

Katalogizace a propojená data

Barbora Drobíková

18. 5. 2023

Úvod

- Představení
- Téma
- Katalogizace
 - Lístkové katalogy
 - Automatizace
 - MARC(y)
 - Formáty propojených dat; 2018 LC oznámila přechod z MARC 21 na BIBFRAME; stejně tak Národní knihovna ve Švédsku

Proč propojená data

- knihovní data - často uložena v izolovaných databázích ve formátech, jako je MARC
 - -> mimo knihovnickou komunitu nejsou snadno srozumitelné
 - -> nedostupné pro webové vyhledávače a další systémy
- lepší zviditelnění knihoven na webu
- lepší vyhledávání - předkládána bohatší a s oblastí zájmu uživatelů související data (obohacená o další zdroje)
- lepší interoperabilita dat mezi různými komunitami (nakladatelské prostředí, služby zpřístupňující různé typy zdrojů, LAM sektor ...)

- Propojená data – základem pro tzv. sémantický web
- Technologie sémantického webu (RDF, OWL, SKOS, SPARQL atd.)- prostředí, kde se aplikace mohou dotazovat na data, vyvozovat závěry pomocí slovníků atd.
- Slovníky - pojmy a vztahy (označované také jako "termíny") používané k popisu a reprezentaci oblasti zájmu.
- Slovníky slouží ke klasifikaci pojmů, které lze použít v určité aplikaci, charakterizují možné vztahy a definují možná omezení při používání těchto pojmů.
- V praxi mohou být slovníky velmi složité (s několika tisíci pojmů) nebo velmi jednoduché (popisující pouze jeden nebo dva pojmy)
- *Čerpáno z webu konsorcia W3C:*
- <https://www.w3.org/standards/semanticweb/data>

Pokr. – viz W3C

- "slovníky" a "ontologie,, - není jasné rozdělení
- "ontologie" pro složitější a případně zcela formální soubory pojmů
- "slovník" se používá v případech, kdy se takový přísný formalismus nemusí nutně používat nebo jen ve velmi volném smyslu; slovníky jsou základními stavebními kameny pro techniky odvozování na sémantickém webu

Cituji z webu W3C:

- <https://www.w3.org/standards/semanticweb/data>:
- příklad je použití slovníků k uspořádání znalostí - knihovny, muzea, noviny, vládní portály, podniky, aplikace sociálních sítí a další komunity, které spravují rozsáhlé sbírky knih, historických artefaktů, zpravodajských zpráv, obchodních glosářů, příspěvků na blogu a dalších položek
- pomoci může obecný příklad - knihkupec může chtít integrovat data pocházející od různých vydavatelů; jedna databáze však může používat termín "autor", zatímco druhá může používat termín "tvůrce". Aby byla integrace úplná, měla by být do dat RDF přidána další definice popisující skutečnost, že vztah popsáný jako "autor" je stejný jako "tvůrce". Tato dodatečná informace je ve skutečnosti slovníkem (nebo ontologií), i když velmi jednoduchým.

Inference

- objevování nových vztahů
- data modelována jako soubor (pojmenovaných) vztahů mezi zdroji
- automatické postupy mohou generovat nové vztahy na základě dat a na základě propojování slovníků
- nové vztahy explicitně přidány do množiny dat, nebo jsou vráceny v době dotazu - otázka implementace

Znalostní grafy, sémantický web, RDF a knihovní data

- <https://data.gov.cz/%C4%8DI%C3%A1nky/znalostn%C3%AD-grafy-01-%C3%BAvod>
- 2012 – výsledek propojování dat - znalostní graf Google
- Reálné věci, osoby – na webu -propojení
- Původně jen propojování dokumentů na webu
- Znalostní grafy lze různě propojovat (DBPedia, Wikidata)

Základy sémantického webu

- Podle T. Berners Lee
- <https://www.w3.org/DesignIssues/LinkedData.html>
- 1) identifikovat objekty na webu pomocí URI
- 2) používat HTTP URI za účelem možného vyhledání objektů
- 3) při přístupu k HTTP URI jsou uživateli poskytnuta data o objektu reprezentovaná ve strojově čitelném formátu
- 4) data o objektu obsahují odkazy na HTTP URI jiných objektů, a tak lze dohledávat související propojená data. (TDKIV)

