

ÚSKALÍ TVORBY A VYUŽÍVÁNÍ BÁZE JMENNÝCH AUTORIT

Mgr. Helena Dvořáková
Národní knihovna ČR

První etapa budování báze národních autorit je konečně za námi. Knihovnické i neknihovnické veřejnosti nabízí Národní knihovna ČR na internetu (<http://omega.nkp.cz:4001/ALEPH/CZE/NKC/CLA/CLA/START>) přes 60 000 záznamů českých i zahraničních autorů či jiných původců dokumentů s revidovanou formou záhlaví, v převážné většině s charakteristikami a odkazy. Prvních cca 34 000 záznamů českých autorů s biografickými daty (tzv. „definitivních“ záhlaví, i když co je definitivní...) je k dispozici na CD-ROM Česká národní bibliografie, v nejbližší době budou veškeré záznamy také na FTP serveru NK. Zájemci tedy již mohou převzít tato data pro své vlastní báze autorit. Pro efektivní využití záznamů autorit by bylo dobře, aby se knihovníci seznámili jednak s jejich vznikem a strukturou, jednak se všemi pozitivními i negativními důsledky využívání těchto dat při budování bibliografické báze. Hlavně proto vznikl tento článek.

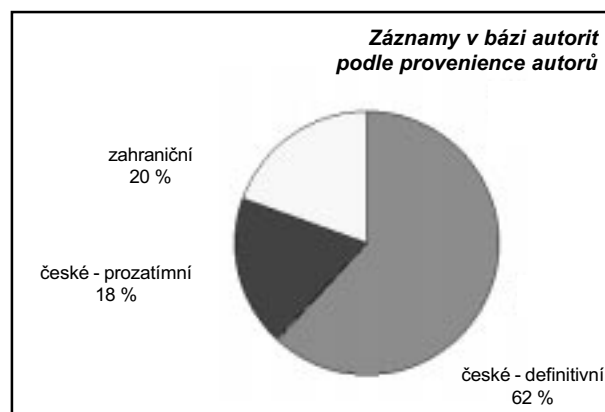
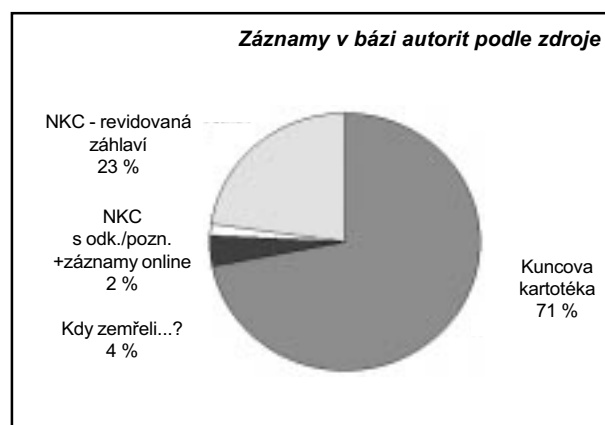
Když jsme v NK práci na bázi autorit zahájili, netušili jsme, kolik problémů a jakých se v jejím průběhu vyskytne. Dost nám zkomplikoval život fakt, že funkčnost báze autorit připojené k bibliografické bázi v současné verzi systému ALEPH je omezena na využití informací zde obsažených pro vyhledávání a že záhlaví není možno stahovat do záznamů. Byli jsme tedy nuceni ponechat bázi autorit (AUJ) zcela samostatnou a zajistit dávkové přenášení údajů z této báze do přístupových souborů bibliografické báze (NKC) a naopak pomocí speciálních programů.

Tím, že NK nepracuje s regulérní připojenou bázi autorit, nachází se při využívání dat z báze autorit v bibliografické bázi v podobné situaci, v jaké budou všechny ostatní knihovny, které se rozhodnou převzít data z báze autorit. Naše zkušenosti ukazují, že počátky nejsou lehké. Přes veškerý svůj přínos standardizaci a urychlení katalogizace nepřinese báze autorit knihovnám tolik, kolik si od ní zřejmě slibují. Knihovníci by měli počítat s tím, že v prvním stadiu tu bude spíše hodně práce navíc, výhody využívání souborů autorit se projeví teprve po určité době. Je zároveň třeba upozornit, že některé problémy spojené s bázi autorit budou společné pro všechny knihovnické systémy, jiné se projeví jen v některých, další budou pro určitý systém specifické.

