

Soubory autorit a internet

Radka Římanová*

r.rimanova@stk.cz

Zdeněk Bartl**

Zdenek.Bartl@nkp.cz

Abstrakt: Autoři článku vycházejí z publikovaných výsledků práce Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), studie návazující na studii FRBR. Upozorňují na problematiku autoritního záhlaví jako propojovacího prvku mezi databázemi různého typu a využívaného při vyhledávání na internetu prostřednictvím SFX služeb. Informují o dvou experimentálních studiích o vyhledávání osob a korporací na Internetu. Autoři uvažují o možné formě podchycování zdroje informací dostupných na internetu, který je často využíván při ověřování autorit, v záznamech ve formátu MARC21/Authorities Format, a to v souvislosti s projektem „Tvorba a využívání souboru národních autorit“.

Klíčová slova: FRANAR, MARC21 670\$, autority, vyhledávání na internetu

Portály budované knihovnami integrují knihovní katalog do kontextu dalších možností vyhledávání informací využíváním služeb SFX, propojení na faktografické databáze, elektronická knihkupectví a různé typy elektronických vyhledávačů ve volném webu. Při tomto typu propojení je limitujícím faktorem použití optimálního řetězce slov, získaného z bibliografického záznamu a použitého k vyhledání v dalším zdroji. Soubory autorit budované jako soubory selekčních prvků pro vyhledávání v bibliografických databázích sehrávají v tomto procesu důležitou roli. Autority jsou charakterizovány jako můstek mezi vnějším světem a databází. Tato charakteristika byla publikována v článku G. E. Pattona „FRANAR: a conceptual model for authority data“, který byl přednesen na konferenci „Authority Control: Reflections and Experiences“, pořádané v roce 2003 ve Florencii. Význam této konference pro tvorbu autoritních souborů dosud trvá, o čemž svědčí i skutečnost, že časopis *Cataloging and Classification Quarterly* věnoval v roce 2004 dvě monotematická čísla vydání příspěvků z této konference. Příspěvky jsou zároveň volně přístupné na internetových stránkách konference, některé z nich vyšly separátně v odborném tisku.

G. E. Patton je členem pracovní skupiny Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). Tato pracovní skupina si stanovila ve svých úkolech rozpracování prvků studie Functional Requirements for Bibliographic Records (FRBR) v oblasti autoritních entit (jmenné a předmětové). Výchozím materiálem pro zpracovávání autoritních souborů je druhé vydání (2001) *Guidelines for authority records and references*, kde byly popsány základní charakteristiky prvků zpracovávaných v rámci procesu „authority control“. Toto druhé vydání pravidel (první vydání *Guidelines...* bylo vydáno v roce 1984) je již ovlivněno studií FRBR. V tomto materiálu byla definována anglická odborná terminologie, interpunkce záznamů a oblasti popisu. Studie je zpracována jako univerzální materiál nezávislý na konkrétních katalogizačních instrukcích, a proto může být

* Státní technická knihovna, Mariánské náměstí 5, 110 01 Praha 1

** Národní knihovna ČR, Klementinum 190, 110 00 Praha 1

propojovacím materiálem při sdílení (intelektuálním nebo fyzickém) výsledků produkce jednotlivých bibliografických agentur. Ve strategickém plánu rozvoje IFLA je podporou dalšího rozvoje studie FRBR i FRANAR pověřena British Library, z jejíchž zpráv vyplývá, že úkol přijala, ale ne již, v jakém časovém sledu budou očekávané zprávy prezentovány.

G. E. Patton ve svém vystoupení zdůraznil nutnost participace jednotlivých národních bibliografických agentur na vytváření „vlastních autorit“.

Problematice technického řešení sdílení národních souborů autorit se věnoval projekt Linking and Exploring Authority Files (LEAF), koordinovaný Staatsbibliothek zu Berlin a ukončený v roce 2004.

Uveřejněný konceptuální model FRANAR (v současné době prezentovaný jako FRAR - Functional Requirements for Authority Records and References) se zabývá pouze jmennými autoritami (osoby, rodiny, korporace, název a entity typu jméno-název). V roce 2002 zpracoval kanadský knihovník Tom Delsey první návrh konceptuálního modelu, který v následujícím období přepracoval v souladu s výsledky výzkumu pracovní skupiny FRANAR. Zájemce o tuto studii odkazují na článek G. E. Pattona, ze kterého jsme převzali grafickou prezentaci modelu, která je nejvýstižnější formou jeho vyjádření (obr. 1, obr. 2).