RDF, klíčové pojmy

Model pro výměnu dat na webu, strojová reprezentace znalostních grafů

Vztahy mezi zdroji, lidmi a věcmi na webu

Trojice – subjekt – predikát - objekt

URI, IRI

Reprezentace entit/reálných věcí

RDF XML – jen jedna z možných serializací

Slovníky

<https://data.gov.cz/%C4%8DI%C3%A1nky/znalostn%C3%AD-grafy-02-rdf>

William Shakespeare

je autorem

Hamlet

<https://id.loc.gov/authorities/names/n78095332.html>

<http://purl.org/dc/terms/creator>

<https://id.loc.gov/resources/works/7848605.html>

Příklady knihovních dat

- Knihovnické slovníky vyjádřené formou propojených dat:
- MESH (<https://www.nlm.nih.gov/mesh/meshhome.html>)
- GETTY Research Institute (<http://vocab.getty.edu/>)

Pokr.

- University of Wisconsin – Madison
- <https://search.library.wisc.edu/search/system>
- Marie Curie Sklodowska (Information from the web)

Pokr.

- Hong Kong University of Science and Technology
- Information Science – the basics
- <https://lbdiscover.hkust.edu.hk/bib/991013146942103412>
- <https://catalog.hkust.edu.hk/bf/991013146942103412>
- Knowledge card

- <https://catalog.hkust.edu.hk/bf/>
 - Discover with knowledge cards
 - Bibframe – marc conversion tool

Nový výměnný? formát

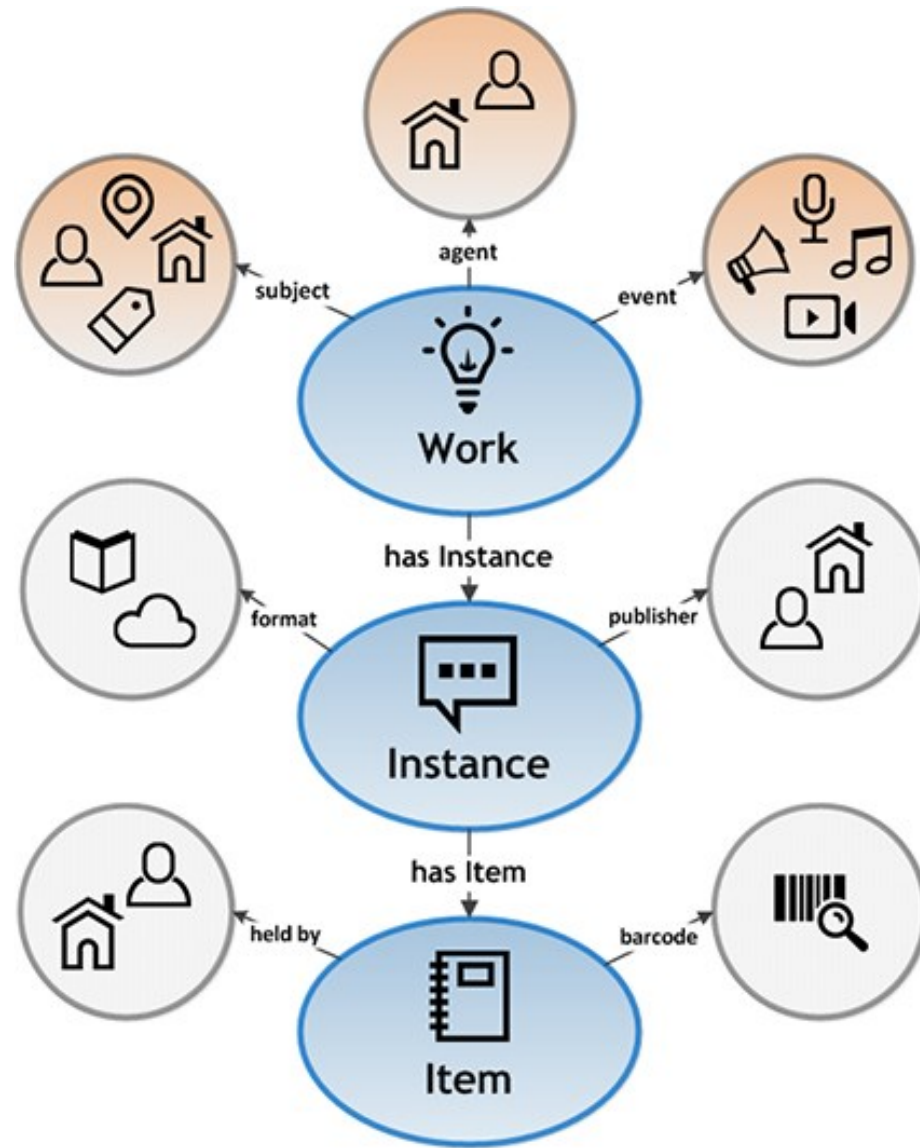
- založen na principu propojených dat
- není interním formátem, **formát pro publikování a sdílení dat na webu**

