

NTK

50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Library of Technology

Czech Academic and Research
Discovery Services
CARDS

Seminář k datům v PNG

Tým pro data

NTK Praha

16. 11. 2023



Spolufinancováno
Evropskou unií

MS
MT
MINISTERSTVO ŠKOLSTVÍ,
MLÁDEŽE A TĚLOVÝCHOVY

Úvod...

...aneb lehké opakování

Axiom Centrální katalogizace

Tvorba fyzického centrálního katalogu s jedinou instancí bibliografického záznamu pro každý dokument.

Snaha o sjednocení číselníků a rejstříků pro maximální přesnost vyhledávání.

- sloučení samostatných databází zapojených knihoven do jedné společné databáze (centrálního katalogu), která obsahuje vše
- sdílení této databáze a dat v ní obsažených všemi knihovnami
- deduplikace záznamů s cílem sloučit stejné záznamy z různých knihoven
- katalogizace probíhá v souladu s platnými standardy
- různé úrovně oprávnění přístupu k záznamům / lokální extenze

Základní standardy pro zpracování záznamů v PNG

Hana Vochozková, MU Brno

Základní standardy zpracování Formát

- MARC 21. Bibliografický formát
- MARC 21. Formát pro autority
- MARC 21. Formát pro exemplářové údaje

- Minimální záznam pro SK ČR

<http://www.nkp.cz/o-knihovne/odborne-cinnosti/zpracovani-fondu/katalogizacni-politika/minimalni-doporucene-zaznamy-rda-marc-21>

- NK ČR - ... - Katalogizační politika – MARC 21

<http://www.nkp.cz/o-knihovne/odborne-cinnosti/zpracovani-fondu/katalogizacni-politika/marc21-soubor>

Základní standardy zpracování

Pravidla

- RDA Toolkit (online verze) <http://www.rdatoolkit.org/> (přístup je placený)
- RDA: Resource Description and Access (tištěná verze)
- Katalogizace podle RDA ve formátu MARC 21 – tištěné a elektronické monografie - katalogizace na úrovni minimálního/doporučeného záznamu

Aktualizace září 2023

<http://www.nkp.cz/o-knihovne/odborne-cinnosti/zpracovani-fondu/katalogizacni-politika/katalogizace-podle-rda-ve-formatu-marc-21-tistene-a-elektronicke-monografie-katalogizace-na-urovni-minimalniho-doporuceneho-zaznamu>

- NK ČR - ... - Katalogizační politika
<http://www.nkp.cz/o-knihovne/odborne-cinnosti/zpracovani-fondu/katalogizacni-politika/standard1>

Deduplikace záznamů

Martin Kybal, NTK Praha

Návrh řešení deduplikace záznamů

- řešení deduplikace pomocí Souborného katalogu ČR
- deduplikační klíč = systémové číslo záznamu z SK
- na základě tohoto klíče se příslušné záznamy z lokálních katalogů nahradí jedním společným záznamem

K tomu je třeba:

- dostat do SK maximum záznamů z jednotlivých knihoven
- pravidelně synchronizovat data knihoven s SK (z důvodu možných změn v lokálním katalogu i SK)
- automaticky doplňovat deduplikační klíč do lokálních katalogů knihoven (a v případě změny ho aktualizovat)
- spustit čištění lokálních katalogů knihoven (bude pokračovat i v PNG)
- připravit nástroje na podporu čištění dat

Pole 993 pro potřeby řízení čištění dat v zapojených knihovnách

Antonín Vaishar, VUT Brno

- data knihovny lze rozdělit do 3 množin:
 - ✓ záznamy odeslané do SK (tyto záznamy jsou v určité míře navázané na záznamy v SK)
 - ✓ záznamy dosud neodeslané do SK
 - ✓ záznamy, které se do SK nebudou posílat
- vznikla tak potřeba vytvořit filtr, který bude separovat jednotlivé množiny záznamů
- a to jak pro knihovny, tak i pro NTK, která bude realizovat kontroly
- s každou ze tří vzniklých množin se bude zacházet úplně jinak

Nové kontrolní pole

- bude obsahovat buď informaci o odeslání do SK, konkrétně systémové číslo navázaného záznamu z SK
- nebo bude obsahovat informaci o tom, že se záznam do SK posílat nebude
- nebo v záznamu nebude, což bude značit, že se záznam do SK má odeslat a bude primárním cílem kontrol a oprav
- úzká návaznost na kontroly, které budou probíhat na straně Týmu pro data, resp. Servisního centra PNG
- zároveň nechceme zasahovat do existujících mechanismů odesílání do SK, což může znamenat duplikaci informací v záznamu