Model popisuje prostředí, ve kterém je ukotvena „autorita“ (name + identifier). Na levé straně modelu jsou umístěny entity skutečného světa – osoba, rodina, korporace, dílo, vyjádření, zhmotnění, jednotka, pojem, předmět, událost a místo. Tyto jsou popsány v termínech vytvořených pro definování metadat, <index> model (nazývané „primitives“): bytost – životná entita; věc – neživotná entita; pojem – existuje nezávisle na času, prostoru a je nehmotný; událost – dynamický vztah zahrnující dvě a více entit ve vzájemném vztahu, který je proměnný a nestálý v čase a místě; situace – statický vztah zahrnující dvě a více entit, které jsou pevně zakotveny v čase a místě.

Na pravé straně modelu jsou prvky směřující ke konkrétním záznamům a jejich částem.

Toto pojetí otevírá principy zpracování autorit nejen potřebám vytváření selekčních prvků pro bibliografické potřeby, ale i pro propojování v dalších typech databází (muzejních, archivních, faktografických).

Struktura záhlaví autoritního záznamu (nebo odkazované záhlaví) je pojata jako samostatný prvek a respektuje specifický tvar vycházející z konkrétních katalogizačních pravidel. Identifikátor autority není přímo součástí záhlaví.

V článku G. E. Pattona je naznačen další směr rozpracování tohoto základního modelu směrem k dalšímu zjemnění prvků, zahrnutí problematiky různojazyčných jmen (pojmenování) v různých abecedách.

Konceptuální model zahrnuje také uživatelské určení autoritního záznamu. Jedná se o přímé využití v knihovnické práci (katalogizace, referenční služby), službu uživatelům knihovny (vyhledání informace) a prostředek pro správu databáze (redakce a údržba katalogů). Procesy, které probíhají v rámci manipulace s autoritními záznamy, jsou rozděleny do dvou základních skupin: zjišťování zdrojů (vyhledávání, identifikace, kontrola výrazů a vytvoření vazeb k dalším prvkům databáze) a správa dat (zpracování, třídění, zobrazení a integrace do struktur databáze). Ve svém příspěvku G. E. Patton předpokládá předložení výsledků práce FRANAR k návrhu rozvoje FRBR. Potvrzení realizace se nám však nepodařilo v odborném tisku vypátrat.

