

Authority v Evergreenu aneb Automatizace především

Mgr. Livia Vrzalová
Knihovna Jána Langoše
livia.vrzalova@ustrcr.cz

PhDr. Linda Jansová, Ph.D.
pražská organizace SKIP, Národní knihovna ČR
linda.jansova@gmail.com

Václav Maixner
Matematicko-fyzikální fakulta Univerzity Karlovy
vaclavmaixnerj@gmail.com

Seminář Evergreen v českých knihovnách 2017
9. 10. 2017, Národní knihovna ČR

Nastavení práce s autoritami v Evergreenu

- **autoritní záznamy jsou přiřazovány k tezaurům a tezaury do větších skupin (tzv. control setů)**
- **je výhodné používat přednastavenou skupinu označenou zkratkou LoC**
- **ta přesně odpovídá specifikaci formátu MARC 21 a jsou pro ni připraveny i další nástroje**

Nastavení vlastní skupiny

- **nastavení je možné, je však třeba počítat s tím, že:**
 - aktuální verze nástroje pro automatické propojování bibliografických a autoritních záznamů pracuje pouze s přednastavenou skupinou
 - jako hodnota (zkratka) tezauru se uvádí velké písmeno, což není v souladu se specifikací formátu MARC 21

Pozice 11 v poli 008

11 - Subject heading system/thesaurus

- a - Library of Congress Subject Headings
- b - LC subject headings for children's literature
- c - Medical Subject Headings
- d - National Agricultural Library subject authority file
- k - Canadian Subject Headings
- n - Not applicable
- r - Art and Architecture Thesaurus
- s - Sears List of Subject Heading
- v - Répertoire de vedettes-matière
- z - Other
- | - No attempt to code

Zdroj: <http://www.loc.gov/marc/authority/ad008.html>, získáno 7. 10. 2017

Nastavení tezaurů

- **tezaury v Evergreenu odpovídají předmětovým heslářům / tezaurům uváděným na pozici 11 v poli 008**
- **výjimkou jsou případné další tezaury (ty jsou označovány velkým písmenem)**

Propojení s poli v bibliografických záznamech

- **je třeba zkontrolovat (popř. doplnit) nastavení polí z bibliografických záznamů, která mají mít vazbu na záznamy autoritní**

Dostupnost souboru českých národních autorit

- je k dispozici na serveru Národní knihovny ČR v řadě variant
- zatím pracujeme především se souborem všech jmenných i věcných autoritních záznamů (s výjimkou záznamů smazaných) ve formátu MARCXML
- aktualizovaná verze tohoto velkého souboru (po rozbalení cca 1,6 GB) je vystavována jednou měsíčně

Předzpracování dat

- **odstranění písmenkových tagů**
- **s ohledem na dostupné hardwarové vybavení:**
 - předvýběr (potenciálně) relevantních autoritních záznamů a jejich import (podle údajů v příslušných polích, resp. podpolích bibliografických záznamů)
 - import všech záznamů

Možné způsoby importu

- **jednotlivě nebo hromadně:**
 - přes příkazovou řádku
 - prostřednictvím služebního klienta
- **z praktických důvodů je vhodnější hromadný import přes příkazovou řádku**

Identifikátory

- **při importu Evergreen autoritnímu záznamu přiřadí interní identifikátor a uloží se do pole 001**
- **původní identifikátory (trvalé identifikátory českých národních autorit) jsou uchovány v:**
 - poli 016 (v podpoli a)
 - v poli 035 (v podpoli a, a to s prefixem CZ-PrNK v závorce)
 - v podpolích 7 (v různých polích, např. u autorů - osob v poli 100)
- **identifikátor přidělený Evergreenem je ještě v poli 901c**

Aktualizace autoritních záznamů

- **je třeba počítat s tím, že v souboru českých národních autorit dochází v průběhu času ke změnám:**
 - přibývají nové autoritní záznamy
 - dochází k úpravám existujících autoritních záznamů
 - dochází ke smazání (např. duplicitních) záznamů

Projekt v rámci programu VISK 3

- **název:**

- Zkvalitnění služeb Knihovny Jána Langoše poskytovaných prostřednictvím online katalogu Evergreen