Jak se v NK buduje báze autorit

Bázi autorit AUJ je možno v zásadě plnit třemi způsoby: dávkovými importy dat získaných z vnějších zdrojů, importem záznamů vytvořených transformací záhlaví, odkazů a poznámek z bibliografické báze a konečně vytvá-

řením záznamů online. Prvním způsobem bylo vytvořeno „jádro báze“, cca 45 000 záznamů z Kuncovy kartotéky, v nedávné době přibyly záznamy ze soupisů „Kdy zemřeli...?“, které bylo nutno důkladně „profiltrovat“ vůči stávajícím záznamům, aby nevznikly duplicity (čemuž se ovšem zcela vyhnout pochopitelně nepodařilo). Z báze NKC se speciálními programy převádějí jednak všechna záhlaví s odkazy a poznámkami, jednak revidovaná záhlaví s biografickými daty, opět samozřejmě s kontrolou na duplicitu. Tvorba nových záznamů v bázi AUJ online byla z důvodů problémů s oboustranným přenosem dat zmiňovaným dále omezena na nezbytné případy rozdělování původně shodných záhlaví, kdy se však zároveň vytvoří nová záhlaví i s příslušnou poznámkou či odkazem v přístupovém souboru NKC. Data z báze AUJ získaná z vnějších zdrojů se jiným programem poté přenášejí do přístupového souboru báze NKC, aby zde mohla být využívána. Pokud v NKC totožné záhlaví již existovalo, připojí se k němu odkazy a poznámky.



Již na počátku práce se ukázala potřeba opravovat a doplňovat již existující záhlaví. Postupem času by k těmto změnám mělo docházet jen výjimečně (např. při úmrtí autora), ale v době prvotní redakce báze autorit se změny vyskytují dosti často. Protože záhlaví v přístupových souborech nemají svá systémová čísla, není možné změnu provedenou v záznamu báze AUJ automaticky promítnout do záhlaví v bázi NKC: přenesením údajů by vzniklo heslo nové, staré by bylo třeba zrušit a záznamy na něj napojené převést na heslo nové. Tím by se sice provedla automatic-

ká změna v bibliografických záznamech, ale je s tím spojeno také mnoho ruční práce a tudíž riziko chyby. Zvolili jsme proto praxi, že veškeré změny se provádějí paralelně na obou místech (do AUJ se po základním importu přenášejí již pouze nově vzniklá, nikoli změněná záhlaví). Tuto komplikaci však považujeme pouze za dočasnou, protože nová verze systému ALEPH (ALEPH 500) by již měla mít regulérní interaktivně spolupracující bázi autorit.

Proč záznamy v bázi autorit nejsou ve standardním formátu UNIMARC/Autority

Základem báze AUJ se staly záznamy z Kuncovy kartotéky, které při tagování byly převedeny do formátu UNIMARC/Autority bez jakýchkoli úprav - pouze biografická poznámka byla vzhledem ke svému rozsahu, netypickému pro zahraniční autoritní báze, uložena do zvláštního „národního“ pole 907. Aby mohla být vytvořena vazba na záznamy v bázi NKC, kde se ověřená záhlaví označují písmenem *r* v přidaném nestandardním podpoli *t*, a záznam při tom zůstal korektní, byla tato rozšířená forma záhlaví uložena navíc do pole AU. Když však byla část těchto dat pokusně přenesena programem do přístupového souboru báze NKC, ukázaly se problémy, na něž se v online budované bázi autorit většinou nenarazí, protože záznamů je relativně málo a shody ve jménech nenastávají příliš často; pokud k nim dojde, ihned se projeví a nežádoucím efektem se může zabránit již v zárodku. Důsledkem bylo, že ani další pole nemohla zůstat v korektním tvaru. Šlo zejména o dva problémy: odkazy a nerozlišená záhlaví.