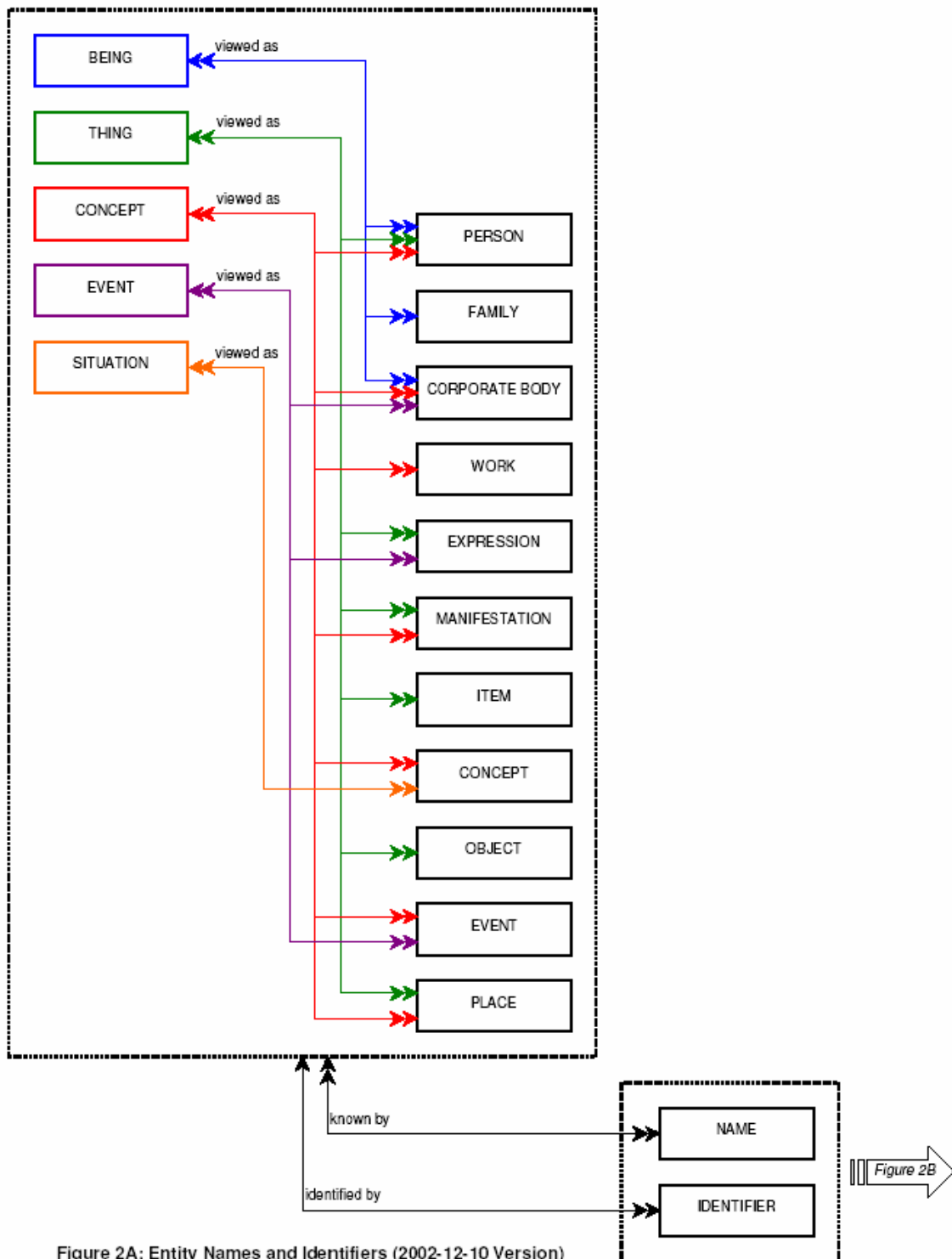


Figure 2A: Entity Names and Identifiers (2002-12-10 Version)

