

AUTORITNÍ ZÁZNAMY, ČIŠTĚNÍ REJSTRÍKŮ ANEB LZE SE POUČIT Z PROJEKTŮ EU?

Radka Římanová, Státní technická knihovna

Několik projektů řešených v minulých letech za podpory Evropské unie se věnovalo autoritním souborům (projekt AUTHOR, LEAF, VIAF). Společným rysem těchto projektů je snaha sdílet náklady na vytváření autoritních databází, které jsou budovány především národními bibliografickými agenturami a které jsou stále pocitovány jako jistý stupeň luxusu v práci katalogizátorů.

Hlavní cíl budování autoritních databází, jak je definován v právě připomínkové konceptuální studii Funkční požadavky na autoritní záznamy [IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR), 2007] je vytvoření selekčního prvku (nebo jeho odkazu) pro třídění a vyhledávání v bibliografických databázích. Je vyjádřeno těmito funkcemi „najít entitu“ (Find) – „identifikovat entitu“ (Identify) – „dát do souvislosti s jinou entitou“ (Contextualize) - „přiradit správnou autoritní podobu záhlaví pro daný dokument“ (Justify). Každá autoritní entita je určována sadou vlastností a vzájemnými vazbami mezi nimi.

Autoritní záznam vzniká při bibliografickém popisu dokumentu a v souladu s konkrétní katalogizační instrukcí. Tato skutečnost je velice podstatná pro kooperaci mezi databázemi vznikajícími podle různých pravidel pro bibliografický popis (viz projekt VIAF) nebo paměťovými institucemi (viz projekt LEAF).

Evropská komise financovala v letech 1995–1997 projekt **AUTHOR**. Projekt měl za cíl zjistit možnosti mezinárodní spolupráce v oblasti tvorby a zpřístupňování osobních a korporativních autoritních hesel. Hlavní řešitelskou institucí byla Národní knihovna Francie (Bibliothèque nationale de France). Během projektu byla připravena konverze autoritních záznamů různých národních MARC formátů do komunikativního formátu UNIMARC.

V oblasti technického řešení projektu byl testován protokol Z39.50. Bylo úspěšně ověřeno virtuální simultánní prohledávání více autoritních databází zároveň [ZILLHARDT, 1998, s. 12].

Projekt odhalil významné rozdíly vyplývající z odlišného pojetí principů tvorby záhlaví pro autoritní záznam. Některé národní bibliografické

agentury ukládaly údaje do strukturovaných polí záznamu, jiné agentury stejné údaje zapisovaly do polí poznámek.

Další významné rozdíly byly zaznamenány u popisu korporativních entit. Způsob práce s variantním záhlavím nebyl shodný u žádných ze zkoumaných databází. Některé národní bibliografické agentury tvořily i pro záhlaví dočasných akcí (festivaly, konference) záhlaví podle pořadajících korporací stálých, nebo ho netvořily vůbec a dokumenty popisovaly pod názvem. Další okruh problémů vyplýval z transliterace jmen v jiných abecedách. Významné rozdíly byly zaznamenány také ve škále polí, která jednotlivé národní agentury vyplňovaly [TILLET. 2001].

Ve shrnutí výsledků projektu je mimo jiné uvedeno [ZILLHARDT. 1998. s. 20]: **Prohledávání autoritní databáze je plně efektivní pouze za předpokladu, že záznam bude prohledáván celý, včetně poznámkových polí.**

Pod záštitou IFLA byl v roce 2003 zahájen projekt **Virtuální mezinárodní autoritní soubor (Virtual International Authority File, VIAF)**. Do projektu se zapojily Německá knihovna (Die Deutsche Bibliothek, DDB), Kongresová knihovna (Library of Congress, LC) a OCLC počítačové online knihovní centrum (OCLC Online Computer Library Center, OCLC) (BENNETT, 2006a).

Projekt měl pět fází: vytvořit společný autoritní záznam z databáze Personennormdatei (PND) a autoritní databáze Kongresové knihovny. Dalším cílem bylo vytvoření spojovacího algoritmu pro porovnávání obou databází. Technologický princip pro práci s databázemi byl stanoven na standardech Open Archive Initiative (OAI). Čtvrtým cílem bylo vytvoření aktualizací souboru, který bude reagovat na změny v obou souborech a zabezpečí sdílení dat. Poslední fází projektu bylo vytvoření uživatelského rozhraní pro prohlížení záznamů databáze VIAF.