Odkazy

Systém ALEPH pracuje s odkazy typu *viz* tím způsobem, že katalogizátorovi, který by nepřebíral záhlaví z přístupového souboru, nedovolí použít vyloučený tvar jména - při uložení ho převede na preferovanou formu

Příklady:

```
AU   $$a Bukovský $$b Adolf Vojtěch $$f 1889-1953 $$t r
200-1 $$a Bukovský $$b Adolf Vojtěch $$f 1889-1953
400-0 $$a Ave $$f 1889-1953 $$t o
AU   $$a Stoøes $$b František $$f 1895-1942 $$t r
200-1 $$a Stoøes $$b František $$f 1895-1942
400-0 $$a Ave $$f 1895-1942 $$t o
```

(Rozlišení stejných pseudonymů)

```
AU   $$a Plaidy $$b Jean $$f 1906- $$t r
200-1 $$a Plaidy $$b Jean $$f 1906-
400-1 $$a Holt $$b Victoria $$f 1906- $$t o
400-1 $$a Carr $$b Philipa $$f 1906- $$t o
```

(Autorka publikující pod více pseudonymy)

```
AU   $$a Cimrman $$b Jára da $$t r
200-1 $$a Cimrman $$b Jára da
500-1 $$a Smoljak $$b Ladislav $$f 1931- $$t r
500-1 $$a Svěrák $$b Zdeněk $$f 1936- $$t r
500-1 $$a Šebánek $$b Jiří $$f 1930- $$t r
```

(Kolektivní pseudonym)

(obdobným způsobem může zpracovat i importované záznamy, buď přímo při importu, nebo po dávkovém znovuvytvoření přístupového souboru). Proto systém pochopitelně nepřijme stejný vyloučený tvar směřující na dvě různá záhlaví - odkaz *viz* musí být jednoznačný.

V záznamech převáděných z báze autorit se objevily následující problematické případy:

1. několik různých autorů (uvedených pod vlastním jménem nebo nejčastějším pseudonymem) používalo stejné pseudonymy, které byly uvedeny v poli 400
2. několik autorů publikovalo pod kolektivním pseudonymem
3. dva pseudonymy použité jako samostatná záhlaví měly tentýž vyloučený tvar (např. vlastní jméno, pod nímž autor nepublikuje)
4. pseudonym jednoho autora byl totožný s vlastním jménem jiného autora

Ukázalo se, že v čisté „unimarcové“ podobě systém následující případy nemůže řádně zpracovat: první tři případy odmítá zařadit, v posledním případě by dokonce došlo ke spletení a záměně různých autorů. Proto bylo nutno vyloučené tvary ošetřit tak, aby nemohlo k záměnám ani rozporům docházet, i když se tím ztratila možnost automatických oprav v bázi (při důsledném přebírání záhlaví z přístupových souborů však ani nejsou nutné). Veškeré odkazované tvary z pole 400 byly doplněny o biografická data ze záhlaví (podpole *f*) a písmeno *o* do pomocného podpole *t* k označení odkazu; pro výjimečné případy, kdy by biografická data chyběla a vyskytl se vícenásobně stejný odkaz, bylo rozhodnuto do pomocného podpole *s* převést jméno autora z pole 200 (v invertovaném tvaru). Pro kolektivní pseudonymy je určen odkaz *viz též* (pole 500), který se jinak používá spíše výjimečně - při více pseudonymech se stává záhlavím ten tvar jména, pod nímž autor publikoval nejčastěji a ostatní se uvádějí do pole 400.