Obr. 1

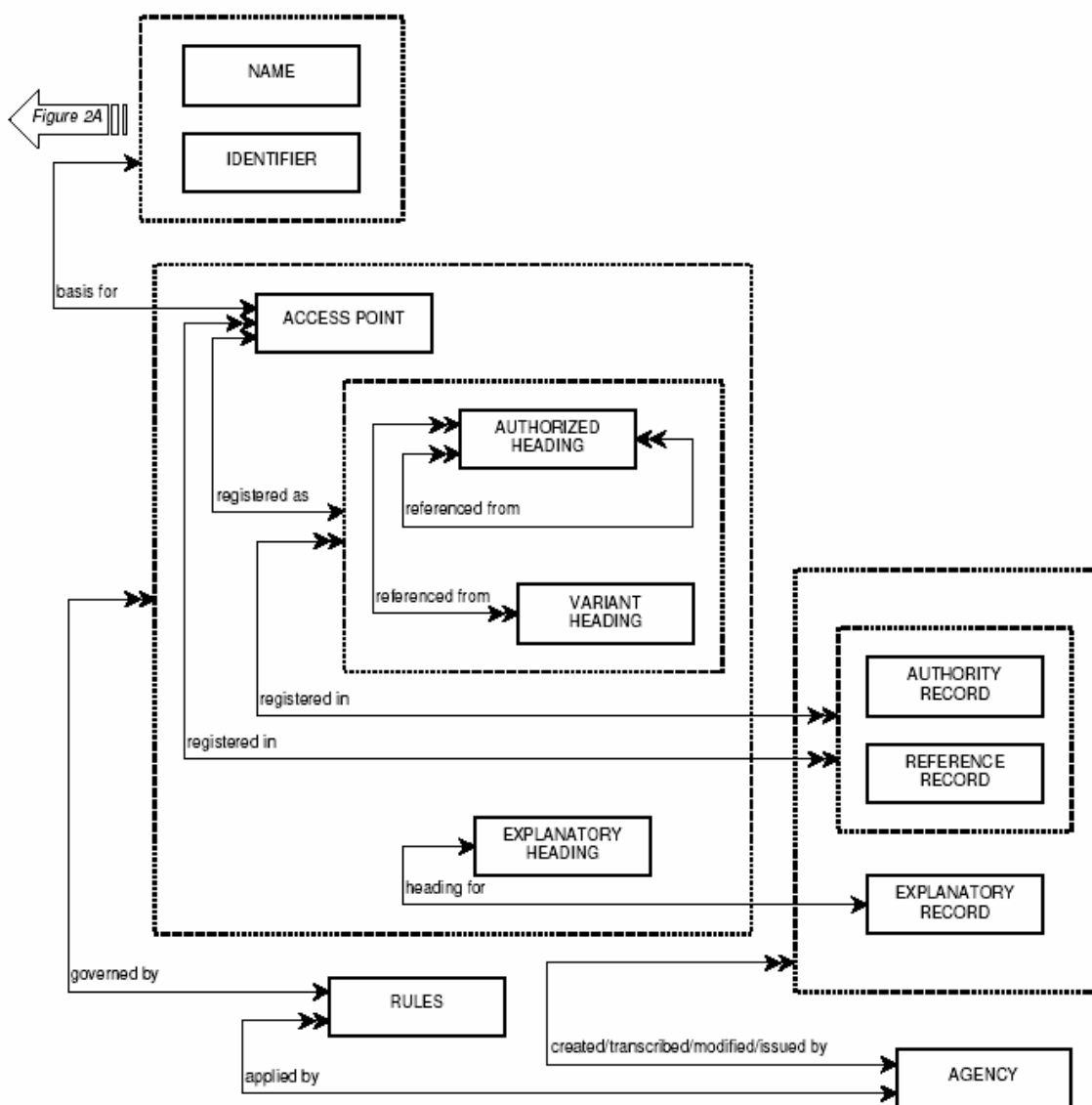


Figure 2B: Access Points and Authority Records (2002-12-10 Version)

Obr. 2

Vraťme se však k nastolenému problému použití záhlaví autoritního záznamu jako selekčního řetězce v prostředí volného webu.

Z údajů studie [11] zkoumající laickou rešeršní strategii ve vyhledávacích vyplývá, že toto vyhledávání nepřináší vždy relevantní výsledky. Uživatelé převážně používají jednoduché vyhledávání, bez kombinace termínů a operátorů. Ve studii je popsán pokus, kdy v náhodně vybraném vzorku 20 000 dotazů položených ve vyhledávacích Alta Vista a AlltheWeb.com byly zkoumány dotazy směřující k vyhledání informace o osobnosti. Bylo zjištěno, že z celkového počtu dotazů celkem 4 % směřovala na vyhledání informace o osobnosti, z toho

bylo 26 % zaměřeno na informace o celebritách, 18 % dotazů bylo upřesněno jako souvislý řetězec („John Smith“), 14 % dotazů bylo sestaveno jako kombinace jméno + přidaný termín („John Smith“ Atlanta).

Autoři studie se pozastavili nad tím, že ačkoli kombinace jméno + přidaný termín přináší nejlepší výsledky, žádný vyhledávač tento způsob definování dotazu nijak transparentně nedoporučuje a uživatele k němu nenaviguje.

I to může být jeden z důvodů, vedle dalších aspektů, že je tato strategie opomíjena. Z pohledu informačních odborníků se laický uživatel bohužel spokojí s málem.

Pokud navigují elektronické katalogy (OPAC) na další vyhledávání ve faktografických databázích (např. encyklopediích) nebo volném webu, měli bychom dbát i na to, aby termín automaticky vkládaný do vyhledávače přinesl relevantní výsledek.

Problém stanovení vhodného selekčního prvku v záznamu je nutné řešit již při vytváření záznamu.

Jedním z kroků tímto směrem je v současnosti doporučovaná praxe vytváření předmětových hesel ve tvaru přirozeného jazyka (bez inverze).

Pro oblast jmenných záhlaví však platí přesné katalogizační instrukce, které není možné okamžitě změnit, i když z pohledu laika vytváří knihovník v katalogích někdy záhlaví zcela nepochopitelná – zápis korporací pod jurisdikcí, přehazování pořadí částí jmen u jmen složených z více než dvou částí apod.

Jedním z řešení by mohlo být nabídnutí všech odkazovaných variant výrazu k vložení do formuláře a uložení jména při vyhledávání v přirozeném pořadí (křestní jméno, příjmení).

