

Systemy organizace znalostí a jejich typologie

Knowledge organization systems and their typology

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:

Termín „systém organizace znalostí“ (angl. Knowledge Organization System, zkratka KOS) není zatím součástí české odborné terminologie (ta používá termín „selekční jazyk“). Ve světové literatuře i praxi se však používá již více než 15 let pro souhrnné označení slovníků, seznamů autorit, předmětových heslářů, klasifikací, tezurů, ontologií a dalších nástrojů organizace znalostí, použitelných v prostředí digitální síťové komunikace, reprezentované aktuálně technologií propojených otevřených dat. Studie předkládá předběžné výsledky výzkumu aktuálního stavu současných systémů organizace znalostí, realizovaného v rámci projektu DF13P01OVV013 programu NAKI „Znalostní báze pro obor organizace informací a znalostí“. Výzkum je metodologicky založen na empirické analýze systémů organizace znalostí, které jsou registrovány v prototypu navrhované znalostní báze. Vychází také z analýzy navržených nebo již v praxi aplikovaných typologií prezentovaných v odborné literatuře nebo v provozovaných registrech systémů organizace znalostí. Studie představuje návrh typologie systémů organizace znalostí pro účely jejich popisu ve znalostní bázi. Dílčím výsledkem je i pracovní vymezení termínu „systém organizace znalostí“ a jeho předložení k diskusi české odborné veřejnosti.

Klíčová slova: organizace informací a znalostí, systém organizace znalostí, typologie SOZ, registry

Summary:

The term of knowledge organization system (with acronym KOS) does not form part of the present day Czech terminology, using in its place the term „information retrieval language“. The world literature and practice, though, has applied it well over 15 years to designate the sum of vocabularies, author lists, subject headings, classifications, thesauri, ontology and other knowledge organization tools that are applicable in the digital network communication environment, currently represented by linked open data technology. The study presents interim results of current state-of-the-art research in the field of knowledge organization systems, conducted within the project DF13P01OVV013 „Knowledge base for the subject area of knowledge organization“, as part of the NAKI Programme. The research methodology is based upon an empirical analysis of the knowledge organization systems, as registered in the prototype of the suggested knowledge base. It also draws on the analysis of proposed or implemented typologies presented in specialized literature or in the operating KOS registries. The study presents a typology of knowledge organization systems for the purpose of their description within a knowledge base. Yet another outcome appears to be a working definition of the term „knowledge organization system“, as offered for debate among the Czech information professionals.

Keywords: knowledge organization, knowledge organization system, KOS typology, registries

Úvod

Studie předkládá předběžné výsledky výzkumu aktuálního stavu současných systémů organizace znalostí (angl. *Knowledge Organization Systems*, zkratka KOS; v textu používáme doslovný překlad do češtiny a iniciálovou zkratku SOZ, termín ani jeho zkratka však zatím nejsou součástí české odborné terminologie). Cílem výzkumu, realizovaného v rámci projektu programu NAKI „Znalostní báze pro obor organizace informací a znalostí“, bylo navrhnout typologii systémů organizace znalostí pro účely jejich popisu

a kategorizace v prototypu znalostní báze. Výzkum je metodologicky založen na empirické analýze systémů organizace znalostí, které jsou registrovány v prototypu navrhované znalostní báze. Vychází také z analýzy navržených nebo již v praxi aplikovaných typologií prezentovaných v odborné literatuře nebo v provozovaných registrech systémů organizace znalostí.

Text studie je členěn do tří částí. V první části vymezujeme obsah a rozsah pojmu „systém organizace znalostí“ a jeho vztah k tradičně v českém prostředí používanému termínu „selekční jazyk“. Druhá část představuje výsledky komparace nejvýznamnějších typologií, uplatněných v rámci aktuálních výzkumných aktivit a v nově založených registrech systémů organizace znalostí. Ve třetí části nabízíme návrh vlastní typologie systémů organizace znalostí, založený na jejich sémantické síle.

1 Vymezení pojmu „systém organizace znalostí“

Termín *systém organizace znalostí* je relativně novým¹ označením pro nástroje a pomůcky, které se pod různými jmény tradičně používají při práci s informacemi a znalostmi jak v paměťových institucích, tak v ostatních oblastech společenské i individuální sféry, všude tam, kde je potřeba zajistit organizované uložení informací a efektivní přístup k nim. O tom, že jde o nový termín s neustáleným výkladem, svědčí i skutečnost, že s výjimkou Tezauru ASIS&T² zatím není zařazen do žádného z dalších oborových tezaurů ani klasifikací³. Vymezení pojmu je zatím vágní a navíc se pojetí různých autorů liší. To názorně dokládají následující příklady definic, lišících se jak přístupem k vymezení, tak svými formulacemi. První příklad definice SOZ se soustředí na určení jeho funkce: „*Systémy/služby organizace znalostí [...] modelují sémantickou infrastrukturu oboru. Jejich začlenění do webových služeb usnadňuje objevování a vyhledávání zdrojů. Fungují jako sémantické mapy a umožňují společnou orientaci indexátorů a budoucích uživatelů (lidských nebo strojových)*“⁴. Ve druhém příkladu je definice prostřednictvím funkce doplněna vyjmenováním typických instancí (příkladů) SOZ: „*Termín systém organizace znalostí je určen k označení všech typů schémat pro organizaci informací a podporu znalostního managementu. Systémy organizace znalostí zahrnují klasifikační a kategorizační schémata, jež organizují materiál na všeobecné úrovni, předmětová hesla, jež umožňují detailnější přístup, a soubory autorit, jež slouží k řízení variantních*

¹ Podle G. Hodgeové byl vytvořen v roce 1998 na ustavujícím setkání Pracovní skupiny pro síťově propojené systémy organizace znalostí (angl. *Networked Knowledge Organization Systems Working Group – NKOS WG*), pořádaném 27. 6. 1998 v rámci konference *ACM Digital Libraries '98* v Pittsburghu (HODGE, Gail. *Systems of knowledge organization for digital libraries: beyond traditional authority files*. Washington: The Digital Library Federation, Council on Library and Information Resources, April 2000, s. 3. ISBN 1-933645-06-7 (DLF). Dostupné také volně z webu vydavatele: <http://www.clir.org/pubs/reports/pub91/pub91.pdf>).

² Termín „*knowledge organization systems*“ zde figuruje jako jeden z vrcholových termínů ve fasetě „*knowledge and information*“.

³ Konkrétně byly prověřeny tyto tezaury a klasifikace: Tezaurus LISA, Tezaurus LISS, Tezaurus INSPEC, Tezaurus PASCAL, KO Literature (*Classification System for Knowledge Organization Literature* [online]. ISKO 2011-2012, last update 2012-04-13 [cit. 2014-08-25]. Dostupné z: <http://www.isko.org/scheme.php>).

⁴ TUDHOPE, Douglas, KOCH, Traugott. New applications of knowledge organization systems: introduction to a special issue. In: *Journal of digital information* [online]. 2004, 4(4) [cit. 2014-08-25]. ISSN 1368-7506.

Dostupné z: <https://journals.tdl.org/jodi/index.php/jodi/article/view/109/108>.

verzí klíčových informací, jako například geografických nebo osobních jmen. [...] zahrnují rovněž vysoce strukturované slovníky, jako jsou tezaury, a méně tradiční schémata jako sémantické sítě a ontologie“⁵. Třetí příklad navzdory své stručnosti vymezuje SOZ jak pomocí jeho funkce, tak jeho struktury: „Systém organizace znalostí považujeme za reprezentaci znalostí založenou na pojmech s různými stupni vzájemných vztahů“⁶.

Pro účely našeho výzkumu pohlížíme na systém organizace znalostí jako na nástroj vytvořený za účelem podpory procesu organizace znalostí. Za *organizaci znalostí* považujeme jakoukoli záměrnou činnost spočívající v zavádění struktury do existujících zaznamenaných znalostí s cílem umožnit jejich uložení a usnadnit k nim přístup. *Aktivita spojené s uložením* zasahují takřka celý životní cyklus znalostí, od jejich reprezentace (vyjádření) přes případné sdělování či sdílení, zaznamenání, publikování, zpracování (identifikaci, popis, obsahovou analýzu, indexaci, transformaci) až po samotné uložení. V opačném směru jde o sadu *aktivit zajišťujících přístup* ke znalostem, od vyjádření potřeby znalostí přes případnou formulaci dotazu k vyhledávání požadovaných či objevování nových informací a k prezentaci a uspořádání výsledků vyhledávání z organizované kolekce, případně generování nových znalostí. V obou případech plní SOZ funkci pomůcky, která má formu modelu budoucí struktury organizované množiny. V současné době se tyto pomůcky, původně používané v izolovaném prostředí jednotlivých institucí, transformují na nástroje digitální síťové komunikace, reprezentované aktuálně technologií propojených otevřených dat (Linked Open Data, LOD). Značný počet a rozmanitost aktivit organizace znalostí se analogicky odráží i v širokém vymezení extenze pojmu SOZ. K tomu přistupuje rozmanitost prostředí, kde se mohou jednotlivé aktivity realizovat – kromě tradiční oblasti paměťových institucí se jedná o řadu dalších profesionálních i laických oblastí. Pokud přijmeme takto široce definovaný pojem, musíme akceptovat, že zahrnuje značně rozsáhlou množinu často dramaticky rozdílných instancí, u nichž budeme jen obtížně hledat společné vlastnosti.

Jak ukazují výše uvedené příklady, samotná definice systému organizace znalostí může být vzhledem k široké extenzi pojmu jen velmi obecná. Považujeme proto za vhodné ji v následující části konkretizovat detailnějšími pohledy jak na funkci, tak na strukturu a atributy SOZ, jak to uvádějí významní autoři oboru.

1.1 Funkce, struktura a rozlišující atributy systémů organizace znalostí

D. Soergel v publikaci věnované organizaci znalostí v digitálních knihovnách⁷ specifikoval šest oblastí funkcionality systémů organizace znalostí, doplněných konkrétními příklady: *sémantická* funkce (např. sémantická mapa jednotlivých oborů a vztahů mezi nimi, objasnění významu pojmů jejich zasazením do kontextu v rámci klasifikace, slovníky datových prvků a pojmové báze znalostních systémů), *komunikační* funkce (např. pomoc

⁵ HODGE, Gail. *Systems of knowledge organization for digital libraries: beyond traditional authority files*. Washington: The Digital Library Federation, Council on Library and Information Resources, April 2000, s. 1. ISBN 1-933645-06-7 (DLF). Dostupné též z: <http://www.clir.org/pubs/reports/pub91/pub91.pdf>.

⁶ SOUZA, Renato Rocha, TUDHOPE, Douglas, ALMEIDA, Maurício Barcellos. Towards a taxonomy of KOS: dimensions for classifying knowledge organization systems. In: *Knowledge organization*. 2012, 39(3), 181. ISSN 0943-7444.

⁷ SOERGEL, Dagobert. Digital libraries and knowledge organization. In: Sebastian Ryszard KRUK, Bill McDANIEL, ed. *Semantic digital libraries*. Berlin: Springer, 2009, s. 29. ISBN 978-3-540-85433-3 (Print). ISBN 978-3-540-85434-0 (Online). Dostupné také komerčně z (DOI): http://dx.doi.org/10.1007/978-3-540-85434-0_2.

při učení a osvojování znalostí, při správné formulaci otázky, při porozumění čtenému textu, při formulaci srozumitelného textu), *konceptuální* funkce (např. při stanovení výzkumných cílů a při objasňování problémů, při konzistentním sběru dat a sestavování srovnávacích statistik), „*akční*“ funkce (podpora realizace procesů a činností, např. využití klasifikace nemocí pro diagnostiku), *selekční* funkce (např. vyhledávání informací, zboží a služeb, přístup ke znalostem), *lingvistická* funkce (např. tvorba slovníků využitelných lidmi i stroji při práci s jazykem).