Nerozlišená záhlaví

V Kuncově kartotéce byl poměrně častý také vícenásobný výskyt stejnojmenných autorů, u nichž se nepodařilo zjistit biografická data. Když byly záznamy těchto autorů pokusně přeneseny do báze NKC, pod jedním záhlavím se logicky spojily všechny poznámky či odkazy; v některých případech bylo v NKC již před importem stejné revidované záhlaví, a pak se připojila nejen jeho poznámka či odkaz, ale i bibliografické záznamy (v současném systému ALEPH je záhlaví jediným propojovacím údajem mezi bází autorit a přístupovým souborem bibliografické báze). Aby k tomuto nedocházelo, bylo nutné stejnojmenné autory rozlišit. Po dlouhých úvahách bylo rozhodnuto pro tyto účely použít pomocné podpole *s*, do něhož se zapisuje stručná profesní charakteristika (nelze-li ji vytvořit, uvádí se alespoň obor, v němž autor publikuje). Tyto záznamy dostávají status prozatímní, protože je možné, že biografická data budou zjištěna dodatečně, a také proto, aby se knihovny mohly rozhodnout, zda je ve svých bázích chtějí využít (na první CD-ROM záměrně nebyly zařazeny). Nelehké rozlišovací práce dosud probíhají, ale do budoucna by se tvorba prozatímních záhlaví měla omezit na minimum, protože s sebou nese i riziko vzniku dalších duplicit. Jen pro ukázkou, jaké otázky tvůrci záhlaví musí řešit: Je vhodnější užít označení pedagog nebo učitel? Jak např. označit autora, u něhož je v poznámce uvedeno „autor příruček pro podnikatele“? Co s tím, když je jedna osoba činná ve více oborech? Neuděláme pro jednoho a téhož autora více záhlaví omylem jen proto, že to nevíme? Stalo se to prý při prvním vydání slavného slovníku Larousse: ruský básník a vědec Lomonosov byl uveden pod dvěma hesly, navíc s poznámkou „Nezaměňovat se stejnojmenným básníkem/vědcem“.

Příklady záhlaví:

```
$$a Beránek $$b Václav $$$ dramatik
$$a Beránek $$b Václav $$$ ilustrátor
$$a Beránek $$b Václav $$f 1785-1864 $$t r
$$a Beránek $$b Václav $$f 1877-1965 $$t r
```

S čím musí počítat knihovna, která bude chtít s bází autorit NK ČR pracovat

V okamžiku importu záznamů autorit do báze autorit již bude mít každá knihovna v této bází své záznamy s ověřenými či neověřenými záhlavími (hesly), která byla vytvořena v procesu katalogizace jejich dokumentů nebo která si vytvořila pro potřeby katalogizace sama. Určitá část z nich budou záznamy týchž autorů, což znamená, že prostý import není možný. Tzv. matching neboli porovnání na duplicitu a následné spojení záznamů či vyloučení duplicit není jednoduchá záležitost; jak dosvědčují zkušenosti ze souborných katalogů budovaných „dostředivou metodou“, značné množství duplicit v důsledku chybovosti záznamů vždy unikne, ať je porovnávací klíč postaven sebelépe. Pokud se záznamy „potkají“, tzn. že autoritní forma záhlaví byla stejná, může být „domácí“ záznam zcela přepsán, nebo jen obohacen o odkazy, biografickou poznámku (což katalogizátoři jistě ocení) a identifikační

číslo báze národních autorit (může sloužit k pozdějšímu propojení). Co však v případech, že „domácí“ záznam již nějaké odkazy měl? Některé mohou být totožné, jiné mohou být navíc, není vyloučena ani eventualita, že v domácí bázi byla zvolena varianta *viz též* a v NK *viz* či naopak. S tím se však ještě program může vypořádat a problematické odkazy vyčlenit k dořešení. Neodhalí však případy, kdy půjde o záznam stejného autora, ale v jednom případě byla uvedena biografická data a v druhém ne, popř. měla jinou formu, chybělo datum úmrtí ap. Pak vzniknou záznamy duplicitní, které budou působit katalogizátorům při přejímání autoritní formy záhlaví potíže. Tyto duplicity se odhalí náhodně při katalogizaci, jinak se dají zlikvidovat jen při důkladné „hlavuruční“ redakci báze. (Kolik dalších duplicit programem neodstranitelných vzniká v důsledku běžných překlepů, o tom je lépe se nezmiňovat.)