Další možností pro zpřesnění vyhledávání v prostředí volného webu je vepsání internetové adresy do autoritního záznamu. Formát MARC21 pro autoritní záznam má pro tyto údaje definovaná pole 856 \$u a 670 \$u. Katalogizátor zpracovávající nový záznam s dokumentem v ruce má oproti uživateli používajícímu pro vyhledání pouze záhlaví autoritního záznamu výhodu širší možnosti zvolení slov „přidaného termínu“.

Pokud je tímto termínem název knihy, kterou autor napsal, budou na prvních místech výsledků vyhledávání uvedeny vstupy do internetových knihkupectví nabízejících titul. Tyto stránky obsahují velice skromné údaje o autorech. Mnohem relevantnějších výsledků dosáhneme, pokud přidaným termínem bude pracoviště autora, místo jeho působení, případně slovo charakterizující jeho publikační obor. Význam této strategie se zvyšuje, jde-li o často frekventované jméno (John Smith) nebo o jmenovce z oblasti show byznysu, politiky atd.

Problém formy jména korporace v záhlaví a jeho vyhledávání na internetu z pozice knihovnické vědy se zabývala studie publikovaná autorem Quing Jin v článku „Creating up-to-date a corporate name authority records by using official a corporate home web page“, která popisuje pokus porovnání 100 autoritních záznamů významných korporací (záhlaví + odkazy) s názvy organizací na internetu. Bylo zjištěno, že 25 % záhlaví nebylo možné na internetu vyhledat, protože obsahovalo jiné řetězce slov. Autor pokládá otázku, zda při zjištění této situace má katalogizátor vytvořit nové záhlaví s propojovacím polem 510. Efektivita takového počínání je velmi sporná, protože pojmenování uvedené na vstupní internetové stránce nemusí mít charakter změny názvu, jak jej známe z prostředí tradičních dokumentů, a může se změnit na název „původní“ nebo modifikovat do dalších variant.

Podle našeho názoru dojde ke zjištění takové změny vážící se k již vytvořenému a mnoha knihovnami sdílenému záznamu pravděpodobně jen náhodou, a to jen tehdy, kdy se moduluje název uváděný na zpracovávaných dokumentech (v ČR se jedná nejčastěji o tištěné dokumenty a elektronické dokumenty s off-line přístupem).

Autor studie předkládá jako jednu z variant uvést takovou informaci do pole 670 („poznámka o ověření“), a to konkrétně do podpole \$a („zdroj citace“) a podpole \$b („nalezená informace“).

Nabízí se tu však možnost využití již zmiňovaných polí určených pro propojování. O podpoli 670 \$u („unifikovaný identifikátor zdroje“) jsme zjistili, že toto podpole bylo do struktur formátu MARC21 autority definováno roce 2002 a v českém překladu manuálu MARC21 – Formát pro autority je charakterizováno: „Podpole \$u obsahuje Unifikovaný identifikátor zdroje (URI), např. URL či URN, který uvádí adresu zdrojů ve standardním syntaxu. Tato se uvádějí za účelem automatizovaného zpřístupnění elektronických jednotek používajících jeden z internetových protokolů“.

V příkladech, které doplňují rozhodnutí o zavedení 670 \$u na stránkách Library of Congress, je toto podpole vždy uvedeno v samostatném výskytu. Současně s podpolem 670 \$a, kde je slovně uveden název zdroje, včetně data zobrazení stránky.

Také jsme zjistili, že toto podpole nemá svůj identický ekvivalent ve formátu UNIMARC (pole 810 – Zdroj ověření dat obsahuje pouze \$a Zdroj citace; \$b Nalezená informace).

Příklad užívání podpole 670 \$u jsme zatím objevili pouze v prezentaci B. B. Tillett : „Authority Control“ přednesené na již zmíněné konferenci ve Florenci. Ta prezentovala záznam Library of Congress se záhlavím „Confucius“ a uvedením odkazu na stránku anglického životopisu autora. Ovšem v záznamu v katalogu Library of Congress jsme již tento záznam našli bez uvedení propojení. Ana Lupe Cristán (koordinátor kooperativní katalogizace BICO) a B. Tillett, působící v Library of Congress, potvrdily, že v autoritní databázi této knihovny přímé adresy nejsou uvedeny. Prvním důvodem jsou technické problémy systému, pokud adresy obsahují nějaké speciální znaky (zmiňován znak „tilde“ ~), a druhým důvodem je nestabilita internetových zdrojů a navigace uživatele na neexistující informaci. Naznačené problémy hodlají řešit implementací kontrolního software, který by sledoval a opravoval platnost linků. Programová úprava je očekávána do pololetí roku 2006. Poté bude zahájeno testování ukládání internetových adres do autoritních záznamů (670 \$u).