R. Szostak definuje jako základní stavební prvky struktury SOZ slova (*termíny*) vyjadřující pojmy, jejichž význam je jednoznačně definován v řízeném slovníku, a *syntaktická pravidla*, stanovená pro jejich spojování.⁸ V souladu se současným trendem směřování k propojeným datům v sémantickém webu považuje za užitečné uvažovat o obecné struktuře systému organizace znalostí v pojmech vyjadřovaných jazykem RDF (Resource Description Framework), jehož jádrem jsou tvrzení tvořená trojicemi subjekt (podmět) – predikát/vlastnost⁹ (přísudek/atribut) – objekt (předmět). Jak subjekt, tak objekt jsou v systému organizace znalostí reprezentovány společnou hierarchií pojmů (např. třídami v klasifikaci) zastupujících věci. Predikát, označovaný v RDF též jako vlastnost, je reprezentován buď jako atribut věci, nebo jako vztah subjektu a objektu (což je z teoretického pohledu specifický typ atributu, který je sdílen více entitami). Dalšími strukturními prvky systému organizace znalostí jsou pak hierarchie vlastností a hierarchie vztahů. Prvky těchto tří klasifikací – *věci*, *vlastnosti* věcí a *vztahy* věcí, resp. slova (*termíny*), jež je vyjadřují, lze navzájem libovolně kombinovat do trojic RDF.

Atributy představují v aristotelské tradici rozlišovací vlastnosti (*differentia specifica*), potřebné pro správnou definici. M. Zengová a M. Žumerová je uvedly a komentovaly ve svém referátu věnovaném přípravě Metadatového profilu pro registry slovníků SOZ.¹⁰ Podle autorek jsou systémy organizace znalostí specifickými informačními zdroji, které vykazují následující společné vlastnosti, jež je odlišují od jiných typů tvůrčích děl: *kontinuita* (průběžná aktualizace a rozvoj, sledující změny reálného světa), směřující až ke *správě na mikroúrovni* (tj. verzování, datace a údaje o odpovědnosti a původu na úrovni jednotlivých pojmů či termínů), *různorodost* členů „rodiny“ (ze stejného kmenového systému se dynamicky odvozují zkrácené, rozšířené nebo upravené verze či překlady), *sdílené autorství* (v průběhu znovupoužívání, mapování, spojování a derivování jak v rámci stejné „rodiny“, tak i mezi nimi, se mění i autorství), *složitě vzájemné vztahy* (např. mezi různými vydáními, různými jazykovými verzemi včetně různých typů překladů a lokálních úprav a různými tištěnými a elektronickými formáty, jak je to možné pozorovat např. u Deweyho desetinného třídění). Tyto atributy jednoznačně svědčí o tom, že systémy organizace znalostí spadají do komplikované a obtížně zpracovatelné kategorie integračních informačních zdrojů. Situaci dále ztěžuje skutečnost, že mnohdy nefungují jako samostatné jednotky, ale jsou integrovány do jiných zdrojů, typicky do online rešeršních systémů a vyhledávacích aplikací.

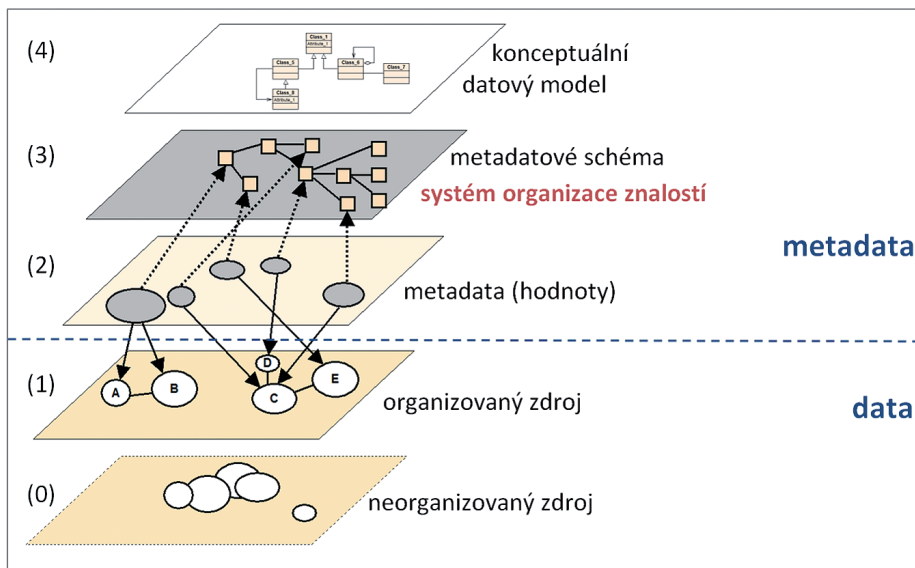
⁸ SZOSTAK, Rick. Classification, ontology, and the Semantic web. In: *Advances in classification research online* [online]. 2013, 24(1) [cit. 2014-08-25], 30-37. DOI:10.7152/acro.v24i1.14674. ISSN 2324-9773. 24th ASIS SIG/CR Classification Research Workshop. Dostupné z: <https://journals.lib.washington.edu/index.php/acro/article/view/14674>.

⁹ V angl. *predicate/property*.

¹⁰ ZENG, Marcia Lei, ŽUMER, Maja. A metadata application profile for KOS vocabulary registries. In: *Knowledge organization: pushing the boundaries: ISKO UK [3rd] Biennial Conference, 8th - 9th July 2013, London* [online]. London: ISKO UK, 2013 [cit. 2014-08-25]. Dostupné z: <http://www.iskouk.org/conf2013/papers/ZengPaper.pdf>.

1.2 Úrovně (vrstvy) systémů organizace znalostí

Na obrázku 1 spojujeme dva pohledy na systém organizace znalostí: 1) funkční pohled, který vidí systém organizace znalostí jako pomůcku/nástroj organizace znalostí, a 2) objektový pohled, který nahlíží na systém organizace znalostí jako na specifický typ informačního zdroje, který je rovněž možné zpracovávat/organizovat¹¹.



Obr. 1 Systém organizace znalostí (vrstvy organizace znalostí)

Jednotlivé vrstvy schematicky oddělují významné komponenty organizovaného zdroje znalostí: pro názornost je na nejnižší úrovni uvedena vrstva (0), představující neorganizované zdroje (např. dosud nezpracované archiválie z pozůstalosti). Ve vrstvě (1) je znázorněna struktura, která vznikla uspořádáním zdrojů do skupin podle nějakého kritéria (např. archiválie rozříděné podle svého obsahu do složek). Nad těmito vrstvami, znázorňujícími organizovaná a neorganizovaná data, jsou metadatové vrstvy: vrstva (2) představuje metadata, jež jsou typickým produktem organizace znalostí (např. obsahová charakteristika, aktuální cena, rozměry); ta mohou být fyzickou součástí zdroje (např. tiráž, ex libris, metadata v HTML dokumentu), nebo na něj odkazovat (např. bibliografická citace, třídník MDT). Zatímco vrstva (1) znázorňuje strukturu organizovaných zdrojů, vrstva (3) znázorňuje strukturu metadat, jež je oproti neustále se dynamicky proměňující množině metadat relativně trvalá. Díky této relativně stabilní sémantické struktuře je možné využít při práci s organizovanými zdroji nejen hodnoty metadat, ale i vztahy mezi metadatovými prvky a jejich případné vlastnosti, definované v tzv. metadatovém

¹¹ Na tomto místě je vhodné připomenout, že navzdory abstraktní povaze všech tří slov, z nichž je vytvořen termín systém organizace znalostí, neoznačuje tento termín žádnou abstrakci, ale konkrétní artefakt (tj. záměrně za určitým účelem vytvořenou věc) – pomůcku pro některý z procesů organizace znalostí.

schématu. Tato vrstva je tedy schematickým znázorněním struktury a funkce systému organizace znalostí (např. klasifikace, tezauru, seznamu autorit, číselníku). Vrstva (4) ukazuje další metaúroveň – „strukturu struktury“ metadat (např. FRBR, CIDOC CRM). Konceptuální modely systémů organizace znalostí zobecňují jejich strukturu a umožňují tak teoretický výzkum a zejména implementaci systémů organizace znalostí do současné informační infrastruktury síťového prostředí.

Rozdíl mezi funkčním a objektovým pohledem na systém organizace znalostí se pokusíme naznačit prostřednictvím dvou příkladů.

Příklad 1 uplatňuje **funkční pohled** na SOZ. Předpokládá, že ve vrstvě dat se nacházejí „klasické“ primární dokumenty, například časopisecké články z oboru informační vědy v PDF formátu. Ve vrstvě (0) na obrázku 1 jsou neorganizované články od různých vydavatelů, ve vrstvě (1) články v oborové databázi, např. v agregované plnotextové kolekci producenta EBSCO, vrstva (2) označuje množinu použitých deskriptorů z Tezauru LISS, jimiž jsou články indexovány, ve vrstvě (3) je Tezaurus LISS a vrstva (4) obsahuje obecný model struktury tezauru podle standardu ISO 25964-1¹².

Příklad 2 nahlíží SOZ jako **objekt zpracování**. Předpokládá, že ve vrstvě dat jsou systémy organizace znalostí, může tedy například jít o repozitář systémů organizace znalostí, jako je BioPortal¹³, který slouží jako úložiště pro SOZ z oblasti biomedicíny. Vrstva (0) představuje množinu zatím nijak neuspořádaných systémů, vrstva (1) jsou systémy organizace znalostí, uložené v repozitáři. V metadatových vrstvách je možné si představit například metadata podle standardu Dublin Core (DC): ve vrstvě (2) jsou hodnoty metadatových prvků z připravovaného Metadatového aplikačního profilu pro registry slovníků SOZ (DC-AP NKOS)¹⁴ a Slovníků NKOS¹⁵, vrstva (3) pak obsahuje samotný profil DC-AP NKOS. V metavrstvě (4) je Abstraktní model iniciativy DCMI (DCMI Abstract Model, DCAM)¹⁶.

1.3 Problémy české terminologie

Na terminologické problémy organizace znalostí v českém prostředí detailně poukázal samostatně publikovaný článek¹⁷, proto zde uvádíme jen stručné shrnutí.

Stávající český terminologický systém, zpracovaný Blahoslavem Kovářem v 70. letech 20. století, je zachycen v České terminologické databázi knihovnictví a informační vědy (TDKIV). Pro pojmenování systémů organizace znalostí používá termín *selekční*

¹² Náhled schématu je dostupný z: http://www.niso.org/schemas/iso25964/Model_2011-06-02.jpg.

¹³ National Center for Biomedical Ontology (Spojené státy americké). *BioPortal* [online]. © 2005-2014 [cit. 2014-10-15]. Dostupné z: <http://bioportal.bioontology.org/>. -- K datu citování obsahoval tento repozitář 389 systémů organizace znalostí.

¹⁴ 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. 2014-08-25]. Dostupné z: http://wiki.dublincore.org/index.php/NKOS_AP_Elements.

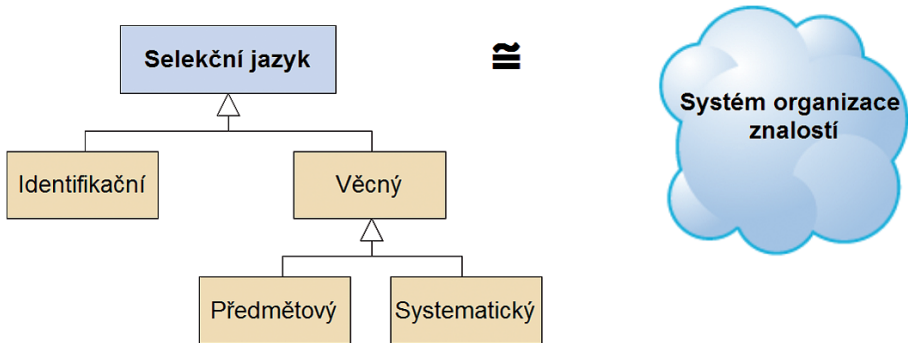
¹⁵ Dublin Core Metadata Initiative. NKOS Task Group. NKOS Vocabularies. In: *DCMI NKOS Task Group* [online]. 2013, updated 2013-12-16 [cit. 2014-08-25]. Dostupné z: http://wiki.dublincore.org/index.php/NKOS_Vocabularies.

¹⁶ Dublin Core Metadata Initiative. *DCMI Abstract Model – DCAM* [online]. 2007-06-04 [2014-08-25], sect. 2. Dostupné z: <http://dublincore.org/documents/abstract-model/#sect-2>.

¹⁷ KUČEROVÁ, Helena. České termíny pro věcné pořádkání informací po 40 letech: příspěvek k terminologické diskusi. In: *ProInflow: časopis pro informační vědy* [online]. 2013, 5(Speciál), [cit. 2014-02-17], s. 1-19. ISSN 1804-2406.

