knihovny – předmětového hesláře (Library of Congress Subject Headings – D1 LCSH) a klasifikace (Library of Congress Classification – D2 LCC, <http://viaf.org/viaf/203733980>). Heslář LCSH je tentokrát popisován ve své komerční verzi, zpřístupňované v rámci portálu Classification Web, opět až do úrovně jednotlivých záznamů (vztah P6 je část P1 vyjadřuje, že předmětové heslo „Classification, Universal decimal“ s identifikátorem „sh85026810“ je součástí hesláře LCSH). Součástí portálu je i klasifikace LCC. Integrovaní obou SOZ do portálu Classification Web znázorňují vztahy P1 je část P5 a P2 je část P5. V případě klasifikace LCC jsou zachyceny i další dvě formy vyjádření – V3 je volně dostupný statický formát PDF a MS Word, V4 popisuje LCC ve formátu propojených



Obr. 11 Schéma vazeb klasifikace a předmětového hesláře ve znalostní bázi

otevřených dat. Formát, v němž jsou propojená data dostupná, je zaznamenán ve vzájemně kombinovatelných fasetách *datový formát* a *struktura dat*.

## Závěr

V průběhu sběru dat o systémech organizace znalostí se potvrdila hypotéza, že klasická katalogizace ve formátech typu MARC a podle stávajících pravidel (např. AACR aj.) není pro účel popisu SOZ v kontextu projektované znalostní báze vyhovující. Existující proprietární řešení popisu SOZ (například systém TaxoBank) jsou určitě zajímavá a inspiřující, ale veřejnosti nejsou poskytnuty relevantní metadatové specifikace a dokumentace. Byla proto testována hypotéza o použitelnosti teoreticky koncipovaných modelů FRBR a FRASD a také aplikačního profilu AP NKOS. V obou případech se ovšem ukázalo, že tyto modely jsou vysoce abstraktní a před vlastní implementací v konkrétním softwarovém prostředí je nezbytné provést jejich analýzu a doplnit vlastní specifikace. Naplňování takto vytvořené struktury metadat konkrétními daty přineslo praktické zkušnosti a umožnilo zaznamenat problematické aspekty. Model FRBR je intelektuálně náročný co do určení, čím naplnit jednotlivé entity nejvyšší úrovně. Potíže nastávají už v případě určení identifikátorů a forem názvů pro dílo a vyjádření, protože obecně akceptované názvy a identifikátory např. v systému VIAF jsou zatím k dispozici jenom v malé míře. Problematické je rovněž rozhodnout, jak rozdělit údaje v prvcích *popis* a *datum* v entitách dílo a vyjádření, aby nedocházelo k zbytečné redundanci. Tento problém souvisí s obecným problémem, jak oddělit obsah (dílo) a formu (vyjádření) a je třeba ho řešit individuálně v každém případě popisu. Další problémy přináší rozhodnutí, kdy popisovanou jednotku považovat za nové dílo nebo vyjádření. Aplikace obecných principů formulovaných ve studii FRBR často nabízí více řešení a v praxi je ovlivněna mnoha faktory. Je například vhodnější považovat všechna vydání tezauru AGROVOC za jedno dílo, nebo jsou rozdíly mezi 1. vydáním z roku 1982 a současnou verzí, přepracovanou do ontologické

struktury<sup>31</sup> a formátu propojených otevřených dat, už tak výrazné, že je vhodné je považovat za dvě různá, byť vzájemně propojená díla? Jak postupovat v případě četných typů vydání Mezinárodního desetinného třídění? Je „UDC Summary“ v 52 jazycích třeba považovat za 52 vyjádření? Co považovat za kritérium určující, že se jedná o specifické vyjádření MDT – jazyk, vydání nebo typ (úplné, střední, zkrácené ad.)?

Specifický problém představuje použití entity „Thema“ z modelu FRSAD pro vyjádření obsahu SOZ. Systémy organizace znalostí často spadají do kategorie víceoborových děl, u nichž se v praxi obvykle téma vyjadřuje z hlediska formálního (např. MDT je klasifikace, LCSH je předmětový heslář, dokonce nejsou ojedinělé případy typu DDT je Deweyho desetinné třídění). Takový přístup by ovšem ve znalostní bázi vedl k duplicitě, téma popsané tímto způsobem by se krylo s určením typu SOZ. To ostatně potvrzuje i studie FRSAD, která používá pro tento údaj termín „*isness*“ s tím, že jej nepovažuje za určení tématu, nýbrž za jeden z atributů zdroje. Ani u doménově specifických SOZ, u nichž lze určit, jaké věcné oblasti se týkají termíny tvořící jejich slovník (např. AGROVOC zahrnuje problematiku zemědělství), nejde o totéž, jako při určování obsahu („*aboutness*“) ve standardních odborných textech (AGROVOC bezesporu není o zemědělství, DDT bezesporu není o Deweyho desetinném třídění).