Jedním z problémů, který řešil projekt VIAF bylo navazování souvisejících autoritních záznamů na záznamy bibliografické. Z tohoto důvodu si řešitelé projektu všimli všech atributů, které jsou společně autoritní i bibliografické databázi a směřují k určení autoritní entity a to i v případě, že je v záhlaví autoritního záznamu uložena jiná varianta jména nebo názvu (různá pravidla → různá záhlaví).

Za velmi významné lze považovat zjištění prvků, které vedou k přesné strojové identifikaci autorů. Při vytváření souborného záznamu, tkz. vyšší autority (Enhanced Authority), byly údaje z obou záznamů spojovány do jediného záznamu údaje z PND i autoritní databáze LCC.

Pro tyto údaje byly ukládány ve formátu MARC21 do speciálních polí z řady 9XX. Např. pole 90X - identifikační čísla typu ISBN, 91X - názvová pole, 92X - údaje o vydavatelích, 93X - údaje o odpovědnosti z polí 245 \$c

MARC21, 94X - atributy (např. jazyk dokumentu, autorská role, znak Konceptu), 95X - spoluautoři, 96X - předmětové údaje (z pole 600 MARC21 bibliografického záznamu), 99X - speciální pole (např. číslo připojeného bibliografického záznamu) [BENNETT. 2006].

Údaje, které vedly k přesnému spojení záznamů byly rozděleny do dvou skupin: **přesné vlastnosti**: název díla, standardní číslo, přesná biografická data, spoluautoři, předmětové heslo personální; **slabší vlastnosti**: pouze jedno z biografických dat (datum narození a úmrtí), oblast působení, typ publikací, ke kterým se autor váže (knihy, filmy, hudebniny,...), jazyk, nakladatel, dílčí názvy děl, datum publikování, země působení, autorská role.

Při vytváření vyššího autoritního záznamu byly údaje získávány z autoritní i bibliografické báze. Celkem bylo spojováno 2 659 276 DDB jmen s 4 187 973 LC jmény. Úspěšně bylo spojeno celkem 558 618 záznamů [BENNETT. 2006]. Projekt **Propojování a prohledávání autoritních souborů (Linking and Exploring Authority Files, LEAF)** [Consortium LEAF. 2001] byl řešen od roku 2001 do roku 2004. Ve své metodice navázal na projekt AUTHOR. Na jeho řešení se podílely knihovny a archivy z mnoha zemí Evropské unie.

V rámci projektu byl vytvořen centralizovaný repozitář autoritních záznamů získaných z lokálních národních databází různých paměťových institucí (knihovny, archivy), v různých datových formátech, zpracovaných podle různých pravidel zpracování a v různých jazycích. To vše se vzájemným propojením záznamů o téže autoritní entitě. Toto propojování bylo stanoveno nejen na úrovni propojování řetězců slov v záhlaví, jehož nespolehlivost projekt prověřil, ale i na bázi identifikačních čísel, případně identifikátorů autoritních entit, z jednotlivých lokálních systémů. Každý záznam v centrální databázi po přijetí obdržel své stálé identifikační LEAF číslo [KAISER, 2003].

Jedním z principů řešení projektu byla podmínka, že v centrální databázi po uložení nebudou ztraceny žádné údaje, které záznam obsahoval v lokálním systému [Weber, 2004, s. 231].

Při aktualizaci dat pracuje systém metodou sklizení – „harvesting“ (protokol OAI-PMH), nové a aktualizované záznamy byly pravidelně sbírány z lokálních databází do centrální databáze, kde byly ukládány v jednotném výměnném formátu, který je odvozen od formátu XML EAC, metadatový formát Encoded Archival Context (EAC) ve formátu XML.

Tento formát, jehož prezentace je dostupná na internetu [Encoded Archival Context Working Group, 2004], vytváří mezinárodní pracovní skupina EAC Working Group. Požadavkem projektu bylo umožnit uživatelům zobrazit jakýkoli záznam z databáze v různém autoritním formátu bez ohledu na to, jaký byl vstupní formát záznamu. Tyto zobrazovací formáty byly

pojmenovány s přídatkem jako „like“ (MARC 21-like, UNIMARC-like, MAB2-like). Tato funkce je natolik potřebná, že řešitelé projektu k jejímu definování přistoupili navzdory tomu, že automaticky generovaný formát nedosahuje kvality původního záznamu [KAISER, 2003].