Dostupné z: http://pro.inflow.cz/sites/default/files/pdfclanky/ProInflow_Kucerova_final_0.pdf.

jazyky. Terminologie vychází z procesního pohledu na organizaci znalostí a selekční jazyk chápe jako nástroj procesů pořádání informací. V souladu s členěním pořádání na identifikační a věcné se i selekční jazyky dělí na dvě základní skupiny – identifikační a věcné, jež se dále člení na předmětové a systematické. Za identifikační jsou považovány selekční jazyky orientované na formální atributy zdrojů, věcné selekční jazyky jsou zaměřeny na atributy obsahu. Jako předmětové jsou označeny ty selekční jazyky, jež nezachycují paradigmatické vztahy mezi jednotkami selekčního jazyka. Ty jazyky, jež některý typ paradigmatických vztahů (především hierarchii a asociaci) vyjádřit umožňují, jsou označeny jako systematické. Pojmy „identifikační“ a „věcný“ a hlavně „systematický“ a „předmětový“ poskytují užitečný teoretický mezičlánek mezi abstraktním pojmem nejvyšší úrovně (tj. selekční jazyk) a konkrétními jmény jednotlivých případů procesů organizace znalostí (např. katalogizace, tagování, klasifikace, kategorizace) a jejich nástrojů (např. MDT, soubor autorit VIAF). Základní uvažovanou jednotkou organizace je (klasický tištěný) dokument. V členění selekčních jazyků se uplatňuje výrazná dualita – jsou vymezeny ostré hranice mezi jazyky identifikačními a věcnými, a také mezi jazyky systematickými a předmětovými. Paradigmatické změny, vyvolané v 90. letech 20. století rozšířením elektronických zdrojů a technologií internetu, nejsou zatím v české terminologii reflektovány.



Obr. 2 Česká a světová terminologie systémů organizace znalostí

V citovaném článku věnovaném terminologii věcného pořádání byl předložen návrh nahradit termín „selekční jazyk“ termínem „systém organizace/pořádání znalostí“, případně analogickou rozšířenou verzí „systém organizace/pořádání informací a znalostí“. Bylo konstatováno, že termíny „selekční jazyk“ a „systém organizace znalostí“ lze považovat z pohledu jejich klíčových charakteristik za významově ekvivalentní – oba označují zároveň nástroj i produkt procesu organizace znalostí. Ve skutečnosti nejde o pravou synonymii, spíše o částečné, byť dosti rozsáhlé, překrývání významů termínů. Například přirozený jazyk je možné použít ve funkci selekčního jazyka, už by ho ovšem nebylo možné označit za systém organizace znalostí. Na druhé straně existují systémy organizace znalostí, které by bylo obtížné považovat za selekční jazyky, například sémantické sítě nebo konceptuální schémata. Na obrázku 2 je schematicky naznačen další rozdíl mezi zřetelně definovanou typologií selekčních jazyků a zatím nejednotně a vágně vymezeným pojmem systém organizace znalostí (diverzita stávajících typologií SOZ bude představena v části 2).

Rovněž bylo konstatováno, že oba termíny mají v zásadě metaforickou povahu, přičemž jak metafora „jazyka“, tak metafora „systému“ se jeví jako vhodně zvolené a mají

oporu v teoretických pracích významných představitelů oboru. Metaforu „jazyka“ používají kromě „české školy“ například E. Svenoniová (bibliografický jazyk¹⁸), V. A. Moskovič (informační jazyk¹⁹), používali ji průkopníci informační vědy C. N. Mooers, A. C. Foskett, D. Soergel (indexovací jazyk – *indexing language*) a je obsažena i v současném standardu ISO 25964²⁰ (strukturovaný/řízený slovník – *structured/controlled vocabulary*). Typickým příkladem využití metafory jazyka a jejího přenesení i na široce metaforické chápání termínu slovník (angl. *vocabulary*) komunitou sémantického webu je výklad pojmu ontologie prostřednictvím termínu „slovník“ na stránkách konsorcia W3C²¹. Tento přístup je uplatněn i v rámci schématu Dublin Core. Metadatové schéma tříd, atributů a vztahů je označeno jako „slovník množiny prvků“ (*element set vocabulary*) a číselníky/seznamy hodnot atributů se nazývají „slovníky hodnot“ (*value vocabulary*). Metafora „systému“ je používána členy komunity NKOS²², která v současné době vyvíjí nejvýznamnější výzkumné i aplikační aktivity v oblasti systémů organizace znalostí. Dalším příkladem je použití systémové metafory v názvu schématu SKOS (*Simple Knowledge Organization System*)²³.

Metafory jazyka a systému se vzájemně doplňují a vystihují nutné (je otázka, zda i postačující) podmínky pro to, aby bylo možné nějakou entitu označit za systém organizace znalostí: musí to být systém, tj. daná entita musí být definovatelná prostřednictvím struktury a funkce odpovídajících stanovenému cíli (podpora organizace znalostí), a současně musí splňovat minimální charakteristiky jazyka, tj. musí mít buď slovník, nebo gramatiku (syntaktická pravidla), nebo obojí. Protože zatím chybí „zastřešující“ terminologická metafora, jež by v sobě zahrнула jak jazykové, tak systémové atributy, navrhuje se přiklonit se v české terminologii používáním termínu „systém organizace znalostí“ k metafoře systémové, která by podle našeho názoru usnadnila mezinárodní komunikaci.

1.4 Vlastní pracovní definice systému organizace znalostí

Při formulaci pracovní definice systému organizace znalostí jsme se pokusily využít klasickou aristotelskou metodu definování pomocí bezprostředně nadřazeného rodového pojmu a rozlišujících druhových vlastností. Systém organizace znalostí je v tomto

¹⁸ SVENONIUS, Elaine. Bibliographic languages. In: *The intellectual foundation of information organization*. Cambridge (MA): MIT Press, 2000, chapter 4, s. 53-55. ISBN 0-262-19433-3.

¹⁹ MOSKOVIČ, Volf Abramovič. *Informacionnyje jazyki*. Moskva: Nauka, 1971. 143 s.

²⁰ ISO 25964-1:2011. *Information and documentation – Thesauri and interoperability with other vocabularies – Part 1: Thesauri for information retrieval*. 1st ed. Geneva: International Organization for Standardization, 2011-08-08. 152 s.

²¹ World Wide Web Consortium. Vocabularies. In: *World Wide Web Consortium (W3C)* [online]. Cambridge (MA): World Wide Web Consortium, © 2012 [cit. 2014-08-25]. Dostupné z: <http://www.w3.org/standards/semanticweb/ontology>.

²² LYKKE, Marianne. Networked Knowledge Organization Systems/Services (NKOS). In: Marcia J. BATES, Mary Niles MAACK, ed. *Encyclopedia of library and information sciences*. 3rd ed. Boca Raton (Florida): CRC Press, © 2010, s. 3911-3915. doi:10.1081/E-ELIS3-120044732. ISBN 978-0-8493-9712-7 (soubor, Print). ISBN 978-0-8493-9711-0 (Online). Dostupné komerčně také z: <http://www.tandfonline.com/doi/pdf/10.1081/E-ELIS3-120044732>.

²³ MILES, Alistair, BECHHOFFER, Sean, ed. *SKOS Simple Knowledge Organization System Reference* [online]. W3C Recommendation 18 August 2009 [version]. Cambridge (MA): World Wide Web Consortium, © 2009. Dostupné z: <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>.

případě možné považovat za specifický typ systému, který je od ostatních systémů odlišitelný svými rozlišujícími vlastnostmi. Tato specifika jsou dána jednak jeho funkcí, jednak strukturou, jednak atributy, jež byly charakterizovány v části 1.1.

Výsledkem naší analýzy specifík SOZ, podložené literárním průzkumem, je návrh následující definice: *Systém organizace znalostí je schéma modelující strukturu (tj. prvky a vzájemné vztahy) organizované množiny znalostí. Funkcí systému organizace znalostí je podpora procesů organizace znalostí a přístupu k znalostem. Základním strukturálním prvkem systému organizace znalostí je pojem. Jádrem fyzické reprezentace každého systému organizace znalostí je slovník, tj. formální vyjádření pojmů. Ten je používán pro vyjádření jak sémantiky, tak syntaxe organizovaného celku, případně i pravidel určujících používání struktury.*

2 Typologie systémů organizace znalostí v zahraničí

V návaznosti na předcházející vymezení systémů organizace znalostí (SOZ) je v této části sledována související problematika jejich typologie. Diskuse k možnému dělení SOZ jsou vedeny v zahraničí již více než 15 let. Doposud bylo předloženo několik významných návrhů. Tato část si klade za cíl představit, porovnat a vyhodnotit vybrané typologie SOZ zjištěné ve světové literatuře a také již v konkrétních provozech jejich registrů. Výsledek malé srovnávací analýzy vybraných typologií poslouží jako východisko pro návrh pracovní typologie SOZ, jež bude představena v další části (3). Zpracování této typologie je důležité též pro přípravu znalostní báze v rámci již zmíněného projektu NAKI. Navržené hodnoty typů SOZ jsou potřebné navíc pro kompletní metadatový popis těchto systémů.

Doposud předložené typologie ukazují, že existuje určité společné jádro typů SOZ. Jsou vymezovány na základě podstatných kritérií, jimiž jsou struktura, funkce a případně i uživatelské určení. Přestože je možné zvažovat i jiná hlediska dělení těchto systémů, ta hlavní doposud uplatňovaná jsou podstatná a dominantní. V první podčásti proto budou analyzovány a vyhodnoceny hlavní typologie, tak jak jsou dokumentovány v publikované literatuře včetně standardů. Teoretický pohled bude doplněn v druhé podčásti o analýzu a vyhodnocení typologií z vybraných provozů „živých“ registrů SOZ, které se vyznačují praktickým či pragmatickým pohledem a někdy i „nadhledem“ nad danou problematikou (vybrány byly pouze ty registry, které typologii SOZ skutečně zahrnovaly). Analyzované typologie jsou představeny také formou srovnávací tabulky. Tabulka obsahuje jména typů SOZ v původním znění (v angličtině), textové komentáře pak uvádějí české překlady.

2.1 Typologie systémů organizace znalostí ve světové literatuře

V rámci vymezování pojmu „systém organizace znalostí“ podala G. Hodgeová ve své zprávě jako jedna z prvních také komentovanou typologii těchto systémů²⁴. Typologie vychází z dlouholeté praxe autorky²⁵, navržené typy doporučovala v dané chvíli

²⁴ HODGE, Gail. *Systems of knowledge organization for digital libraries: beyond traditional authority files*. Washington: The Digital Library Federation, Council on Library and Information Resources, April 2000, s. 4-7.

²⁵ Gail Hodgeová působila dlouhá léta v USA v oblasti informačního průmyslu se zaměřením na bibliografické databáze (například ve společnosti BIOSIS aj.) a také v oblasti standardizace informačních systémů.

jako nosné pro aplikaci v rozvíjejících se digitálních knihovnách. Návrh jedenácti v té době dobře známých typů SOZ (viz tab. 1) nebyl novinkou. Významný byl ale fakt, že autorka uvedené typy vymezila kombinací několika jejich podstatných vlastností: strukturou, komplexností, vztahy mezi zahrnutými termíny a také historickou rolí. Důraz na uvedené vlastnosti se promítl i do návrhu uspořádání zahrnutých typů SOZ do tří skupin. Do skupiny nazvané „*Seznamy termínů*“ (*Term Lists*) jsou zahrnuty jednodušší strukturované systémy se známými základními funkcemi, a to *autoritní soubory* (jmenné povahy), dále specializované *glosáře*, obecněji zaměřené *slovníky* (angl. dictionaries) a také v anglosaských zemích často používané *zeměpisné slovníky* (angl. gazetteers). Do druhé skupiny nazvané „*Klasifikace a kategorie*“ (*Classification and Categories*) jsou zahrnuty již více strukturované SOZ s důrazem na vytváření předmětově organizovaných souborů. Autorka do ní zařadila jednak *schémata předmětových hesel*, vyznačující se mělkou strukturou s limitovanou hierarchií, jednak *klasifikační schémata*, *taxonomie* a *kategorizační schémata*, jež bývají podle ní v praxi často zaměňovány. Třetí skupina s názvem „*Seznamy vztahů*“ (*Relationship Lists*) zahrnuje již vysoce nebo komplexně strukturované systémy, pro které jsou typická spojení mezi pojmy. Zařazeny jsou především nejpočetnější SOZ – *tezaury*, založené na pojmech a bohatých vztazích mezi nimi. Hodgeová nově do této skupiny přiřadila i zcela nové typy SOZ, a to *sémantické sítě*, které podle ní strukturují pojmy nikoliv v hierarchickém uspořádání, ale jako síť nebo web s četnými specifickými vztahy, a také *ontologie*, které jsou podle ní specifickými pojmovými modely, jež reprezentují komplex vztahů mezi objekty včetně pravidel a zásad.