Problémem aplikačního profilu AP NKOS v jeho stávající verzi je, že se obtížně zjišťují některé specifické hodnoty – zřejmě se předpokládá, že údaje budou zapisovat producenti SOZ a nikoli zpracovatelé, např. knihovníci. Naplňování prvků popisu SOZ podle AP NKOS třetí stranou vyžaduje často základní průzkum na úrovni badatelského popisu muzejních nebo archivních jednotek. Tento aspekt není na závadu v případě projektované znalostní báze, kde se předpokládá výzkumná aktivita, ale byl by problematický v případě rutinního popisu. Problémy působí i zdánlivě jednoznačný údaj *datum* – u elektronických zdrojů je na úrovni díla a vyjádření často obtížně určitelný časový okamžik vzniku SOZ. Navíc vzhledem k dynamické povaze SOZ jsou kromě data vzniku zajímavé i údaje o změnách, mnohdy i na mikroúrovni jednotlivých záznamů, takže se počet údajů zvyšuje. V případě údaje *datum* se ukazuje být na místě definice ze slovníku DC-terms (dc:date) – „*určitá doba nebo časové období spojené s nějakou událostí v životním cyklu zdroje [...] na jakékoli úrovni granularity*“. Pro využití potenciálu této definice bude ovšem zapotřebí analyzovat a popsat komplikovaný životní cyklus systémů organizace znalostí. Ukazuje se, že prvky specifické pro SOZ by potřebovaly mít vytvořené slovníky (číselníky), nebo alespoň velmi přesnou specifikaci obsahu daného prvku, jinak hrozí nekonzistence zapisovaných údajů.

Přes tyto dílčí problémy se podle našeho názoru model FRBR a na něm založený aplikační profil AP NKOS při popisu SOZ osvědčil. Jako silná stránka modelu FRBR se ukazuje jeho potenciál vyjadřovat vztahy. Zatím se podařilo všechny vztahy, které byly v rámci analýzy jednotlivých SOZ zjištěny, zaznamenat do znalostní báze pomocí standardizovaných typů vztahů, obsažených ve specifikaci modelu.

Pro budoucí vývoj znalostní báze v jejím popisném modulu považujeme za klíčové určit míru explicitnosti popisu. Teoreticky by bylo možné veškeré vlastnosti, specifické pro systémy organizace znalostí, shrnout do volného textu v rámci jednoho popisného prvku. V současné době je v naší bázi takto implicitně zakomponován do prvku *bibliografická citace* údaj o verzi (vydání), informace o autorských právech a údaj o dostupnosti. V rámci prvku *popis* jsou uváděny formou nestrukturovaného textu údaje o historii, frekvenci aktualizací, uživatelském určení, poskytovaných službách, vnitřní struktuře a vnějších vztazích systému. Výhodou nestrukturovaného volného textu je možnost vyjádřit veškerá specifika, často unikátní pro určité SOZ. Nevýhodou je obtížná

<sup>31</sup> Viz bohatou strukturu definovaných vztahů ve slovníku vztahů Agrontology na: <http://aims.fao.org/aos/agrontology>.

zpětná vyhledatelnost implicitních údajů a praktická nemožnost kvantitativních analýz, filtrování podle zadaného kritéria apod. U strukturovaných údajů jsou výhody a nevýhody opačné. Snadná vyhledatelnost strukturovaných údajů je doprovázena nutností jejich zcela přesné a jednotné specifikace, což je obvykle možné jen za cenu určitého zobecnění a tedy ztráty informace o často unikátních a atypických vlastnostech jednotlivých SOZ. Už například zdánlivě jednoznačný údaj o rozsahu/počtu jednotek SOZ vyžaduje přesně definovat, co je považováno za jednotku. Všechny termíny nebo jen preferované? Hlavní znaky nebo i pomocné znaky? Jak postupovat v případě fasetové struktury?