Stanovisko Library of Congress potvrzuje složitost dané problematiky. Nicméně se domníváme, že k využívání aktivního linku v podpoli 670 \$u v nějaké podobě (po vyřešení technických problémů) v blízké budoucnosti nakonec zřejmě dojde. Pokud chceme citovat zdroj ověření dat, konkrétně internetovou stránku, je nepřesnější citací, podle našeho názoru, uvedení internetové adresy. Pole 670 \$u je právě pro informace o zdroji dat určeno.

Koncepcí každého katalogu pak může být rozhodnutí, zda tato adresa bude aktivní, nebo pouze textem, který je možné využít vložením do vyhledávače. Pokud by měla být aktivním prvkem, je nutné pravidelně databázi redigovat a platnost linku kontrolovat.

Nabízí se otázka, zda by bylo uvádění internetové adresy další zátěží pro knihovny kooperující v národním systému. Domníváme se, že by nemuselo, protože při vytváření nového autoritního záznamu popisujícího v současné době existující entitu – osobnost nebo korporaci či akci – je vyhledávání na internetu jedním ze standardních a neekonomičtějších postupů při ověřování informací.

Pokud zůstane vyhledaná adresa skryta např. za formulací „www(osobní stránka)“, je energie vynaložená pro prohledávání internetu využita (pouze) pro ověření biografických a dalších informací.

Předpokládáme ještě jednu námitku proti ukládání internetových adres do záznamů. Autoritní soubor není biografický slovník ani adresář firem. Stanovení meze této podmínky je však otevřenou otázkou, na kterou upozorňuje Françoise Bourdon v příspěvku „Functional requirements and numbering of authority records (FRANAR): to what extent authority control

can be supported by technical means?“ V novější práci G. E. Pattona není tento problém nijak transparentně zmiňován. Ve společné uživatelské skupině, využívající autorit přímo, uvádí autor katalogizátory i referenční knihovníky. Úkol autoritních souborů nespočívá pouze v oblasti bibliografické, ale v širších souvislostech různých typů databází. Užívání autoritních databází ve smyslu vyhledání základních informací o entitě se začíná ve světě objevovat v článcích o referenčních službách v knihovnách.

Domníváme se, že praxí českých knihoven byly ověřeny (při zpracování nových přírůstků a zvláště při retrospektivní katalogizaci) výhody užití rozsáhlejší biografické poznámky, případně navigace na další zdroje informací o osobnosti (biografické články, nekrology). Tento způsob práce má své opodstatnění i v tom, že soubor národních autorit plní zároveň funkci zdroje pro předmětová hesla personální a korporativní. Samozřejmě si uvědomujeme ekonomickou náročnost tohoto modelu a disproporci, která vyplývá z toho, že celá řada specializovaných oborových knihoven (institucí) není zapojena do kooperativního projektu „Tvorba a využívání souboru národních autorit“.

Výrazný rozdíl zatím panuje mezi webovou prezentací osobností české a zahraniční vědy. Souvisí to s prezentačními snahami institucí, které zdůrazňují svůj význam potenciálem osobností v nich působících. Tuto potřebu zřejmě české akademické instituce zatím dostatečně nepocítily. Přesto i na jejich stránkách jsou uváděny alespoň základní vizitky, které informují o jméně a akademických titulech osobnosti, kontaktní informace (e-mail, telefon), příslušnost k části instituce apod. Stránky zahraničních autorů působících ve vyspělých zemích světa obsahují mimo uvedené údaje i fotografii, odborný profil, bibliografii prací, případně curriculum vitae. Nacházíme zde často i osobní informace privátního charakteru. Kontaktní informaci lze využít k navázání komunikace s autorem. Velké procento oslovených osobností reaguje vstřícně na žádost o doplnění informací o sobě (datum a místo narození, publikační obor).

Uživatelům může napomoci uvedení internetového linku získat informace, které jsou nad rámec biografické poznámky o autorovi v autoritním záznamu (kontaktovat autora se žádostí o separát atd.).