Typologie G. Hodgeové byla přijata a dále rozpracována v rámci aktivit Pracovní skupiny NKOS (Networked Knowledge Organization Systems)²⁶. Její čelná představitelka Marcia Zengová z Kentské státní univerzity, kde se věnuje výuce oboru Organizace znalostí, zveřejnila návrh rozpracované typologie v červnu 2000 na webu NKOS²⁷. Modifikovaná verze návrhu této typologie SOZ byla publikována také v roce 2008 v rozsáhlé analytické studii věnované kompletně SOZ²⁸. Zengová do typologie zahrnuje celkem 14 typů SOZ (o tři více než Hodgeová – viz tab. 1). Podstatná jsou ale dvě zvolená základní hlediska, na jejichž základě typy SOZ vymezuje. Jsou jimi struktura systému a funkce systému. V rámci nich rozčlenila typy SOZ do 4 skupin (uspořádanost uvnitř skupin je dána principem od jednoduchého ke složitějšímu co do struktury a od nízkého k vysokému co do funkcí). Typologie M. Zengové je znázorněna ve studii také dvojrozměrným grafickým schématem²⁹ (bývá často citováno). Stejně jako u Hodgeové je první skupina nazvaná „*Seznamy termínů*“ (*Term Lists*). Jsou do ní zahrnuty systémy s jednoduchou strukturou a s minimálními funkcemi (zejména zjednoduštění významů a kontrola synonym). Patří k nim různé jednoduché *seznamy* (vybraných termínů), *slovníky* (angl. dictionaries), *glosáře* a nově také *seznamy či soubory synonym* (angl. Synonym Rings). Do druhé skupiny nazvané „*Metadatové*

²⁶ Prezentace skupiny a výsledky činnosti jsou dostupné na jejím webu: <http://nkos.slis.kent.edu/>.

²⁷ ZENG, Marcia Lei. Taxonomy of knowledge organization sources / systems. In: *Networked Knowledge Organization Systems/Services: NKOS* [online]. Kent: School of Library and Information Science, Kent State University, Draft June 7, 2000, revised July 31, 2000 [cit. 2014-08-25]. Dostupné z: http://nkos.slis.kent.edu/KOS_taxonomy.htm.

²⁸ ZENG, Marcia Lei. Knowledge Organization Systems (KOS). In: *Knowledge organization*. 2008, 35(2-3), 160-182. ISSN 0943-7444.

²⁹ ZENG, Ref. č. 28, s. 161, obr. 1.

modely“ (*Metadata-like Models*) autorka zařadila systémy s vyšší mírou strukturovanosti a přidáními funkcemi (například již také ustanovení jednoduchých hierarchických vztahů). Zařazeny jsou především *autoritní soubory* (jmenné povahy), dále nově *adresáře* (angl. directories) se seznamy jmen a souvisejícími kontaktními údaji a také *zeměpisné slovníky*. Do třetí skupiny nazvané „*Klasifikace a kategorie*“ (*Classification and Categories*) jsou zařazeny systémy s vyšší až vysokou mírou hierarchických vztahů. Na první pozici řadí autorka tradiční *schémata předmětových hesel* a dále *kategorizační schémata*, *taxonomie* a *schémata klasifikační*. Čtvrtá skupina nazvaná pracovním „*Modely vztahů*“ (*Relationship Models*) zahrnuje stejně jako G. Hodgeová na první pozici *tezaury*, u nichž přibývá výrazný a cenný vztah asociační, a dále *sémantické sítě* a zejména rozrůstající se *ontologie*, které se vyznačují vysokou mírou asociativních a jiných vztahů mezi pojmy a také přítomností definovaných pravidel a axiomů.

Typologie SOZ M. Zengové, resp. Pracovní skupiny NKOS se stala východiskem návrhu oficiální typologie SOZ s názvem „*Slovník typů SOZ*“ (*KOS Types Vocabulary*)³⁰, která je připravována v rámci vytváření Metadatového aplikačního profilu pro registry slovníků SOZ³¹ novou Pracovní skupinou DCMI-NKOS. Ustanovení typologie je nezbytné pro naplňování prvku Typ SOZ (Type of KOS, jde o prvek specifikace Metadatových termínů DCMI <dcct:format>) stanovenými hodnotami. V názvu typologie se objevuje frekventovaný termín z oblasti IT „slovník“ (angl. vocabulary) z důvodu sladění terminologie se standardy iniciativy DCMI. Jde o čistě abecedně uspořádaný seznam (viz tab. 1) v tuto chvíli 16 typů SOZ („slovníků“) s jejich stručnými definicemi v původním znění typu systému (užito je v tomto případě jednotného čísla). Hierarchické uspořádání typů SOZ není uplatněno. Oproti výchozí typologii M. Zengové byl vyřazen typ „adresáře“, přibýly naopak 3 nové typy a některé typy byly upraveny v pojmenování. Do výčtu se dostaly následující typy systémů: 1. *kategorizační schéma*, 2. *klasifikační schéma*, 3. *slovník* (angl. dictionary), 4. *zeměpisný slovník*, 5. *glosář*, 6. *seznam* (termínů), 7. *seznam jmenných autorit* (upravené pojmenování Zengové přidáním slova „jmenných“), 8. *ontologie*, 9. *schéma* (jde o nově přidáný typ s významem schémat datových modelů, například pro taxonomie, ontologie aj.), 10. *sémantická síť*, 11. *schéma předmětových hesel* (upravené pojmenování Zengové přidáním slova „schéma“), 12. *seznamy synonym*, 13. *taxonomie*, 14. *terminologie* (jde o nově přidáný typ s významem systému, který obsahuje pojmy a jejich výklad v nějakém specializovaném oboru), 15. *tezaurus* a 16. *slovník* (angl. vocabulary; jde také o nově zařazený typ s významem jednoduchého či komplexního souboru „termínů“ pro zvláštní účely v oblasti IT, například v oblasti sémantického webu a propojených otevřených dat). Poslední zařazený typ je problémový, protože je zároveň užít pro označení všech dalších typů systémů (viz název celé typologie). Jde o střešní termín, který je v tomto případě synonymem termínu „systém organizace znalostí“. V českém kontextu je tu navíc problém mnohovýznamovosti pojmu „slovník“.

³⁰ Dublin Core Metadata Initiative. NKOS Task Group. NKOS Vocabularies. 2., KOS Types Vocabulary. In: *DCMI NKOS Task Group* [online]. 2013, updated 2013-12-16. Dostupný z: http://wiki.dublincore.org/index.php/NKOS_Vocabularies#KOS_Types_Vocabulary.

³¹ 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. 2014-08-25]. Dostupné z: http://wiki.dublincore.org/index.php/NKOS_AP_Elements.

Hlavní (jádrové) typy SOZ (viz tab. 1) ze seznamu pracovní skupiny DCMI-NKOS jsou i s definicemi kodifikovány také v nové mezinárodní normě ISO 25964-2:2013³², která se týká interoperability tezaurů s jinými „slovníky“ (norma užívá jenom tento termín, popř. termín „řízený slovník“). Typologie SOZ není v normě přímo obsažena, ale v kontextu výkladu reálií týkajících se tezaurů prezentuje jiné související „slovníky“ (v kapitolách 17 až 24). Několik typů v normě definováno není ani v úvodním seznamu termínů (kategorizace, slovník (angl. dictionary), zeměpisný slovník, glosář, seznam a sémantická síť). V seznamu termínů jsou ale naopak krátce definovány termíny „řízený slovník“ a dokonce „strukturovaný slovník“ (controlled vocabulary, structured vocabulary – oba typy jsou pracovní uvedeny v tab. 1 na stejné řádce jako typ „vocabulary“ z typologie skupiny DCMI-NKOS, i když nejsou definovány totožně) a „datový model“ (data model – tento typ je v typologii skupiny DCMI-NKOS uveden jako „schéma“). Citovaná norma i seznam skupiny DCMI-NKOS vznikaly přibližně ve stejné době (na tvorbě normy se navíc podíleli někteří odborníci skupiny DCMI-NKOS včetně M. Zengové), takže termíny a zejména definice z normy byly, vedle jiných, přebírány také do definic uvedených v seznamu skupiny DCMI-NKOS.

Jak je vidět z tab. 1, nejrozsáhlejší typologii SOZ prozatím předložili ve své nejnovější studii R. Souza, D. Tudhope a M. Almeida³³. Na základě analýzy a vyhodnocení do té doby prezentovaných podstatných typologií SOZ v literatuře včetně mnoha vlastních předchozích příspěvků připravili (též formou pojmové mapy) jinou zajímavou typologii těchto systémů. Prezentované typy jsou uspořádány hierarchicky, dokonce na více úrovních, a tato uspořádanost je dána opět hlediskem struktury systémů, jejich funkcí a také typem komunit užívajících SOZ. Do první skupiny nazvané „*Nestrukturované texty*“ jsou zařazeny systémy nestrukturovaných textů, jako jsou *abstrakta* a další *náhrady úplných textů* (Surrogates). Větší počet typů je uveden v rámci druhé skupiny nazvané „*Seznamy termínů a/nebo pojmů*“. Dominují v ní *autoritní soubory* (seznamy jmen autorů, seznamy názvů děl, seznamy kódů zemí, jazyků aj.), zařazeny jsou ale i *slovníky* (dictionaries), *zeměpisné slovníky*, *glosáře* a *seznamy synonym*. Navíc jsou pak začleněny také *folksonomie*, *seznamy tagů* a *konkordanční tabulky*. Posledním zařazeným typem jsou *řízené slovníky*, které ale zároveň patří i do skupiny třetí nazvané „*Struktury pojmů a vztahů*“. Do této skupiny autoři zařadili *klasifikační schémata*, *kategorizační schémata*, *předmětové hesláře*, *taxonomie* (různých dílčích typů), *tezaury*, *sémantické sítě* a *ontologie*, navíc pak doplnili také *rejstříky* (indexy) *rešeršních systémů* (v rámci databází), *datové slovníky* a *rámce* (angl. Frames). Poslední skupina nazvaná „*Struktury pojmů, vztahů a úpravy vzhledu*“ obsahuje ryze specifické typy SOZ. Zařazeny jsou *myšlenkové mapy*, *argumentační mapy*, *pojmové mapy* a *obohacené obrázky* (Rich Pictures), dále *datové modely* (například *podnikové datové modely*), *modely entit a jejich vztahů* (například *konceptuální datová schémata* aj.) a četné dílčí druhy referenčních modelů (například *obchodní referenční modely* nebo *technické referenční modely* anebo i kombinace s předchozím typem – *datové referenční modely*). Analytická studie je uzavřena zajímavým novým návrhem „taxonomie“ všech dimenzí (hledisek) rozdělování SOZ.

³² ISO 25964-2:2013. *Information and documentation – Thesauri and interoperability with other vocabularies – Part 2: Interoperability with other vocabularies*. 1st ed. Geneva: International Organization for Standardization, 2013-03-04. 99 s.

³³ SOUZA, Renato Rocha, TUDHOPE, Douglas, ALMEIDA, Maurício Barcellos. Towards a taxonomy of KOS: dimensions for classifying knowledge organization systems. In: *Knowledge organization*. 2012, 39(3), 179-192. ISSN 0943-7444.