Ve studii byl podán přehled dosavadních fází řešení projektu znalostní báze organizace znalostí v oblasti návrhu metadatového popisu systémů organizace znalostí. Lze konstatovat, že se podařilo dosáhnout většiny z „pětihvězdičkových“ parametrů popisných prvků, jež jsou zapotřebí k tomu, aby mohly být popisné údaje vystaveny ve formátu propojených otevřených dat. Závěrečným, už víceméně administrativním krokem bude přidělení URI popisným prvkům, tak aby byly jednoznačně identifikovatelné a technicky propojitelné se sémanticky mapovanými daty z ostatních relevantních jmenných prostorů.

---

Studie je dílčím výstupem řešení projektu  
*NAKI DF13P01OVV013 Znalostní báze pro obor organizace informací a znalostí,*  
realizovaného na ÚISK FF UK v Praze.

---

### Použitá literatura:

- BRATKOVÁ, Eva, KUČEROVÁ, Helena. Systémy organizace znalostí a jejich typologie. In: *Knihovna*. 2014, **25**(2), 5-29. ISSN 1801-3252 (Print). ISSN 1802-8772 (Online). Dostupný také z: <http://oldknihovna.nkp.cz/pdf/1402/142005.pdf>. Anglická verze dostupná z: <http://oldknihovna.nkp.cz/pdf/1402sup/142001.pdf>.
- COYLE, Karen, BAKER, Thomas. *Guidelines for Dublin Core Application Profiles* [online]. Dublin (Ohio): DCMI, 2009-05-18 [cit. 2015-02-26]. Dostupné z: <http://dublincore.org/documents/profile-guidelines/>.
- Dublin Core Metadata Initiative. NKOS Task Group. NKOS AP Elements. In: *DCMI NKOS Task Group* [online]. 2013, updates 2014-04-03 final, polished 2014-08-03 [cit. 2015-02-26]. Dostupné z: [http://wiki.dublincore.org/index.php/NKOS\\_AP\\_Elements](http://wiki.dublincore.org/index.php/NKOS_AP_Elements).
- Dublin Core Metadata Initiative. NKOS Task Group. NKOS Vocabularies. In: *DCMI NKOS Task Group* [online]. 2013, updated 2013-12-16 [cit. 2015-02-26]. Dostupné z: [http://wiki.dublincore.org/index.php/NKOS\\_Vocabularies](http://wiki.dublincore.org/index.php/NKOS_Vocabularies).
- Dublin Core Metadata Initiative. NKOS Task Group. NKOS Vocabularies. 2., KOS Types Vocabulary. In: *DCMI NKOS Task Group* [online]. 2013, updated 2013-12-16 [cit. 2015-02-26]. Dostupné z: [http://wiki.dublincore.org/index.php/NKOS\\_Vocabularies#KOS\\_Types\\_Vocabulary](http://wiki.dublincore.org/index.php/NKOS_Vocabularies#KOS_Types_Vocabulary).
- FONS, Ted, PENKA, Jeff, WALLIS, Richard. OCLC's Linked Data Initiative: Using Schema.org to Make Library Data Relevant on the Web. In: *Information Standards Quarterly*. Spring/Summer 2012, **24**(2/3), 29-33. ISSN 1041-0031. Dostupné také z: [http://www.niso.org/apps/group\\_public/download.php/9408/IP\\_Fons-et-al\\_OCLC\\_isqv24no2-3.pdf](http://www.niso.org/apps/group_public/download.php/9408/IP_Fons-et-al_OCLC_isqv24no2-3.pdf).
- ISO 24156-1:2014. *Graphic notations for concept modelling in terminology work and its relationship with UML -- Part 1: Guidelines for using UML notation in terminology work*. 1st ed. Geneva: International Organization for Standardization, 2014. 24 s.
- PALMA, Raúl, HARTMANN, Jens, HAASE, Peter. *OMV: Ontology Metadata Vocabulary for the Semantic web* [online]. Version 2.4.1. S.l.: OMV Consortium, March 2009 [cit. 2015-02-28]. 76 s. OMV Report. Dostupné z: <http://freerf.dl.sourceforge.net/project/omv2/OMV%20Documentation/OMV-Reportv2.4.1.pdf>.
- World Web Consortium. *Data catalog vocabulary (DCAT)* [online]. W3C Recommendation 16 January 2014. Ed. Fadi MAALI, John ERICKSON. 2014-01-16 [cit. 2015-02-28]. Dostupné z: <http://www.w3.org/TR/vocab-dcat/>.