Jutta Weberová [2004, s. 235] ve shrnutí poznatků a závěrů projektu, mimo jiné uvádí:

- Informace získané o jedné osobě z lokálních systémů musí být zobrazeny společně.
- Všechny informace získané k jedné osobě musí být porovnatelné.
- Sada údajů, které se mají uvádět do autoritního záznamu osoby by měla být standardizována.
- Šance pro úspěšné vyhledání záznamu o autoritní entitě úměrně narůstá s každým jednotlivým záznamem nebo údajem, který je přidán do centrální databáze.
- Nové a přidávané údaje jsou přínosem i pro uživatelskou veřejnost.

V České republice se podařilo do budování národního autoritního souboru v rámci projektu **Kooperativní tvorba a využívání souboru národních autorit** koordinovaného Národní knihovnou ČR zapojit více než 60 knihoven. Velice kladně ke spolupráci přispěl dotační program Ministerstva kultury VISK, který má od roku 2001 pro tento projekt vyhrazen samostatný podprogram v programu VISK9 (v roce 2000 byl projekt řešen v rámci programu pro souborný katalog ČR). Tento program umožňuje získat prostředky knihovnám na nákup protokolu Z39.50, který je zásadní podmínkou pro spolupráci knihoven, které mají jiný knihovnický software než Národní knihovna ČR (Aleph, firmy Ex Libris). Jedná se o velmi úspěšný projekt, který významně přispívá ke kultivaci české katalogizační obce a bibliografických databází na lokální i souborné úrovni.

Právě úspěšnost tohoto projektu dává naději, že těchto pevných základů lze využít pro další rozvoj a další posunutí práce na projektu.

Typickou činností, která je podporována ve zmiňovaném dotačním programu VISK9 je „čištění lokálních autorských přístupových rejstříků“. Toto čištění je velmi pracné a nákladné a probíhá především na bázi ruční práce, i když automatizované knihovnické systémy nabízejí pro tuto činnost určitý stupeň podpory (např. systém Advance Rapid Library). Cílem čištění je odstranit duplicitní výskyty téhož záhlaví v rejstříku (stejná autoritní entita), opravit chybně zapsaná záhlaví. Součástí těchto projektů je i harmonizace databáze autoritních záznamů s národní databází, položky rejstříků generované z autorských polí bibliografického záznamu jsou propojeny s autoritním záznamem vložením záhlaví z autoritního záznamu do pole 1XX nebo 7XX záznamu bibliografického. Ovšem míra automatizovaného propojování je velmi omezená, protože v současné době nemůže nikdy nahradit inte-

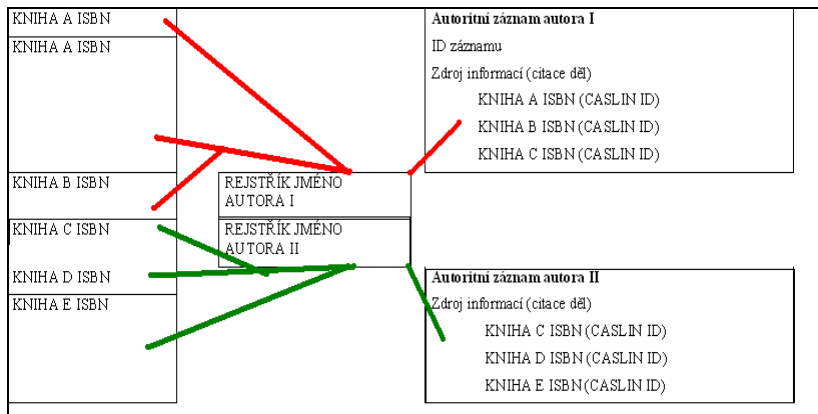
lektuální práci při rozhodování, zda konkrétní autor je autorem konkrétní publikace, neboť tuto informaci nenese v sobě v současné době autoritní záznam vytvářený pro český národní soubor (viz projekt VIAF).

Další knihovna, která k čištění rejstříků přistoupí, **nemůže plně automatizovaně** využít práce jiné knihovny k propojení svých bibliografických záznamů s autoritní databází.