Tab. 1 Přehled vybraných typologií SOZ (položky jsou uvedeny v původním znění a v abecedním pořádku podle prvního sloupce; význam barevných pozadí je uveden v závěrečné části 2.3)

	DCMI-NKOS, 2013-	ISO 25964-2, 2013	Souza, Tudhope aj., 2012	Zeng, 2008	Hodge, 2000	BARTOC, 2013-	VEST (AIMS), 2012-	TaxoBank, 2009-
category scheme			Abstracts	Argument Maps	Category Schemes	Category Schemes	Classification Schemes	Classification Schemes
classification scheme		classification scheme	Category schemes	Classification systems	Classification Schemes	Classification Schemes	Classification Scheme / System	classification system
			Concept Maps	Concept Maps				concept map
dictionary			Concordance Lines	Concordance Lines				
			Data Dictionaries	Directories	Directories	Directories	Dictionary	
			Dictionaries	Dictionaries	Dictionaries	Dictionaries	Dictionary	
			Entity-Relationship Models	Entity-Relationship Models			Encyclopedia	
			Folksonomies	Folksonomies				
gazetteer			Frames	Frames				
glossary			Gazetteers	Gazetteers	Gazetteers	Gazetteers	Glossary	gazetteer
			Glossaries	Glossaries	Glossaries	Glossaries	Glossary	glossary
			IR Indexes	IR Indexes				
list			Lists (Pick lists)	Lists (Pick lists)			Lexical database	
name authority list		name authority list	Mind Maps	Mind Maps				
ontology		ontology	Authority files (Aut., Ttl., Codes)	Authority Files	Authority Files	Authority Files	Authority file / list	ontology
			Ontologies	Ontologies	Ontologies	Ontologies	Ontology	ontology
			Reference Models	Reference Models				
schema		data model	Rich Pictures	Rich Pictures				
semantic network			Data Models	Data Models				
subject heading scheme		subject heading scheme	Semantic networks	Semantic Networks	Semantic Networks	Semantic Networks	Subject Headings	subject headings
			Subject headings	Subject Headings	Subject Headings	Subject Headings	Subject Headings	subject headings
			Surrogates	Surrogates				
synonym ring		synonym ring	Synonym rings	Synonym Rings	Synonym Rings	Synonym Rings		
taxonomy		taxonomy	Tag Lists	Tag Lists				
terminology		terminology	Taxonomies	Taxonomies	Taxonomies	Taxonomy	Taxonomy	taxonomy
thesaurus		thesaurus	Thesauri	Thesauri	Thesauri	Thesaurus	Thesaurus	thesaurus
			Topic maps	Topic maps			Topic Tree	
vocabulary		controlled vocabulary	Controlled vocabularies	Controlled vocabularies	Controlled vocabularies	Controlled Vocabulary	Controlled Vocabulary	controlled vocabulary
		structured vocabulary						

2.2 Typologie systémů organizace znalostí v online registrech

Pro srovnání je v této části podán analytický komentář k typologiím SOZ, které byly připraveny v rámci „živých“ online dostupných databází registrů těchto systémů. Sledován je praktický či pragmatický přístup tvůrců registrů k otázce typologie či kategorizace SOZ. Vybrány byly tři registry (viz také tab. 1), které vznikly v rozmezí let 2009–2013 a stále se rozvíjejí. Jeden je oborově profilovaný, další dva jsou univerzálního charakteru.

Registrační systém *VEST (Vocabularies, mEtadata Sets and Tools)*³⁴ je novým online registrem systémů pro komplexní správu a řízení informací z oboru zemědělství a příbuzných oborů včetně SOZ (v registru je užit v kontextu daných technologií taktéž termín „slovník“ (vocabulary)). Byl vytvořen a je spravován odbornou skupinou světové organizace FAO (Food and Agriculture Organization of the United Nations) a je zpřístupňován v rámci komplexního portálu AIMS (Agricultural Information Management Standards). K srpnu 2014 registr obsahoval 180 záznamů. Jeho specifikem je, že vedle „slovníků“ (136 záznamů) registruje zvlášť také metadatové soubory (61 záznam) a také nástroje pro správu informací (62 záznamy). Všechny zaznamenané „slovníky“ jsou rozděleny podle typu a podle předmětové oblasti (jde zároveň o vyhledávací údaje). Typologie „slovníků“ je relativně bohatá (celkem 12 typů). Jde o (zahrnuta je také aktuální statistika a hodnotící poznámky): 1. *Autoritní soubor/seznam* (15 záznamů, zařazeny jsou ale nejenom soubory jmenných autoritních dat a seznamy kódů jazyků a zemí, ale také některé taxonomické databáze organismů nebo dokonce i věcná kategorizační schémata, např. předmětová kategorizace AGRIS (Agricultural Information System)), 2. *Klasifikační schéma/systém* (10 záznamů), 3. *Slovník* (angl. dictionary, 7 záznamů, vedle různorodých druhů slovníků jsou ale zařazeny i některé seznamy synonym nebo dokonce známý ontologický systém UMLS (Unified Medical Language System)), 4. *Encyklopedie* (zaznamenána je zatím jediná Encyklopedie života, <http://eol.org/>), 5. *Glosář* (16 záznamů), 6. *Lexikální databáze* (prozatím ale není zaznamenána žádná), 7. *Ontologie* (18 záznamů řady úspěšných ontologií v tomto oboru), 8. *Předmětová hesla* (zatím pouze 2 záznamy), 9. *Taxonomie* (14 záznamů pro daný obor významných taxonomických systémů), 10. *Terminologická databáze* (obsahuje 10 záznamů různých výkladových slovníků odborných termínů, též například databázi „Geonames“), 11. *Tezaurus* (40 záznamů, nejpočetnější typ v daném registru, zahrnuty jsou i některé vzdáleně příbuzné tezaury, například Tezaurus ERIC (Education Resources Information Center), zatím ale chybí i některé podstatné, například Tezaurus NAL (National Agricultural Library)) a 12. *Stromy témat* (registr je ojedinělý v zařazení tohoto typu, zatím také neobsahuje žádný záznam).

Zajímavé parametry vykazuje univerzálně profilovaný registr SOZ s názvem „*TaxoBank*“³⁵, který je v provozu od roku 2009 a spravuje ho odborný tým „taxonomistů“ americké společnosti „Access Innovations“ specializované na softwarové aplikace se zaměřením na lingvistické aspekty dat ukládaných v databázích (harmonizace dat

³⁴ Food and Agriculture Organization of the United Nations. *Agricultural Information Management Registry of Vocabularies, mEtadata Sets and Tools. VEST Registry. Vocabularies* [online]. Rome: FAO, Office of Knowledge, Exchange, Research and Extension, © 2012 [cit. 2014-08-25]. Dostupné volně z Portálu AIMS: <http://aims.fao.org/vest-registry>.

³⁵ Access Innovations. *TaxoBank Terminology Registry: TaxoBank* [online]. Albuquerque (New Mexico, USA): Access Innovations, 2009- [cit. 2014-08-25]. Dostupné z: <http://www.taxobank.org/>.

apod.). Registr obsahuje velmi podrobné a kvalitně strukturované údaje o „řízených slovnících“ (užit je i v tomto případě právě tento termín namísto termínu „systém organizace znalostí“). Databáze zatím obsahuje 240 zaznamenaných systémů. V registru TaxoBank je prozatím ustanoveno 9 typů „slovníků“ (Vocabulary type) – viz tab. 1. Jde o (v závorce jsou na první pozici uvedeny zkratky typů užívané v popisu, zahrnuta je i statistika a poznámky): 1. *Klasifikační systém (clsssystem, zatím 9 záznamů, zahrnuta jsou známá schémata, ale také například „Taxonomie organismů NCBI“)*, 2. *Pojmová mapa (concmp, zatím jenom 2 záznamy)*, 3. *Řízený slovník (contrvoc, zahrnuje zatím 29 záznamů, jde o druhý nejpočetnější typ; pojmenování tohoto typu se ale shoduje s pojmenováním všech zahrnutých typů; při bližším prozkoumání uživatel pod tímto typem najde i různé tezaury, například Tezaurus DTIC (Defense Technical Information Center), seznamy předmětových termínů, specifické terminologické glosáře, některé klasifikace, například Klasifikaci INSPEC (INformation Service in Physics, Electrical & electronic, and Computer control), nebo i bibliografickou Antropologickou databázi)*, 4. *Zeměpisný slovník (gaz, zatím pouze 2 záznamy)*, 5. *Glosář (glos, zatím 3 záznamy)*, 6. *Ontologie (ont, zatím 3 záznamy, uvedena je například „Ontologie DBpedie“)*, 7. *Předmětová hesla (subjh, zatím pouze 6 záznamů)*, 8. *Taxonomie (taxon, zatím nevelká skupina 15 záznamů známějších i méně známých taxonomií)* a 9. *Tezaurus (thes, nejpočetnější skupina cca 140 záznamů známých i méně známých tezaurů)*. Registr TaxoBank je doplňován nepravidelně a náměty k registraci zasílají převážně uživatelé internetu. Systém zatím nemá zcela efektivní rešeršní rozhraní (chybí například možnost vyhledávat podle typů slovníků). Předností registru je kvalitně strukturovaný záznam, který bývá často velmi bohatý na údaje, ocenit lze zejména přítomnost podstatného údaje o typu zobrazení dat, a to ve zkrácených výrazech (např. „hier“ = hierarchický, alph = abecední, perm = permutovaný, other = jiný), nebo velmi cenný údaj o typech vztahů (užívá se zkrácených hodnot, např. eq_pri_eq = ekvivalence, eq_lang = jazyková ekvivalence, hier_bn = rododruhová hierarchie, hier_inst = třída instance, rel_t = asociace, othr = ostatní).

Posledním reprezentantem registrů SOZ je nový systém „BARTOC“ (*BA*sel *R*egister of *T*hesauri, *O*ntologies & *C*lassifications)³⁶. Je v provozu od roku 2013 a je spravován na serveru Univerzitní knihovny v Basileji. V srpnu 2014 zaznamenal již 667 záznamů registrovaných systémů, což je do značné míry dáno iniciativním přístupem správce báze (A. Ledl). I v případě tohoto registru jsou registrované jednotky často označovány termínem „řízené a strukturované slovníky“ (angl. controlled and structured vocabularies), nebo jenom jednoduše „slovníky“. To je jistý problém, jelikož i jeden dílčí typ nese označení „řízený slovník“. Služba BARTOC ustanovila pragmaticky pouze 5 typů systémů organizace znalostí (slovníků). Čtyři jsou známé a všude jinde uznávané hlavní typy těchto systémů (tezaury, klasifikace, ontologie a taxonomie), pátým typem jsou již zmíněné „řízené slovníky“ (do tohoto typu je zařazen například německý autoritní soubor GND (Gemeinsame Normdatei), ale i specifikace metadat DCMES). Jiný významný typ SOZ (schémata předmětových hesel) služba vůbec nevede, a tak jsou významné předmětové hesláře (MeSH, LCSH aj.) jednoduše přiřazeny k tezaurům. Většine záznamů je přidělena pouze jedna kategorie, v některých případech je uveden ještě jeden další typ. Aktuální statistika sděluje, že nejvíce systémů (cca 360) patří do skupiny *tezaurů*. Druhou příčku obsazují *klasifikace* (cca 185 systémů). Na třetím místě jsou *ontologie*

³⁶ Universitätsbibliothek Basel. *BARTOC.org: BA*sel *R*egister of *T*hesauri, *O*ntologies & *C*lassifications [online]. Projektleiter Andreas LEDL. Basel: Universitätsbibliothek Basel, 2013- [cit. 2014-08-25]. Dostupné volně ze serveru Basilejské univerzity: <http://www.bartoc.org/>.

(cca 65 systémů), na místě čtvrtém „řízené slovníky“ (cca 33 systémů) a na místě pátém *taxonomie* (cca 22 systémů). Veřejný přístup k registru zatím neposkytuje v rámci rešeršního rozhraní mnoho možností pro přesné či detailní vyhledávání záznamů SOZ (chybí například právě vyhledávání podle typu „slovníku“). Záznamy obsahují relativně dostatečný počet metadat, specifikem a velkou předností systému je doplňování třídniců v německy hovořících zemích dnes velmi používaného Deweyho desetinného třídění (až na třetí hierarchickou úroveň) a také deskriptorů z vícejazyčného tezauru EUROVOC.