Pro samotné autory má přesné bibliografické určení jejich publikační činnosti velký význam. V dnešní době je publikační činnost a z ní vycházející citační analýza jedním z prvků hodnotících význam odborníka. I když autentifikace osobnosti v citačních databázích je zatím postavena na principu identifikace autora vůči pracovišti, je pravděpodobné, že v budoucnu dojde k propojování knižních katalogů, soupisů publikovaných prací, článkových bibliografií, přičemž společným prvkem všech těchto databází bude autoritní záhlaví.

Použitá literatura a WWW odkazy

1. Authority control : definition and experiences : international conference : Florence, Italy, February 10-12, 2003 [online]. Dostupné na www: <<http://www.sba.unifi.it/ac/en/home.htm>>.
2. BOURDON, F. Functional requirements and numbering of authority records (FRANAR): to what extent authority control can be supported by technical means? In: *67th IFLA General Conference and Council, August 16th-25th, 2001, Boston, USA [on line]*. The Hague: International Federation of Library Associations and Institutions, 2001. Dostupný na www: <<http://www.ifla.org/IV/ifla67/papers/096-152ae.pdf>>.
3. BRATKOVÁ, E. [autor statě]: Informační entity, jejich znaky a vztahy : podle modelu FRBR. In *Informační studia a knihovnictví v elektronických textech I [CD-ROM]*. 2002. Praha: Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví, s. 1-57.
4. IFLA Working Group on GARE Revision. Guidelines for Authority Records and Reference. München : K.G. Saur, 2001. - (UBCIM Publications : NewSeries, vol.23). Dostupné na www:

- <<http://www.ifla.org/VII/s13/garr/garr.pdf#search='Guidelines%20for%20authority%20record%20and%20references%20second%20edition'>>.
5. JIN, Q.: Creating Up-to-Date Corporate Name Authority Records by Using Official Corporate Home Web Pages. *Cataloging & Classification Quarterly*, 2004, vol. 38, no. 3-4, s. 287-290.
 6. JIN, Q. Authority Control in the Online Environment: Celebrating the 20th anniversary of LITA/ALCTS CCS Authority Control in the Online Environment Interest Group. *Cataloging & Classification Quarterly*, 2004, vol. 38, no. 2, s. 101-109.
 7. KAISER, M.; LIEDER, H.-J.; MAJCEN, K.; VALLANT, H. New ways of sharing and using authority information: the LEAF Project [online]. *D-Lib Magazine*. 2003, vol. 9, no. 11. Dostupné na [www](http://www.dlib.org/dlib/november03/lieder/11lieder.html): <<http://www.dlib.org/dlib/november03/lieder/11lieder.html>>.
 8. MARC Proposals. Proposal no. : 2002-01. [online]. Washington. Library of Congress. 1995-. Dostupný na [www](http://www.loc.gov/marc/marbi/2002/2002-01.html): <<http://www.loc.gov/marc/marbi/2002/2002-01.html>>.
 9. PATTON, G. E. FRANAR : a conceptual model for authority data. *Classification Quarterly*, 2004, vol. 38, no. 3-4, s. 91-103. Dostupný též na [www](http://eprints.rclis.org/archive/00000312/): <<http://eprints.rclis.org/archive/00000312/>>.
 10. RUST, G.: The <indec>model : some key issues. MODELS workshop, 1999. Dostupný na [www](http://www.ukoln.ac.uk/dlis/models/models10/gr-mod10.ppt): <<http://www.ukoln.ac.uk/dlis/models/models10/gr-mod10.ppt>>.
 11. SPINK, A.; JANSEN, B. J.; PEDERSEN, J. Searching for people on Web search engines. *Journal of documentation*, 2004, vol. 60, no. 3, s. 266-278.
 12. TILLET, B. B. Authority control : state of the art and new perspectives. *Cataloging & Classification Quarterly*, 2004, vol. 38, no. 3-4, s. 23-41. Dostupný též na [www](http://www.sba.unifi.it/ac/relazioni/tillett_eng.pdf): <http://www.sba.unifi.it/ac/relazioni/tillett_eng.pdf>.
 13. UNIMARC Manual : authority format. München. Saur. 2nd ed. 2001. 200 s. (UBCIM Publications – New Series, vol. 22).
 14. WADHAM, R. Authority Records in Reference Service. *Library Mosaics*, 2004, vol. 15 no. 5, s. 23.