Čištění rejstříků a přiřazování bibliografických záznamů k autoritním je pro operátora jednodušší, pokud je v autoritních záznamech uloženo více citací děl v poli zdroj nalezených informací (pole 670 MARC21/Autority), nebo je dokonce dostupná kompletní bibliografie autora. U autorů odborné literatury považují takový soupis děl za nejpřesnější vlastnost určující autoritní entitu. Problémem je, že citace děl, tak jak je nyní v českých autoritních záznamech uváděna, není přesným propojovacím prvkem, neboť má velmi různorodý způsob zápisu, který se mění v průběhu dnes již více než desetileté existence databáze (první práce zahájeny v roce 1996) a i současné metodické pokyny Národní knihovny se na tento údaj (v poli 670 \$a) dívají pouze pohledem pravidel AACR2 (např. „připouští možnost zápisu citace ve formě Jeho: Toppics in the theory of generative grammar“) [LUFFER, 2006. s. 25].

V letošním roce byla z popudu Národní knihovny ČR a jejich digitalizačních aktivit zahájena jednání o volbě jednoznačného identifikátoru díla pro potřeby identifikace digitalizovaných děl [Národní knihovna České republiky. 2007] V rámci jednání o trvalém identifikátoru je též uvažováno o jednoznačném identifikátoru záznamů v souborném katalogu. Tento návrh má několik problémů směrem k využití k identifikaci digitalizovaných děl (v tomto směru se přikláním spíše k identifikaci na úrovni digitalizované jednotky, neboť digitalizovaná kopie je vždy obrazem konkrétního fyzického díla), ale jeho naplnění by přineslo výhody právě směrem k propojování bibliografických a autoritních záznamů a možnostem je sdílet.

I když projekt VIAF považuje ISBN za „přesnou vlastnost“, z katalogizační praxe víme, že se nemusí vždy jednat o jednoznačnou identifikaci, protože publikace může mít uvedeno ISBN chybně, duplicitně vůči jiné publikaci, ze záznamů pokračujících děl se ISBN může zcela vypustit a mnoho knih, jejichž záznamy katalogy knihoven obsahují, ISBN nemá (v ČR systém zaveden v roce 1989). Vytvořením jednoznačného identifikátoru provedení díla (FRBR terminologie: manifestation) v rámci souborného katalogu by celou akci zpřesnilo.



Obrázek naznačuje možnosti vytváření automatických vazeb mezi autoritní a bibliografickou bází dat, které vzniknout automaticky jako výstup „čistění rejstříků“.

Knihovna, která bude ve svém lokálním systému mít uložena díla tohoto autora, naimportuje do svého lokálního systému autoritní záznam a na základě uložených vazeb se automaticky v rejstříku vytvoří správné vazby na díla, jejichž ID je v autoritním záznamu uloženo.

Naznačené řešení je pouze teoretickým návrhem, který nebyl testován. Na první pohled se otevírá např. otázka zvládnutí vazeb mezi bibliografickými záznamy a autoritními záznamy v případě, že se na autoritní záznam navazuje vysoký počet bibliografických záznamů (např. autoři článků v denním tisku). Pravděpodobně by bylo nutné pro každý automatizovaný knihovnický systém, ve kterém „čistění“ a import záznamů s vazbami na bibliografické záznamy probíhá, vytvářet vlastní softwarové prostředky pro podobnou funkci.

Druhým perspektivním okruhem využití potenciálu autoritních souborů je spolupráce s dalšími typy paměťových institucí a propojování autoritních souborů jako soubor ontologií na další nebibliografické databáze – viz propojení na Wikipedii [BARTL, 2006].

V České republice se podporou spolupráce paměťových institucí za pomoci využití autoritních záznamů uložených v databázi zabývá systematicky firma Cosmotron [ANDREJČÍKOVÁ, 2006]. Vzhledem k tomu, že se jedná o firmu, která úzce spolupracuje s koordinačním centrem v Národní knihovně ČR lze předpokládat, že tato možnost rozvoje nebude opominuta. V tomto směru jsou jistě inspirativní výsledky projektu LEAF. Domnívám se, že zásadním rozhodnutím pro další rozvoj českého souboru bude muset být změna koncepce, a to ve smyslu, zda se budeme na český národní auto-

ritní soubor dívat pouze zorným polem ověřování záhlaví či jeho působení rozšíříme i směrem k faktografickému popisu.