- **BIBFRAME – Library of Congress**
- OCLC – Schema.org
- stále se vyvíjí

BIBFRAME

- Vyvíjen v rámci LC
- Dnes široce přijímán jako vhodný formát pro knihovní data
- Share-VDE, LD4PC, PCC – spolupráce
- <https://www.svde.org/>

- hlavní informace o Bibframe
- <https://www.loc.gov/catworkshop/bibframe/index.html>
- <http://www.loc.gov/bibframe/faqs/>



<https://www.loc.gov/bibframe/docs/bibframe2-model.html>

Entity BIBFRAME

- **Dílo** – zdroj odrážející konceptuální základ katalogizovaného zdroje. Z hlediska modelu IFLA-LRM má nejbližší k entitě **vyjádření**.
- **Instance** – individuální, hmotné provedení díla. Z hlediska modelu IFLA-LRM má nejbližší k entitě **provedení**.
- **Jednotka** - jednotka je skutečná kopie (fyzická nebo elektronická) instance. Odráží informace, jako je jeho umístění (fyzické nebo virtuální), značka police a čárový kód.

Pokr.

- **Další klíčové koncepty:**
- **Aktéři** - aktéři jsou lidé, organizace, jurisdikce atd., Spojené s dílem nebo instancí prostřednictvím rolí, jako je autor, editor, umělec, fotograf, skladatel, ilustrátor atd.
- **Předmět** - dílo může být „o“ jednomu nebo více pojmech/konceptech – tj. „předmětech“ díla. Koncepty zahrnují témata, místa, časové výrazy, události, díla, instance, položky, aktéry atd.
- **Událost** - události, jejichž záznam může být obsahem díla.
- **Hub** – nová třída (téměř) = **dílo** dle IFLA-LRM

Official RDA

- 2017 – 3R projekt
- Obsah zcela reorganizován, na různých místech
- Není to nový standard, není to RDA 2.0, ale nově nabízí vždy několik cest jak zaznamenat informace -> pro různé komunity využívající RDA

- I když různé komunity:
 - Stejná sada prvků
 - Stejně řízené slovníky
 - Založen na stejném konceptuálním modelu
 - ->fundamentální konzistence podporující interoperabilitu dat

Pokr.

- Odpovídají na otázky:
- Jaká data mám zaznamenat?
- Jak je mám zaznamenat?
- Definuje prvky nutné k popisu i vyhledávání

- Původní pravidla – vedla k tvoření řetězců znaků a slov
- Nyní: data zaznamenána jako prvky, které jsou interpretovatelné na prvním místě stroji

Pokr.

- Strojové zpracování vyžaduje bibliografické zpracování ve formě prvků
- Každý prvek jednoznačně definován, obsahuje konkrétní data
- RDA nezávislé na formátech, v M21 stejně jako v RDF, BIBFRAME
- RDA – otevřený rámec – nové typy zdrojů/typy obsahů

Pokr.