2.3 Výsledek porovnání typologií SOZ

Komparace pěti vybraných typologií SOZ dokumentovaných v teoretické literatuře a standardech a tří vybraných typologií aplikovaných v reálných provozech nových registrů SOZ ukázala, že existuje – po letech diskusí – určité **širší jádro těchto typů** a okolo něho specifické typy SOZ, které se vyskytují spíše ojediněle a bývají diskutabilní. V naší komparativní tabulce (viz tab. 1) jsou v řádcích s oranžovým pozadím uvedeny typy SOZ, na kterých se shodují jak teoretici, tak praktici, tj. vyskytly se ve všech osmi typologiích (pěti teoretických + třech praktických). Jde o nejuzší jádro velmi dobře známých a uznávaných typů SOZ, k němuž patří (upřednostněna jsou pojmenování daná návrhem skupiny DCMI-NKOS, která by měla být schválena jako standard): *klasifikační schémata*, *ontologie*, *taxonomie* a *tezaury*. K nim lze připojit další typy, které tvoří širší jádro typů SOZ (vyskytly se 7 až 3krát). Jde o systémy, jež jsou v komparativní tabulce uvedeny na světle žlutém pozadí. K těmto typům patří: *schémata předmětových hesel* (7 výskytů, tj. 5 teoretických + 2 praktické), *seznamy jmenných autorit* (6 výskytů, tj. 5+1), *glosáře* (6 výskytů, tj. 4+2), *zeměpisné slovníky* (5 výskytů, tj. 4+1), *slovníky (dictionaries)* (5 výskytů, tj. 4+1), *slovníky/řízené slovníky/strukturované slovníky* (celkem 5 výskytů, tj. 3+2), *kategorizační schémata* (4 výskyty, tj. 4+0), *seznamy synonym* (4 výskyty, tj. 4+0), *sémantické sítě* (4 výskyty, tj. 4+0), *schémata/datové modely* (celkem 3 výskyty, tj. 3+0) a *terminologie* (3 výskyty, tj. 2+1). Typy SOZ se dvěma výskyty (jsou uvedeny na světle zeleném pozadí) lze k jádru po zvážení přiřadit také, jde o *pojmové mapy* (1+1) a *seznamy* (2+0). Ostatní typy SOZ s 1 výskytem je možné prozatím označit za okrajové.

Srovnávací analýza typologií SOZ poukázala na problém nejednotného užívání určitých termínů pro pojmenování typů SOZ v různých oborech či uživatelských komunitách (knihovnické, počítačové aj.), ale také na určité nesrovnalosti v chápání významu takových termínů uvnitř jedné typologie, jak bylo naznačeno také v krátkých komentářích části 1. Analýza ukázala, že termín „systém organizace znalostí“ je v určitém pohledu synonymní s termínem „slovník“ (angl. vocabulary). Největší problém představuje zaměňování termínu „slovník“ (angl. vocabulary) nebo „řízený slovník“ (angl. controlled vocabulary) jako termínu určeného pro označení skupiny více dílčích typů SOZ s termínem určeným pro označení jediného typu určité typologie. Konkrétní provozy již existujících registrů SOZ též ukazují reálné příklady i nekorektního nebo problémového zařazování některých SOZ k neodpovídajícím typům.

3 Návrh vlastní typologie systémů organizace znalostí

Předpokladem použitelné typologie je jasné ohraničení členěného celku. V případě systémů organizace znalostí jsme se rozhodly pro vymezení dané nutnými podmínkami, formulovanými v části 1: musí se jednat o záměrně vytvořený nástroj organizace

znalostí a zároveň o systém, jehož struktura odpovídá charakteristice jazyka a základní strukturální jednotkou je pojem. Tím jsme oddělily okrajové a problematické typy, například: soubory volně tvořených klíčových slov a tagů, fulltextové indexy, seznamy morfologických tvarů, jazykové slovníky, abstrakty, encyklopedie, datové a pojmové modely, pojmové mapy i „pouhé“ jazyky (např. RDF). Pro systémy, jež vyhověly všem nutným podmínkám a jež jsme označily jako jádro, jsme se pokusily navrhnout vlastní typologii.

V části 1.1 bylo konstatováno, že systémy organizace znalostí jsou velmi komplikované informační zdroje. Nabízí se tudíž značné množství kritérií členění, použitelných při konstrukci potenciální typologie: *jednotka organizace* (dokument, vědní obor, osoba, instituce, produkt, služba, místo, proces, pojem...), *sémantická síla*, tj. schopnost struktury SOZ reprezentovat význam a složitost (množství vlastností a vztahů), *doména*, jež může být univerzální či polytematická nebo může pokrývat speciální téma, *způsob reprezentace znalostí* (enumerativní či fasetový, prekoordinovaný nebo postkoordinovaný), *typ slovníku* (řízený, volný, jedno- či vícejazyčný), *otevřenost/uzavřenost* (tzv. předpoklad otevřeného či uzavřeného světa), *granularita* (specifičnost), *formát, účel* (funkčnost, uživatelské určení – např. kategorizace se spíše hodí pro listování či prohlížení, tezaury pro vyhledávání).

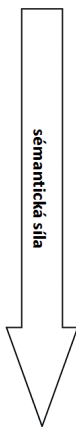
Mezi těmito kritérii se jako nosná jeví zejména první dvě: jednotka organizace a sémantická síla. Jednotka organizace vyjadřuje, co se organizuje. Podle tohoto kritéria se SOZ může stát nástrojem organizace věcí, organizace pojmů či organizace termínů. Zajímavá je zejména skupina věnovaná organizaci pojmů, kterou je možné dále členit na podskupinu organizující „čisté znalosti“ např. v rámci klasifikace věd nebo biologické taxonomie, a na podskupinu tradičně patřící do oblasti zkoumání informační vědy, která je zaměřena na organizování tzv. *aboutness*, tj. pojmů vyjadřujících výsledky obsahové analýzy informačních zdrojů.

Kritérium sémantické síly jsme se rozhodly využít jako základní pro návrh vlastní typologie SOZ. Vycházíme jednak ze závěrů analýzy stávajících významných typologií v předchozí části této studie, jednak z dosavadních zkušeností při vkládání informací o systémech organizace znalostí do prototypu znalostní báze v rámci výzkumného projektu. Předpokládáme dva účely (případy užití) této typologie: 1) kategorizace registrovaných jednotek ve znalostní bázi, 2) terminologická platforma pro odborné diskuse a další teoretický výzkum.

Pro první případ užití při *kategorizaci zdrojů* ve znalostní bázi bude dobře použitelná škála, jež v současné době zahrnuje 16 typů. V tabulce 2 jsou jednotlivé typy uvedeny v pořadí jejich „složitosti“, resp. sémantické síly. Přehled začíná nejjednodušším typem – seznamy slov, a uzavírají ho ontologie – nástroje umožňující reprezentovat nejbohatší strukturu obsahu a vztahů mezi nimi. S určitým zjednodušením by se daly vztahy postupného narůstání komplexnosti mezi skupinami označenými písmeny vyjádřit následovně: skupina A = seznam slov; skupina B = A + definování významu slov; skupina C = B + seskupení významově příbuzných slov (ekvivalence); skupina D = C + určení preferovaných termínů; skupina E = D + určení hierarchických vztahů; skupina F = E + určení asociativních vztahů; skupina G = F + možnost odvozování, usuzování. V tabulce 2 je dále vyznačeno, že skupiny A – C se zaměřují na organizaci slov (termínů), skupiny D – G jsou určeny k organizaci na pojmové úrovni. Typy SOZ uvedené pod čísly 10–16 zhruba odpovídají množině označované v tuzemském kontextu jako „věcné autority“. Předpokládá se, že spolu s tím, jak stoupá vyjadřovací potenciál organizačního systému, stoupá i míra přesnosti a úplnosti vyhledávání realizovaného s jeho pomocí. Zároveň však vzrůstá i obtížnost a pracnost konstruování takového systému.

Tab. 2 Typologie systémů organizace znalostí podle sémantické síly

kategorie	typ	anglicky			
slovníky lineární struktura	organizace termínů	1 nabídkový seznam (menu)	<i>pick list</i>	A	seznam slov
		2 negativní slovník (stop slova, zakázaná slova)	<i>stopwords, exclusion lists</i>		
		3 (řízený) slovník	<i>(controlled) vocabulary, dictionary</i>		
		4 číselník (nomenklatura, seznam, kódovník)	<i>list, code list</i>	B	A + definice
		5 (výkladový) slovník, glosář	<i>glossary</i>		
		6 terminologický slovník	<i>terminology</i>		
		7 seznam synonym	<i>synonym ring, synset</i>		
klasifikace hierarchická struktura	organizace pojmů	8 zeměpisný slovník	<i>gazetteer</i>	D	C + preferované termíny
		9 seznam jmenných autorit	<i>name authority list/file</i>		
		10 předmětový heslář	<i>subject heading scheme</i>	E	D + hierarchie
		11 kategorizace, kategorizační schéma / systém	<i>categorization scheme</i>		
		12 taxonomie (biologie)	<i>taxonomy</i>		
13 klasifikace / klasifikační schéma / systém	<i>classification scheme</i>	F	E + asociace		
14 tezaurus	<i>thesaurus</i>				
15 sémantická síť	<i>semantic network</i>				
pojmové sítě síťová struktura		16 ontologie	<i>ontology</i>	G	F + odvozování



Pro druhý případ užití, tj. *terminologický systém* pro odborné diskuse a výzkum, se v praxi osvědčily dichotomie, limitem počtu kategorií (prezentovaných prvků) pro tyto účely je zřejmě psychologická hranice 7 ± 2 ³⁷ prezentovaných prvků. Proto tedy považujeme za vhodné mít k dispozici ještě další úroveň pohledu na SOZ, jež by jednotlivé typy z naší škály sjednotila do větších skupin. Výše zmíněný článek věnovaný terminologii organizace znalostí poukázal na neudržitelnost dichotomie předmětové – systematické, která již neodpovídá současné realitě. Je ovšem zřejmé, že nějakou „meziúroveň“ uvažování o SOZ potřebujeme. Prozatím používáme mírně upravenou trojici kategorií G. Hodgeové založenou na typologii struktur a vyjadřovaných vztahů – *slovníky* s jednoduchou lineární strukturou, hierarchicky strukturované *klasifikace* a *pojmové sítě*.

Dosavadní zkušenosti ukazují, že ani existující typologie není zárukou jednotného přístupu. Autoři úvodníku k monotematickému číslu časopisu *Applied ontology*, věnovanému vztahu ontologií a terminologických systémů³⁸, názorně doložili problémy se zařazováním jednotlivých instancí do stanovených typových kategorií. Provedli rozsáhlý literární průzkum a na příkladu označování sedmi konkrétních SOZ v odborných textech ukázali na stávající nejednotnost. U systému *WordNet* zaznamenali 10 případů jeho zařazení do kategorie lexikální databáze, 12x byl označen jako (lingvistická) ontologie, 1x jako folksonomie, 2x byl zařazen do více kategorií. Systém *Foundational model of anatomy (FMA)* byl 7x označen jako ontologie a 2x jako terminologie. Systém *Medical subject headings (MeSH)* byl 6x označen jako terminologie nebo tezaurus, 6x jako

³⁷ MILLER, George Armitage. The magical number seven, plus or minus two: some limits on our capacity for processing information. In: *Psychological review*. 1956, **63**(2), 81-97. ISSN 0033-295X (Print). ISSN 1939-1471 (Online).

³⁸ GRABAR, Natalia, HAMON, Thierry, BODENREIDER, Olivier. Ontologies and terminologies: continuum or dichotomy? In: *Applied ontology*. 2012, **7**(4), 375-386. ISSN 1570-5838 (Print). ISSN 1875-8533 (Online).

ontologie, 5x byl sice nazván ontologií, ale byl popisován jako tezaurus nebo řízený slovník. Systém AGROVOC byl 8x označen jako tezaurus nebo strukturovaný řízený slovník, 3x jako ontologie a jednou jako tezaurus přepracovaný na ontologii. Tezaurus *NCI* byl 2x označen jako terminologie nebo tezaurus, 6x jako ontologie a 4x současně jako tezaurus i ontologie. Systém *SNOMED CT* byl v 6 případech zařazen jako terminologie, 5x jako ontologie, 4x současně jako terminologie i ontologie. Systém *UMLS (Unified Medical Language System)* byl 9x považován za metatezaurus nebo oborově specifický terminologický systém, 5x za ontologii, 3x za obojí. Pro názornější představu o jednotlivých typech SOZ zahrnutých do naší typologie uvádíme proto jejich stručné charakteristiky (pořadové číslo se shoduje s pořadovým číslem v tabulce 2) a příklady jejich instancí.