Jaké možnosti se nám pak otevřou nám velmi pěkně ilustruje např. OPAC Střeďočeské vědecké knihovny Kladno [<http://ipac.svkk1.cz>], který v záznamech integruje pole vyplňovaná podle standardů pro katalogizaci i pole faktografická. Problémem však zůstává nekoordinovanost uložení tagování těchto polí na národní úrovni a tudíž nízká míra automatizovaného sdílení těchto databází institucemi, které nemají knihovní systém Advanced Rapid Library. Uživatelský přínos takových databází je jednoznačný. Znamená to však, že knihovny rozšíří svou činnost novým směrem a nabídnou svým uživatelům zcela nové služby. Nelze tedy zakončit můj článek jinak, než citátem ředitele Městské knihovny v Praze Tomáše Řeháka, tj. větou, která musí být leitmotivem naší každodenní práce, tedy i v oblasti katalogizace: „*Existují dva druhy knihoven: ty, které se změňí, a ty, které zaniknou.*“

Literatura:

ANDREJČÍKOVÁ, Naděžda. 2006; ŠAFRÁNEK, Josef M. : *Projekt webu českých památek : [prezentace přednášky]* [online]. Praha : Konference Archivy, knihovny, muzea v digitálním světě 2006, 2006 , 16.04.2007 [cit. 2007-07-07]. Text v češtině. Dostupný z WWW: <http://skip.nkp.cz/KeStazeni/Archivy06/Andrejci-kova.ppt>.

BARTL, Zdeněk. 2006. Authority a Wikipedie našly společnou řeč. *Ikaros* [online]. 2006, roč. 10, č. 11 [cit. 2007-07-17]. Dostupný na World Wide Web: <http://www.ikaros.cz/node/3669>. URN-NBN:cz-ik3669. ISSN 1212-5075.

BENNETT, Rick.2006; HENGEL, Christina; HICKEY, Thomas, O'NEILL, Edward T.;TILLETT, Barbara B. Virtual International Authority File [online]. In: *ALA Annual Conference, 24 June 2006, New Orleans, Louisiana (USA)*. Huron Chicago, American Library Association, 2006 [cit. 2006-12-12]. Dostupné na WWW: <http://www.oclc.org/research/projects/viaf/ala2006c.ppt>

Consortium LEAF. 2001. *Linking and Exploring Authority Files : home* [online]. 2001- [cit. 2006-04-01]. Dostupné na WWW: <http://crxnet.com/leaf/index.html>.

Encoded Archival Context Working Group. 2004. *Encoded Archival Context : beta* [online]. Charlottesville : EAC Working Group, 2004, 2004-08-10 [cit. 2006-03-20]. Dostupné na WWW: <http://www.iath.virginia.edu/eac/>.

IFLA Working Group on Functional Requirements and Numbering of Authority Records (FRANAR). 2007. *Functional Requirements for Authority Data : A Conceptual Model* [online]. 2007, Draft 2007-04-01 [cit. 2007-07-07]. Text v angličtině. Dostupný z WWW: <http://www.ifla.org/VII/d4/FRANAR-ConceptualModel2nd-Review.pdf>.

KAISER, Max, et al. 2003. New ways of sharing and using authority information : the LEAF project. *D-Lib Magazine* [online]. 2003, vol. 9, no. 11 [cit. 2006-03-02]. Dostupné na WWW: <http://www.dlib.org/dlib/november03/lieder/11lieder.html>. ISSN 1082-9873.

LUFFER, Jan. 2006. Metodika tvorby a kontroly jmenných autorit ve formátu MARC21 : Oobní jména. 1. vyd. Praha : Národní knihovna ČR, 2006. 38 s. Standardizace; sv. 28. ISSN edice 1211-7366. Dostupný z WWW: http://www.nkp.cz/pages/page.php3?page=fond_Marc21_osobni-jmena_autority.htm. ISBN 80-7050-507-9.

Národní knihovna České republiky. 2007. Vítejte na portálu pracovní skupiny pro perzistentní identifikátory [online]. Praha : NK ČR, 2007- [cit. 2007-07-07]. Text v češtině. Dostupný z WWW: <<http://pid.ndk.cz/>>.

TILLET, Barbara. 2001. Authority Control on the Web. In *Proceedings of the Bicentennial Conference on Bibliographic Control for the New Millennium : November 15-17, 2000* [online]. Washington : Library of Congress, 2000 [cit. 2006-03-12]. Dostupné na WWW: http://www.loc.gov/catdir/bibcontrol/tillet_paper.html.

WEBER, Jutta. 2004. LEAF: Linking and Exploring Authority Files. *Cataloging and classification quarterly*. 2004, vol. 38, no. 3/4, s. 227-236. ISSN 0163-9374.

ZILLHARDT, Sonia 1998; BOURDON, Françoise. *Author project : final report : report for publication*. Paris : Bibliotheque nationale de France, 1998, 38 s.