Záznamy autorit, které program uznal jako neduplicitní, se naimportují do báze autorit a mohou okamžitě sloužit při katalogizaci, takže všechny nové záznamy budou mít korektní formu záhlaví. V bázi je však jistě velké množství záznamů, které stojí pod neověřeným a blíže nespecifikovaným jménem autora, záznamů, které patří k různým autorům téhož jména. Ty by měly být nyní „rozpleteny“ a přiřazeny ke správným autoritním záhlavím, a to za knihovny také žádný program neudělá. Kromě toho, že půjde o mravenčí práci, je tu ještě jeden problém. Knihovník v tomto okamžiku již nemá v ruce knihu, která by mohla mít na záložce či v doslovu nějaké bližší údaje, takže se bude těžko rozhodovat a asi se snadno zmýlí, i když bude mít k dispozici u každého záhlaví biografickou poznámku. Jak bude vypadat přístupový soubor po importu záhlaví z báze autorit demonstruje následující ukázkou.

Řádek č. Seznam Autoři

| | | |
|-----|----|------------------------------------|
| I1 | * | Rolář, Jaromír, 1926- |
| I2 | 50 | Rolář, Jaroslav |
| I3 | * | Rolář, Jaroslav ekonom |
| I4 | * | Rolář, Jaroslav stavební inženýr |
| I5 | * | Rolář, Jaroslav, 1894-1961 |
| I6 | * | Rolář, Jaroslav, 1910- |
| I7 | * | Rolář, Jaroslav, 1911-1979 |
| I8 | * | Rolář, Jaroslav, 1932-1979 |
| I9 | 1 | Rolář, Jaroslav, 1937- |
| I10 | 35 | Rolář, Jiří |
| I11 | * | Rolář, Jiří beletrista |
| I12 | * | 16 Rolář, Jiří hudebník |
| I13 | * | Rolář, Jiří lékař |
| I14 | * | 86 Rolář, Jiří, 1914- |
| I15 | 74 | Rolář, Josef |

Jestliže knihovna pracuje pouze systémem katalogizace online, dříve či později autoritní bázi a s ní přístupový soubor vyčistí, i když katalogizátor bude mít leckdy problém rozhodnout se, ke kterému záhlaví dílo přiřadit, a jedno „sběrné“ nespecifikované záhlaví pro složitější případy asi stejně zůstane zachováno. Pokud však knihovna importuje data (např. při retrokonverzi skenováním a ta-

gováním, dávkovým stahováním ze souborného katalogu ap.), při každém importu se dostanou do báze záznamy s nespecifikovanými záhlavími, které by se měly ručně „rozplést“. Je to nikdy nekončící práce a než se knihovníkům podaří projít celou bázi, je tu nejspíš další import. Určitým řešením, k němuž jsme se rozhodli v případě větších importů přistoupit do budoucna i v NK ČR, je „předčištění“ importovaných dat v pomocné bázi, která má připojena ověřená záhlaví - přece jen se lépe pracuje s 20 000 záznamy než s půl milionem. Má to však i své nevýhody, záznamy jsou čtenářům k dispozici s určitým časovým skluzem.

Důsledné využívání báze autorit usnadní knihovníkům i uživatelům řešerše na díla konkrétních autorů, nebude třeba provádět složité filtrování a nevyhledají se irelevantní záznamy. Ovšem tak jednoduché to zase není. Ve většině případů uživatel nezná biografická data a musí zjistit, který z oněch např. deseti Janů Nováků je ten hledaný. Pokud systém nabízí přímý přístup z bibliografické báze k biografickým poznámkám v autoritní bázi (což je bohužel u současné verze systému ALEPH možné jen při „pokročilém“ vyhledávání a ve webovském přístupu to není možné vůbec), může uživatel najít toho „svého“ autora otvíráním této poznámky, jinak ovšem musí postupně prohledat všech deset záhlaví. Záznamy pod nerozlišeným záhlavím bude muset pro jistotu prohlednout knihovník i uživatel v každém případě. V případě, že systém zobrazuje uživateli všechna hesla z autoritní báze včetně těch, která nemají připojené žádné záznamy, stává se hledání v přístupovém souboru značně nepřehledným (uvědomme si, že řada knihoven vzhledem ke svému profilu díla od mnohých autorů uvedených v bázi autorit nebude mít ve své bázi nikdy). NK ČR tento problém vyřešila tím, že čtenářům zpřístupňuje logickou bázi, v níž se vyskytují pouze „obsazená“ záhlaví, zatímco katalogizátoři mají přirozeně k dispozici úplné přístupové soubory.