1. *Nabídkový seznam* (v terminologii uživatelského rozhraní softwaru *list box* nebo *combo box*, tj. rozbalovací seznam) lze najít prakticky v každém online systému na webu. Příkladem může být seznam typů zdrojů „Dokumentart“ nabízený v rozhraní pro pokročilé vyhledávání nové metadatové služby BASE (<http://www.base-search.net/Search/Advanced>), seznam fondů k vyhledávání „Sources“ národní nizozemské služby pro vědu a výzkum NARCIS (<http://www.narcis.nl/>) nebo řada dílčích nabídek vyhledávacích údajů v pokročilém vyhledávání digitálního archivu NASA (<http://ntrs.nasa.gov/advSearch.jsp>).
2. *Negativní slovník* obsahuje seznam slov vyřazených z nějakého typu zpracování (typicky z vyhledávacích procedur). Příkladem může být slovník „Stopwords“ systému Web of Science z roku 2009 (http://images.webofknowledge.com/WOK46/help/WOS/ht_stopwd.html) nebo negativní slovník systému Ranks NL Webmaster Tools (<http://www.ranks.nl/stopwords/czech>).
3. *Řízený slovník* je slovník s předem definovaným rozsahem (obvykle stanoveným výčtem termínů) a zpravidla i definovaným obsahem termínů, zajišťující konzistentní označování pojmů. Příkladem může být 1. část licenčního ujednání (Definice) na webu „Creative Commons“ (<http://creativecommons.org/licenses/by-nc-sa/3.0/cz/legalcode>). Dalším příkladem užití řízeného slovníku jsou seznamy termínů pro popis obrázků, např. „Key words (enhancement terms) for digital newsphoto archives“ (<http://www.ibiblio.org/slanews/archiving/terms/photowords.htm>), zpracované pracovní skupinou pro archivaci fotografií sekce novin americké Asociace speciálních knihoven (SLA). Obdobné slovníky jsou na komerční bázi nabízeny i k implementaci do grafických programů (např. The Controlled Vocabulary Keyword Catalog (CVKC), <http://www.controlledvocabulary.com/products/>).
4. *Číselník* je uspořádaný seznam slovních popisů věcí či jevů spolu s číselnými nebo písmennými kódy, nevyjadřující hierarchické vztahy. Příkladem může být Číselník zemí ČSÚ (CZEM, [http://www.czso.cz/csu/klasifik.nsf//ciselnik_zemi_\(czem\)](http://www.czso.cz/csu/klasifik.nsf//ciselnik_zemi_(czem))), kódy jazyků uvedené v normě ISO 3166-3:2013, *Codes for the representation of names of countries and their subdivisions – Part 2: Code for formerly used names of countries* (2nd ed., 2013), dostupné komerčně online z: http://www.iso.org/iso/home/standards/country_codes.htm – zahrnuty jsou i volně dostupné ukázky.
5. *Výkladový slovník* obsahuje seznam slov spolu s informací o jejich významu. Příkladem může být glosář „Linked data glossary“ (HYLAND, B., ATEMEZING, G., PEN-DELETEON, M., SRIVASTAVA, B., ed., W3C, 2013, dostupné z: <http://www.w3.org/TR/ld-glossary>) nebo Glosář evropské soudní sítě (http://ec.europa.eu/civiljustice/glossary/glossary_cs.htm).
6. *Terminologický slovník* je slovník zahrnující terminologii určitého oboru. Příkladem je česká online databáze TDKIV (http://aleph.nkp.cz/F/?func=file&file_name=find-a&local_base=ktld) nebo online zpřístupněný slovník „ODLIS: Online Dictionary for Library and Information Science“ (REITZ, J. M., Santa Barbara, ABC-CLIO, 2014, dostupný z: http://www.abc-clio.com/ODLIS/odlis_A.aspx).

7. *Seznam synonym* představuje množiny synonymních termínů, z nichž každý lze použít pro označení určitého pojmu³⁹. Příkladem je Tezaurus jazyka českého: slovník českých slov a frází souznačných, blízkých a příbuzných (KLÉGR, A., Praha: Lidové noviny, 2007, ISBN 978-80-7106-920-1), systém WordNet (<http://wordnet.princeton.edu/>) nebo seznam gramatických synonym systému Web of Science (http://images.webofknowledge.com/WOKRS57B4/help/WOS/hs_spelling_terms.html).
8. *Zeměpisný slovník* je uspořádaným souborem informací o geografických entitách. Příkladem může být velmi známý „Getty Thesaurus of geographic names® Online“ (<http://www.getty.edu/research/tools/vocabularies/tgn/>) nebo kooperativně vytvářený soubor národních geografických autorit (<http://authority.nkp.cz/vecne-autority/soubor-geografickychautorit-1/>).
9. *Seznam jmenných autorit* je řízený slovník určený k zajištění jednotného pojmenování vymezeného okruhu entit v daném kontextu⁴⁰. Příkladem může být soubor jmenných autorit „Name Authority Headings“ Kongresové knihovny ve Washingtonu, dostupný online ze společného portálu všech autorit (<http://authorities.loc.gov/>), jeho strukturovaná verze ve formátu MARC 21 (Authorities) je dostupná licenčně z webu CDS LC (<https://classificationweb.net/>) a ve formě propojených otevřených dat ho lze získat z URI: <http://id.loc.gov/authorities/names>. Uvést lze také dnes nejvýznamnější mezinárodní soubor jmenných autorit VIAF (<http://viaf.org/>), který je nabízen i ve formě propojených otevřených dat (<http://viaf.org/viaf/data/>). K tomuto typu lze zařadit i významný a velmi rozsáhlý, komerčně dostupný číselník chemických prvků, látek a substancí americké Chemické abstraktové služby „CAS Registry Numbers“ (její derivát je online dostupný zdarma v rámci služby „Common Chemistry“, <http://www.commonchemistry.org/>).
10. *Předmětový heslář* je strukturovaný slovník obsahující termíny, které jsou k dispozici pro předmětové (věcné) indexování, plus pravidla pro jejich kombinování do prekoordinovaných řetězců termínů v případě potřeby⁴¹. Příkladem může být francouzský národní předmětový systém RAMEAU (dostupný v různých formátech z hlavního sídla, <http://rameau.bnf.fr/>, nebo ve struktuře propojených otevřených dat, <http://data.bnf.fr/liste-rameau>) nebo český Polytematický strukturovaný heslář z produkce NTK (<http://www.techlib.cz/cs/82897-polytematicky-strukturovany-heslar>).
11. *Kategorizační schéma* je seznam kategorií (skupin, tříd), sloužící k seskupování organizovaných entit na základě jejich příslušnosti k určité kategorii. Typickým příkladem může být kategorizace oborů „*Web of Science categories*“ společnosti Thomson Reuters užívaná v systému Web of Science (dostupná v nejnovější verzi komerčně v rámci portálu WOS, <http://apps.webofknowledge.com/>, nebo ve starší verzi z roku 2012 jako dílčí slovníky s výkladem kategorií pro soubory SCIE, http://ip-science.thomsonreuters.com/mjl/scope/scope_scie/, SSCI, http://ip-science.thomsonreuters.com/mjl/scope/scope_ssci/, a AHCI, http://ip-science.thomsonreuters.com/mjl/scope/scope_ahci/).
12. *Taxonomie* je hierarchická enumerativní klasifikace používaná především v oblasti přírodních věd (biologie, zoologie). Příkladem může být „Integrated taxonomic information system“ (<http://www.itis.gov/index.html>). V poslední době se termín často

³⁹ Výklad převzat z ISO 25964-2, Ref. č. 32, s. 13, čl. 3.79.

⁴⁰ Výklad zpracován podle ISO 25964-2, Ref. č. 32, s. 8, čl. 3.50.

⁴¹ Výklad převzat z ISO 25964-1, Ref. č. 20, s. 11, čl. 2.57.

používá i k označení klasifikace digitálních objektů, např. webových stránek. V tomto smyslu ho používá například systém Drupal ve svém modulu Taxonomy (<https://www.drupal.org/node/774892>).

13. *Klasifikační schéma* je seznam pojmů a prekoordinovaných kombinací pojmů, uspořádaný do tříd⁴². Nejčastěji se jedná o hierarchickou strukturu s menším počtem vrcholových (kořenových) tříd a s větším počtem hierarchických úrovní. Příkladem může být první tištěné vydání třídění DDT (DEWEY, M. *A classification and subject index for cataloguing and arranging the books and pamphlets of a library*, Amherst (Hampshire County, Mass.), 1876), jež je jako volné dílo dostupné také v digitalizované formě (<http://archive.org/details/classification00dewerich>, <http://catalog.hathitrust.org/Record/001163285>) nebo jeho nejnovější elektronická forma „Dewey-info“ (<http://dewey.info/>), poskytující data také ve formě propojených otevřených dat. Jiným příkladem může být ruský Rubrikátor VINITI dostupný volně na webu (<http://scs.viniti.ru/rubtree/main.aspx?tree=RV>).
14. *Tezaurus* je řízený a strukturovaný slovník, v němž jsou pojmy reprezentované termíny organizovanými tak, že jsou explicitně vyjádřeny vztahy hierarchie a asociace mezi pojmy a preferované termíny jsou doprovázeny odkazy na synonyma nebo kvazisynonyma⁴³. Příkladem může být „ERIC thesaurus“ (dostupný volně v rámci databáze ERIC z portálu: <http://eric.ed.gov/> nebo podstránky pro prohlížení Tezauru z: <http://eric.ed.gov/?ti=all>) nebo „AGROVOC Thesaurus“ (<http://aims.fao.org/standards/agrovoc>), dostupný také již ve struktuře propojených otevřených dat (<http://aims.fao.org/standards/agrovoc/linked-open-data>).
15. *Sémantická síť* je grafické znázornění sémantických vztahů mezi pojmy. Příkladem může být „OpenStreetMap Semantic Network“ (http://wiki.openstreetmap.org/wiki/OSM_Semantic_Network) nebo sémantická síť „UMLS Semantic Network“ (<http://semanticnetwork.nlm.nih.gov/>), zobrazující vztahy biomedicínských pojmů.
16. *Ontologie* je sdílená a opakovaně použitelná pojmová reprezentace vymezené domény, zpravidla doplněná axiomy a ontologickými závazky; kromě organizace a vyhledávání znalostí slouží ke komunikaci, opakovanému využití existujících znalostí a k automatickému odvozování nových znalostí. Příkladem všeobecně zaměřených ontologií jsou „DBpedia Ontology“ (<http://wiki.dbpedia.org/Ontology?v=zj4>) nebo „BFO – Basic formal ontology“ (<http://ifomis.uni-saarland.de/bfo/>) či „YSO – General Finnish ontology“ (<http://finto.fi/yso/en/>), příkladem doménově specifické ontologie je genová ontologie „GO – Gene ontology“ (<http://geneontology.org/>).

Závěr

Dosavadními výsledky výzkumu systémů organizace znalostí v rámci projektu programu NAKI „Znalostní báze pro obor organizace informací a znalostí“ jsou pracovní definice pojmu systém organizace znalostí a návrh typologie SOZ jako podklad pro navrhované přijetí tohoto termínu do české odborné terminologie. V navržené typologii SOZ jsme se pokusily o syntézu přístupů zjištěných v provedené analýze za použití jednotného kritéria sémantické síly, jež by mělo zajistit konzistenci typologie. Jak v případě definice, tak zejména v případě navržené typologie nepovažujeme řešený problém za

⁴² Výklad zpracován podle ISO 25964-1, Ref. č. 20, s. 2, čl. 2.6.

⁴³ Výklad převzat z ISO 25964-1, Ref. č. 20, s. 12, čl. 2.62.

uzavřený, jde o návrh otevřený další diskusi. Spíše než odpovědi nabídly výsledky dosavadního výzkumu nové otázky.

Problémy s nejednotným pojetím pojmové i terminologické základny systémů organizace znalostí jsou patrné nejen na mezioborové úrovni (je například výrazný rozdíl mezi komunitou informační vědy a počítačové vědy), ale i v rámci oborů samotných, což se bohužel týká i informační vědy. Lze konstatovat, že volání po sjednocení terminologie, které se ozývalo už na konci 20. století, zůstalo dodnes nevyslyšeno. Stížnost B. Weinbergové na „závažné nedostatky kontroly slovníku v literatuře o řízených slovnících“⁴⁴ je dodnes aktuální. To může na jedné straně vést ke skeptické úvaze, zda je s takovými nástroji opravdu dosažitelná vize sémantického webu a (sémanticky) spojených dat. Na druhé straně se nabízí otázka, zda zjištěná nejednotnost v chápání systémů organizace znalostí a jejich typů je opravdu chybou, není-li spíše adekvátním obrazem diverzifikované a neustále se dynamicky měnící reality světa znalostí. I současné systémy organizace znalostí jsou v pohybu – klíčovým pohybem je integrace původně distinktivních rysů, jež odlišovaly například klasifikace a předmětová hesla, do společného výskytu v rámci jednoho konkrétního systému (např. MESH, LCSH, AAT, DDC Relative index, fasetizace MDT).