- Musíme nově přemýšlet o jednotlivých prvcích a jejich vztazích
- Nicm. zápis mnoha prvků bude stejný, jako jsme zvyklí
 - Např. rok vydání: 2023
- *"RDA, Resource Description and Access, is an international metadata standard designed to enable the discovery of library and cultural heritage resources in both traditional and linked data environments".*
- Oliver, Chris. *Introducing RDA : A Guide to the Basics After 3R*, American Library Association, 2021. *ProQuest Ebook Central*, <http://ebookcentral.proquest.com/lib/cuni/detail.action?docID=6624803>.
Created from cuni on 2023-05-02 12:16:45

Ukázky

- Marva - <https://bibframe.org/marva/editor>
- Sinopia - <https://stage.sinopia.io/>
- knihovny pracují v editorech třetích stran a importují do systémů

Limity, výzvy

- Podobné jako u přechodu na MARC
- Kompatibilita – MARC->BIBFRAME, nutné testovat
- Trénink katalogizátorů
- Knihovní systémy a podpora BIBFRAME
- BIBFRAME se vyvíjí (v reálném provozu již dlouho)
- Náklady
- Propojená data – miliony záznamů = miliardy propojení - úložiště, nástroje
- Efektivní, pokud jej adoptuje širší komunita

Konverze do BIBFRAME

- Analýza dat
- Příprava slovníků, jejich určení (viz např. Library of Congress Linked Data Service)
- URI
- Entifikace při konverzi nebo před konverzí; \$\$1

Zdroje

- 23 Linked Data Things, MiniTex, University of Minnesota <https://minitex.umn.edu/services/digital-initiatives-metadata/23-linked-data-things>
- BERNERS-LEE, Tim. Linked Data, Up To Design Issues <https://www.w3.org/DesignIssues/LinkedData.html>
- BIBFRAME <https://www.loc.gov/catworkshop/bibframe/index.html>
- “Final Report to the PCC Oversight Group from the BIBCO Mapping BSR to BIBFRAME 2.0 Group,” July 2017. <https://www.loc.gov/aba/pcc/bibframe/TaskGroups/BSR-PDF/FinalReportBIBCO-BIBFRAME-TG.pdf>.
- FRANK, Paul. Official RDA and BIBFRAME. European Bibframe workshop 2022, <https://www.casalini.it/bfwe2022/>
- HAVLOVÁ, Jaroslava. propojená data. In: *KTD: Česká terminologická databáze knihovnictví a informační vědy* (TDKIV) [online]. Praha : Národní knihovna ČR, 2003- [cit. 2023-05-16]. Dostupné z: https://aleph.nkp.cz/F/?func=direct&doc_number=000016064&local_base=KTD
- Library of Congress. Linked Data Service <https://id.loc.gov/>.
- McCallum, Sally. BIBFRAME implementation journey, European Bibframe workshop 2022, <https://www.casalini.it/bfwe2022/>
- NEČASKÝ, Martin. Série Znalostní grafy, Otevřená data, <https://data.gov.cz/%C4%8D%C3%A1nky/znalostn%C3%AD-grafy-01-%C3%BAvod>; <https://data.gov.cz/%C4%8D%C3%A1nky/znalostn%C3%AD-grafy-02-rdf>
- OLIVER, Chris. *Introducing RDA : A Guide to the Basics After 3R*, American Library Association, 2021. *ProQuest Ebook Central*, <http://ebookcentral.proquest.com/lib/cuni/detail.action?docID=6624803>. Created from cuni on 2023-05-02 12:16:45
- Program for Cooperative Cataloging Task Group on URIs in MARC, “Formulating and Obtaining URIs: A Guide to Commonly Used Vocabularies and Reference Sources,” version date 2020-01-15, https://www.loc.gov/aba/pcc/bibframe/TaskGroups/formulate_obtain_URI_guide.pdf.
- RDAToolkit <https://access.rdatoolkit.org/>
- SAMPLES, Jacquie, BIGELOW, Jan (2020). MARC to BIBFRAME: Converting the PCC to Linked Data, *Cataloging & Classification Quarterly*, 58:3-4, 403-417, DOI: [10.1080/01639374.2020.1751764](https://doi.org/10.1080/01639374.2020.1751764)
- SHARE-VDE <https://www.svde.org/>
- W3C – Semantic Data - <https://www.w3.org/standards/semanticweb/data>
- Zapounidou, Sofia. (2020). Study of library data models in the Semantic Web environment [Zenodo]. <https://doi.org/10.5281/zenodo.4018523>