Pole 993

- konzultováno s NK
- proběhlo dotazníkové šetření na obsazenost polí 985 - 999 v zapojených knihovnách
- pole 993 zvoleno, protože ho UK používá k podobnému účelu
- bude k dispozici konfigurace pro Aleph, případně též pro ostatní systémy
- v případě potenciálních kolizí (víme, že systém Clavius pole 993 používá) není problém kontroly v individuálních případech navázat na jiné pole
- v současnosti je hotový základní rámec, struktura pole a slovníkové hodnoty se ještě mohou měnit

Struktura pole

- pole **993** je nepovinné a neopakovatelné
- pole **993** nemá definované žádné indikátory
- podpole **\$a** je nepovinné a neopakovatelné
- podpole **\$m** je nepovinné a opakovatelné
- podpole **\$x** je nepovinné a neopakovatelné
- podpole **\$y** je nepovinné a neopakovatelné

Podpole \$a

- obsahuje systémové číslo navázaného záznamu v SK
- automatické doplnění na základě porovnání exportních souborů (lokální katalog vs. SK, link v poli 910)
- jedná se o objektivní propojení, nikoliv však nutně věcně správné - informace v podpoli nemusí být definitivní

Podpole \$m

- pokud je lokální záznam propojen s více záznamy v SK, např. z důvodu rozdílného způsobu zpracování seriálové publikace v SK a lokálním katalogu, budou tato multiplicitní systémová čísla uložena v opakovatelném výskytu podpole **\$m**
- u takového záznamu musí:
 - ✓ buď dojít k opravě (sjednocení popisu s SK)
 - ✓ nebo musí být vyloučen z deduplikace
- existence podpole **\$m** vylučuje podpole **\$a** a naopak

Podpole \$x a \$y

- obsahují informace o neposlání do SK (\$x) a důvodu (\$y)
- v podpoli \$x je platný text **NEPOSILAT**
- podpole \$y nabývá hodnoty na základě analýzy typů záznamů nevhodných k odeslání do SK
 - ✓ EXEM (exemplářový popis - mapy, historický fond)
 - ✓ VSKP
 - ✓ NOBK (doplňkový sortiment)
 - ✓ SKRYTO (záznam je z nějakého důvodu skryt)
 - ✓ ODPIS
 - ✓ SUBMIN (subminimální záznamy - mělo by se na nich pracovat)
 - ✓ AKVIZICE (akviziční záznamy)
 - ✓ EBOOK (kvůli nežádoucímu slučování s tištěnými verzemi)
 - ✓ EIZ (trvalé nákupy EIZ)
 - ✓ KONVOL (konvoluty)
 - ✓ NESPECIFIKOVANO

Jak s polem 993 pracovat?

- záznam bez pole **993** je standard
- pokud bude nalezena shoda s SK, dodáme pole s vyplněným podpolem **\$a**, knihovny si ho naimportují
- v případě multiplicit dodáme podpole **\$m**
- doplňování podpolí **\$x** a **\$y** je na zodpovědnosti knihovny - většinou je to na základě druhu dokumentu, lze doplnit do:
 - ✓ šablon
 - ✓ importních skriptů
- počítá se s počátečním hromadným doplněním pole 993 (s podpoli **\$x** a **\$y**) na základě analýzy katalogu
- ruční vyplňování pole by mělo být spíše výjimkou (u nových záznamů, po individuálních opravách apod.)

Platné varianty pole, podpolí a hodnot

- pole **993** neexistuje
- pole obsahuje v podpolí **\$a** systémové číslo záznamu v SK (ideální stav)
- pole obsahuje několik podpolí **\$m** se systémovými čísly záznamů v SK
- pole obsahuje podpole **\$x NEPOSILAT** a podpole **\$y** s jednou z vyjmenovaných hodnot

Nástroje na podporu čištění dat

Martin Kybal, NTK Praha

Kategorizace dat knihovny

Data knihovny lze rozdělit na 3 skupiny:

1. záznamy odeslané do SK:
 - ✓ nutná pravidelná synchronizace dat pomocí skriptů
 - ✓ průběžné opravy nalezených chyb
2. záznamy neodeslané do SK:
 - ✓ zpracování na úrovni minimálního záznamu a odeslání do SK
 - ✓ po dohodě s SK lze zkusit import do SK s nízkou váhou záznamu
3. záznamy, u nichž je odeslání nežádoucí:
 - ✓ označení záznamů pro vyloučení z deduplikace
 - ✓ netřeba odesílat do SK
 - ✓ opravy pouze pro zlepšení kvality záznamu

Typy kontrol

- kontroly na pravidla SK ČR:
 - ✓ samostatně pro SK a ANL
- kontroly struktury záznamů:
 - ✓ nepovolené znaky (definované SK)
 - ✓ chyby ve struktuře záznamu (chybná pozice nebo označení pole, podpole) - vzniká při importu chybným naformátováním obsahu pole
- synchronizace dat s SK ČR:
 - ✓ kontrola SK pole 910 proti lokálnímu katalogu
 - ✓ kontrola vybraných polí proti SK (např. LDR(6), rok)
 - ✓ kontrola vybraných polí proti ČNB
 - ✓ automatická aktualizace pole 993 \$a, \$m

Nástroje na čištění dat

Požadavky:

- přehledné výpisy chyb s pravidelnou aktualizací
- možnost řazení výpisů podle různých kritérií (systémové číslo, pole v záznamu apod.)
- statistiky chyb (mj. pro vytipování hromadných oprav dat)
- možnost stažení výpisu:
 - ✓ v textové podobě
 - ✓ v podobě vhodné pro další zpracování v Excelu (tsv formát)
- optimálně ke každé chybě nápověda s popisem a návodem k opravě
- <http://aleph.techlib.cz/kontroly2/ntk/>

Výhody navrženého řešení

- vše na jednom místě (server NTK)
- snadné rozšiřování nástroje o další funkce
- možnost operativního řešení problémů
- kontrola nad celým procesem, která umožní průběžné vyhodnocování zapojení knihoven do čištění dat
- pomoc knihovnám, které by si nebyly schopné zajistit tyto činnosti vlastními silami
- knihovny budou postupovat jednotně (výhoda např. při opravách ve vztahu k SK)
- snadnější podpora knihoven ze strany Týmu pro data, resp. Servisního centra PNG

Postup zpracování dat

- pravidelný (denní) export dat knihovny
- nahrání exportu dat knihovny na ftp server
- zkopírování exportu dat knihovny z ftp serveru na server NTK
- automatické rozdělení exportního souboru dle údajů v poli 993 na menší soubory
- pravidelné (týdenní) stažení exportu dat knihovny z SK
- spuštění skriptů, které vygenerují požadované výstupy:
 - ✓ jednou denně - kontroly na pravidla SK a na strukturu záznamů
 - ✓ jednou týdně - synchronizace dat s SK
- vše bude probíhat zcela automaticky

Postup zapojení knihoven

- nastavení definic pro pole 993
- hromadné doplnění pole 993 \$x \$y do záznamů knihovny
- nastavení automatického exportu dat knihovny
- nastavení automatického kopírování exportu dat knihovny na ftp server
- otestování celého procesu
- spuštění pravidelného generování výstupů
- konzultace k jednotlivým výstupům
- zprovoznění exportu dat do SK ČR (pokud knihovna dosud nepřispívá)

Plán zapojení knihoven do čištění dat

- knihovny budou zapojovány postupně v závislosti na:
 - ✓ migrační vlně
 - ✓ knihovním systémem
 - ✓ stavu přispívání do SK
 - ✓ ...
- 2023/12 - dokončení první verze nástrojů na čištění dat
- 2024/01 - příprava prostředí pro čištění dat (server, ftp server)
- 2024/01-06 - postupné zapojování knihoven
- 2024 - čištění dat v knihovnách, podpora ze strany Týmu pro data
- 2024 - průběžný rozvoj navrženého řešení (mj. na základě připomínek z knihoven)
- 2024/09 - vyhodnocení navrženého řešení, analýzy dalších (nebibliografických) typů dat

Co závěrem?

- 16.11. je mezinárodní den **tolerance**



Tolerance je nutným předpokladem úspěšné spolupráce.
Budme tolerantní a spolupracujme, ať naplníme společný cíl...
...kvalitní společný centrální katalog v PNG.



Nulová tolerance nekvalitním datům:
500 \$\$a Nekvalitní ▼ data ^◀ plná 🤪 chyb
044 \$\$a Belgie

**Děkujeme za pozornost
Tým pro data**