- **realizace:**

- Ústav pro studium totalitních režimů – Knihovna Jána Langoše
- rok 2017

- **výstupy využijí i další knihovny s Evergreenem**

Realizační tým

- **Livia Vrzalová**
- **Václav Maixner**
- **Eva Cerniňáková**
- **Václav Jansa**
Linda Jansová

Hlavní činnosti

- **příprava testovacího prostředí (Evergreen 2.12)**
- **program(y) pro aktualizaci autoritních záznamů (v jazyce Python)**
- **testování**
- **úprava a další testování**
- **kombinace s ostatními programy - skript automaticky spouštěný v určité době**

Hlavní činnosti

- **vytvoření (programátorské i uživatelské) dokumentace**
- **zveřejnění zdrojového kódu i dokumentace na dokumentačním portálu české komunity kolem Evergreenu: <https://eg-wiki.jabok.cz/>**

Výchozí stav na příkladu Společného katalogu Evergreen

- **v databázi Evergreenu jsou:**
 - dříve nainportované autoritní záznamy
 - bibliografické záznamy zatím bez provázání na autoritní záznamy, ale obsahující podpole 7 s identifikátorem autoritního záznamu z Národní knihovny ČR
 - bibliografické záznamy bez provázání na autoritní záznamy a bez podpole 7

Export záznamů z databáze Evergreenu

- **bibliografické záznamy**
- **autoritní záznamy**

Stažení souboru s autoritními záznamy

- **ze serveru NK**
- **obsahuje všechny platné jmenné i věcné autoritní záznamy**
- **neobsahuje záznamy, které byly zcela smazány/zrušeny**

Odstranění písmenkových tagů

- **autoritní záznamy obsahují pole typu MOR, COR apod.**
- **Evergreen počítá pouze se záznamy podle specifikace formátu MARC 21 s tagy tvořenými vždy třemi číslicemi**

Vyhledání aktualizovaných autoritních záznamů

- **program najde:**

- v autoritních záznamech vyexportovaných z Evergreenu identifikátory z Národní knihovny ČR (pole 035, podpole a)
- v souboru z Národní knihovny ČR záznamy odpovídající těmto identifikátorům

Uložení výsledků hledání

- **program vytvoří:**
 - soubor s identifikátory autoritních záznamů z Národní knihovny ČR a aktualizovanými autoritními záznamy ve formátu MARCXML
 - soubor obsahující zbývající (nepoužité) autoritní záznamy

Vyhledání autoritních záznamů podle hodnot podpolí 7

- **program:**

- v exportu bibliografických záznamech z Evergreenu najde (v podpolích 7) identifikátory autoritních záznamů z Národní knihovny ČR, které nejsou součástí exportu autoritních záznamů z Evergreenu
- do souboru uloží všechny takto získané autoritní záznamy, které ještě v Evergreenu nejsou a měly by být naimportovány

Seznam smazaných autoritních záznamů

- **do souboru jsou uloženy:**
 - identifikátory z Národní knihovny ČR, které se vyskytují v Evergreenu, ale už se nevyskytují v souboru s daty z NK
 - odpovídající identifikátory těchto autoritních záznamů z Evergreenu
- **smazání (či kontrola možných vad) se vzhledem k nízkému počtu těchto záznamů provede ručně**
 - jedná se o jednotky záznamů

Výběr dalších záznamů pro import

- do dalšího souboru se uloží zbylé dosud nepoužité záznamy
- v něm se další vhodné záznamy vyhledají na základě řetězců (resp. jejich hashů) z vyexportovaných bibliografických záznamů

Propojování bibliografických záznamů s autoritními

- **ručně:**

- spolehlivé, ale časově velmi náročné

- **automaticky:**

- využití podpůrných skriptů, které jsou součástí Evergreenu:

- `authority_control_fields.pl`

- provázání bibliografického záznamu s nejlépe odpovídajícím autoritním záznamem

- `authority_authority_linker.pl`

- propojení autoritních záznamů mezi sebou

Úprava pro české prostředí

- **přidání podpole 7 (identifikátor Národní knihovny ČR = záruka přesné shody)**
- **(je-li záznam stažen z báze NKC nebo z báze SKC, popř. z jiných českých zdrojů, je velmi pravděpodobné, že podpole 7 obsahuje)**

Tvorba nových autoritních záznamů

- tvorba přímo v Evergreenu
- zaslání do Národní knihovny ČR pomocí programu yaz-client (také otevřený software)

Otázky?

Děkujeme za pozornost!!