Dalším úskalím, které u některých systémů může čtenáři zkomplikovat práci s bázi obsahující velké množství

stejnomených autorů odlišených daty a/nebo přívlastky, je vnitřní řazení přístupového souboru. Pokud je celé záhlaví řazeno jako jeden řetězec bez ohledu na podpole, jako v případě systému ALEPH, mohou se stejnojmenní autoři dostat i poměrně daleko od sebe (někdy i na další stránku seznamu), takže uživatel ani nemusí zjistit, že od toho „jeho“ autora nějaká díla v bázi máme. Viz následující ukázka. (Poznámka k ukázce: čtenářům se záměrně nezobrazuje interní podpole *t*; jeho obsah je však bohužel brán v úvahu při řazení.)

| | | |
|----------|---------------|-------------------------------------|
| I2 | 75 | Novák, Jan |
| I3 | 3 | Novák, Jan Bedřich |
| I4 | * | Novák, Jan Bedřich, 1872-1933 |
| I5 | * | Novák, Jan Bohuslav, 1840-1919 |
| I6 | * | 2 Novák, Jan Viz: Janovský, Jiří |
| I7 | 4 | Novák, Jan práce pro děti |
| I8 | 1 | Novák, Jan V. |
| I9 | 2 | Novák, Jan Václav |
| I10 | * | Novák, Jan Václav, 1853-1920 |
| I11 | 8 | Novák, Jan Zdeněk |
| I12 | * | Novák, Jan, 1857-1907 |
| I13 | * | Novák, Jan, 1866- |
| ----- | | |
| Řádek č. | Seznam Autoři | |
| ----- | | |
| I1 | * | Novák, Jan, 1876-1938 |
| I2 | * | Novák, Jan, 1881-1929 |
| I3 | * | Novák, Jan, 1887- |
| I4 | 4 | Novák, Jan, 1953- |
| I5 | 5 | Novák, Jaromír |

Závěrem

Nechtěla bych, aby můj článek vyzněl v tom smyslu, že budovat bázi národních autorit je mrhání skromnými finančními prostředky knihoven a že přebírat její data do vlastních databází způsobí jenom problémy. Je samozřejmě žádoucí, aby konkrétní autor stál ve všech bázích pod jedinou formou záhlaví a, je-li to možné, aby byl odlišen od jiných stejnojmenných autorů. Ideální by bylo, kdyby společně využívaná báze autorit byla postupně i doplňována kooperativním způsobem, jsem však trochu skeptik, pokud jde o technické řešení problému a především ochotu knihoven spolupracovat. I když k ideálnímu stavu máme daleko, je tu přinejmenším jakýsi elektronický biografický slovník, který může sloužit katalogizátorům jako zdroj správných záhlaví a zároveň jako zdroj informací pro neknihovnické zájemce - ani to však není málo.

Jsem přesvědčena, že myšlenka celosvětové báze autoritních záhlaví má budoucnost; a je dobře, že budeme patřit k těm zemím, které se do jejího budování zapojí v prvním sledu. Při práci s bázi autorit jen občas trochu zapochybuji, zda by někdy méně nebylo více - zda opravdu ke každému autorovi doslovu či např. jediného spisu o svařování mají být doplňována biografická data či zda je vhodné rozlišovat přívlastkem i v případech, kdy biografická data nemáme. Mají být i tyto osoby součástí národní báze autorit a tím v budoucnu i báze celosvětové? Ovšem co je měřítkem významnosti autora?

Nakonec ještě k mému oblíbenému tématu souborných katalogů. Neměli bychom vzbuzovat iluzi, že vytvořením báze autorit rázem postavíme na vyšší kvalitativní úroveň národní souborný katalog. Zatímco pro souborný katalog s katalogizací v centrální databázi je přínos autoritní báze neoddiskutovatelný, v souborném katalogu budovaném převážně z dávkových importů dat může sloužit jen menšině knihoven katalogizujících online a potom jako pomůcka pro pracovníky spravující přístupové soubory - se všemi problémy, které jsem zmínila. Kvalita souborného katalogu se zvýší jen do té míry, o co kvalitnější a jednodušší bude katalogizace v jednotlivých účastnických knihovnách.