V části 1 jsme poukázaly na širokou extenzi významu pojmu SOZ a konstatovaly jsme, že takový pojem lze definovat jen na velmi obecné úrovni a že vymezení kategorie tohoto typu zřejmě nebude možné postavit na množině společných vlastností. Tento problém se v různé intenzitě zopakoval i v případě vymezování kategorií SOZ pro účely jejich typologie. Takové relativně pevně stanovené kategorie, jako jsou tezaury, kde už se lze opřít o konsenzuálně přijatý standard ISO 25964, jsou spíše výjimkou. Uvědomujeme si, že „ospravedlnění“ každého typu v typologii by mělo spočívat nikoli v odlišném pojmenování, ale v přítomnosti distinktivních rysů. Závěry srovnávací analýzy stávajících typologií naznačují, že lingvistickým rozbohem se k jednotnému chápání pojmové základny systémů organizace znalostí nedopracujeme. Pokusily jsme se proto o klasické vymezení SOZ a jeho typů pomocí společných vlastností, zatím jsme jich však našly jen málo a jsou silně abstraktní. Domníváme se, že v dalším výzkumu bude vhodné se zaměřit na „nearistotelské“ principy kategorizace. Z přehledu alternativních přístupů ke kategorizaci, zpracovaného G. Lakoffem⁴⁵, se jako potenciálně užitečné jeví zejména: *rodinné podobnosti* (princip takto pojmenoval L. Wittgenstein, Lakoff ho interpretuje jako vzájemný vztah příbuznosti členů kategorie, daný sérií řetězců rodinných podobností, aniž by existovala množina jejich společných vlastností), *centralita/prototypy* (některé členy jsou lepšími reprezentanty kategorie než jiné), *odstupňované členství* v kategorii (o tento přístup jsme se do jisté míry pokusily při stanovení „jádra“ typů SOZ), *bázová kategorizace* (tzv. postup *middle-out*: východiskem nejsou ani členy na nejnižší úrovni konkrétnosti, ani členy na nejvyšší úrovni abstrakce, nýbrž nejdůležitější zástupci, od nichž se postupuje jak směrem vzhůru, tak směrem dolů), používání *metonymie* způsobem „*pars pro toto*“ (pojmenováním celku jeho částí, což se aktuálně projevuje např. v používání termínů „slovník“ a „ontologie“ pro označení SOZ). J. Tennis ve svém článku věnovaném

⁴⁴ „There is a serious lack of vocabulary control in the literature on controlled vocabulary“. WEINBERG, Bella Hass. ASIS'97: the classification research workshop. In: *Key words*. 1998, 6(2), 21-22. ISSN 1064-1211.

⁴⁵ LAKOFF, George. *Ženy, oheň a nebezpečné věci: co kategorie vypovídají o naší mysli*. 1. vyd. Praha: Triáda, 2006, s. 25-26. ISBN 978-80-86138-78-7 (váz.).

systematicke SOZ⁴⁶ nabízí ještě další perspektivu. Tvrdí, že pro poznání podstaty SOZ není dostačující omezit se na jádro, za něž jsou tradičně považovány systémy fungující jako pomůcka pro vyhledávání dokumentů, ale že je třeba se zaměřit i na okrajové typy, které přinášejí poznatky o vytváření a implementaci systémů mimo oblast zkoumanou informační vědou. Jejich zkoumání pak podle něj umožní lépe „nasvítit“ tradiční SOZ z jádra.

I když překotný vývoj v současnosti zatím zabraňuje zachytit pevné obrysy a komponenty dynamicky se vyvíjejícího fenoménu, lze určitě vyjádřit naději, že budoucí vývoj a souběžně probíhající výzkum systémů organizace znalostí přinese po určitém ustálení jasnější pohled na jejich podstatné a rozlišující vlastnosti a umožní založit na nich typologii užitečnou pro teorii i praxi.

*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:

Access Innovations. *TaxoBank Terminology Registry: TaxoBank* [online]. Albuquerque (New Mexico, USA): Access Innovations, 2009- [cit. 2014-08-25]. Dostupné z: <http://www.taxobank.org/>.

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. 2014-08-25]. Dostupné z: 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. 2014-08-25]. Dostupné z: 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. 2014-08-25]. Dostupné z: http://wiki.dublincore.org/index.php/NKOS_Vocabularies#KOS_Types_Vocabulary.

Food and Agriculture Organization of the United Nations. *Agricultural Information Management Registry of Vocabularies, Metadata Sets and Tools. VEST Registry. Vocabularies* [online]. Rome: FAO, Office of Knowledge, Exchange, Research and Extension, © 2012 [cit. 2014-08-25]. Dostupné volně z Portálu AIMS: <http://aims.fao.org/vest-registry>.

GOLUB, Korajka, TUDHOPE, Douglas, ZENG, Marcia Lei, ŽUMER, Maja. Terminology Registries for Knowledge Organization Systems: Functionality, Use, and Attributes. In: *Journal of the American Society for Information Science and Technology*. September 2014, **65**(9), 903-912. doi:10.1002/asi.23090. ISSN 1532-2882 (Print). ISSN 1532-2890 (Online). Dostupný také komerčně z Wiley Online Library (DOI): <http://dx.doi.org/10.1002/asi.23090>.

⁴⁶ TENNIS, Joseph T. Fringe types and KOS systematics: examining the limits of the population perspective of knowledge organization systems. In: *Advances in classification research online* [online]. 2009, **20**(1) [cit. 2013-08-23], [14 s.]. doi:10.7152/acro.v20i1.12885. ISSN 2324-9773. -- 20th ASIS SIG/CR Classification Research Workshop (Vancouver, Canada, 2009).

GRABAR, Natalia, HAMON, Thierry, BODENREIDER, Olivier. Ontologies and terminologies: continuum or dichotomy? In: *Applied ontology*. 2012, 7(4), 375-386. ISSN 1570-5838 (Print). ISSN 1875-8533 (Online). Dostupný také komerčně z IOS Press (DOI): <http://dx.doi.org/10.3233/AO-2012-0119>.

HODGE, Gail. *Systems of knowledge organization for digital libraries: beyond traditional authority files*. Washington: The Digital Library Federation, Council on Library and Information Resources, April 2000. vi, 37 s. ISBN 1-933645-06-7 (DLF). ISBN 1-887334-76-9 (CLIR, chybné). Dostupné také volně z webu vydavatele: <http://www.clir.org/pubs/reports/pub91/pub91.pdf>.

ISO 25964-1:2011. *Information and documentation – Thesauri and interoperability with other vocabularies – Part 1: Thesauri for information retrieval*. 1st ed. Geneva: International Organization for Standardization, 2011-08-08. 152 s.

ISO 25964-2:2013. *Information and documentation – Thesauri and interoperability with other vocabularies – Part 2: Interoperability with other vocabularies*. 1st ed. Geneva: International Organization for Standardization, 2013-03-04. 99 s.

KUČEROVÁ, Helena. České termíny pro věcné pořádkání informací po 40 letech: příspěvek k terminologické diskusi. In: *ProInflow: časopis pro informační vědy* [online]. 2013, 5(Speciál) [cit. 2014-02-17], s. 1-19. ISSN 1804-2406.

Dostupné z: http://pro.inflow.cz/sites/default/files/pdfclanky/ProInflow_Kucerova_final_0.pdf.

LYKKE, Marianne. Networked Knowledge Organization Systems/Services (NKOS). In: Marcia J. BATES, Mary Niles MAACK, ed. *Encyclopedia of library and information sciences*. 3rd ed. Boca Raton (Florida): CRC Press, © 2010, s. 3911-3915. doi:10.1081/E-ELIS3-120044732. ISBN 978-0-8493-9712-7 (soubor, Print). ISBN 978-0-8493-9711-0 (Online). Dostupné komerčně také z: <http://www.tandfonline.com/doi/pdf/10.1081/E-ELIS3-120044732>.

MILES, Alistair, BECHHOFFER, Sean, ed. *SKOS Simple Knowledge Organization System Reference* [online]. W3C Recommendation 18 August 2009 [version]. Cambridge (MA): World Wide Web Consortium, © 2009. Dostupné z: <http://www.w3.org/TR/2009/REC-skos-reference-20090818/>.

MOSKOVIČ, Volf Abramovič. *Informacionnyje jazyki*. Moskva: Nauka, 1971. 143 s.

SOERGEL, Dagobert. Digital libraries and knowledge organization. In: Sebastian R. KRUK, Bill McDA-NIEL, ed. *Semantic digital libraries*. Berlin: Springer, 2009, s. 9-39. ISBN 978-3-540-85433-3 (Print). ISBN 978-3-540-85434-0 (Online).

Dostupné také komerčně z (DOI): http://dx.doi.org/10.1007/978-3-540-85434-0_2.

SOUZA, Renato Rocha, TUDHOPE, Douglas, ALMEIDA, Mauricio Barcellos. Towards a taxonomy of KOS: dimensions for classifying knowledge organization systems. In: *Knowledge organization*. 2012, 39(3), 179-192. ISSN 0943-7444. Dostupný komerčně také z EBSCOhost (LISS): <http://search.ebscohost.com/>.

SVENONIUS, Elaine. Bibliographic languages. In: *The intellectual foundation of information organization*. Cambridge (Mass): MIT Press, 2000, chapter 4, s. 53-55. ISBN 0-262-19433-3.

SZOSTAK, Rick. Classification, ontology, and the Semantic web. In: *Advances in classification research online* [online]. 2013, 24(1) [cit. 2014-08-25], 30-37. doi:10.7152/acro.v24i1.14674. ISSN 2324-9773. 24th ASIS SIG/CR Classification Research Workshop.

Dostupné volně z: <https://journals.lib.washington.edu/index.php/acro/article/view/14674>.

TENNIS, Joseph T. Fringe types and KOS systematics: examining the limits of the population perspective of knowledge organization systems. In: *Advances in classification research online* [online]. 2009, 20(1) [cit. 2014-08-25], [14 s.]. doi:10.7152/acro.v20i1.12885. ISSN 2324-9773. 20th ASIS SIG/CR Classification Research Workshop.

Dostupné volně z: <https://journals.lib.washington.edu/index.php/acro/article/view/12885/11381>.

TUDHOPE, Douglas, KOCH, Traugott. New applications of knowledge organization systems: introduction to a special issue. In: *Journal of digital information* [online]. 2004, 4(4) [cit. 2014-08-25]. ISSN 1368-7506. Dostupné z: <https://journals.tdl.org/jodi/index.php/jodi/article/view/109/108>.

Universitätsbibliothek Basel. *BARTOC.org: BAseL Register of Thesauri, Ontologies & Classifications* [online]. Projektleiter Andreas Ledl. Basel: Universitätsbibliothek Basel, 2013- [cit. 2014-08-25]. Dostupné volně ze serveru Basilejské univerzity: <http://www.bartoc.org/>.

World Wide Web Consortium. Vocabularies. In: *World Wide Web Consortium (W3C)* [online]. Cambridge (MA): World Wide Web Consortium, © 2012 [cit. 2014-08-25].

Dostupné z: <http://www.w3.org/standards/semanticweb/ontology>.

WEINBERG, Bella Hass. ASIS'97: the classification research workshop. In: *Key words*. 1998, **6**(2), 21–22. ISSN 1064-1211. Dostupné komerčně také z EBSCO (LISS): <http://search.ebscohost.com>.

ZENG, Marcia Lei. Knowledge Organization Systems (KOS). In: *Knowledge organization*. 2008, **35**(2-3), 160-182. ISSN 0943-7444.

ZENG, Marcia Lei. Taxonomy of knowledge organization sources/systems. In: *Networked Knowledge Organization Systems/Services: NKOS* [online]. Kent: School of Library and Information Science, Kent State University, Draft June 7, 2000, revised July 31, 2000 [cit. 2014-08-25]. Dostupné z: http://nkos.slis.kent.edu/KOS_taxonomy.htm.

ZENG, Marcia Lei, ŽUMER, Maja. A metadata application profile for KOS vocabulary registries. In: *Knowledge organization: pushing the boundaries: ISKO UK [3rd] Biennial Conference, 8th–9th July 2013, London* [online]. London: ISKO UK, 2013 [cit. 2014-08-25].

Dostupné z: <http://www.iskouk.org/conf2013/papers/ZengPaper.pdf>.