

Trendy v oblasti bibliografických formátů aneb Má MARC budoucnost? [1]

Autor:

- [RÖSSLEROVÁ, Klára](#) [2]

Číslo:

- [2017, ročník 26, číslo 2](#) [3]

Rubrika:

- [Trendy v knihovnách](#) [4]

Klíčová slova:

- [bibliografické formáty](#) [5], [propojená data](#) [6]

Region:

- [01 Praha](#) [7]

Redakční poznámka: Jedná se o textovou podobu přednášky přednesené na valné hromadě [pražské organizace SKIP](#) [8], která se uskutečnila 18. dubna 2017 a o které Vás informujeme v [samostatném článku](#) [9].

Jak to začalo

[Kongresová knihovna](#) [10] (Library of Congress) vznikla roku 1800. Zákonem z roku 1902 pak byla knihovně dána povinnost distribuovat všechny vytvořené záznamy do amerických knihoven. V 60. letech 20. století bylo oddělení, které tuto agendu zajišťovalo, největším oddělením Kongresové knihovny – jeho prostory zabíraly plochu čtyřikrát větší než fotbalové hřiště. V oddělení pracovalo 600 lidí, jejichž rukama ročně prošlo 78 miliónů záznamů.¹

Již na konci 50. let 20. století se tak Kongresová knihovna začala zajímat o možnost automatizace. Nejprve knihovna získala grant na výzkum možnosti automatizace, postupně byla vytvořena studie možných metod. Roku 1965 knihovna začala hledat osobu se znalostí systémové analýzy a programování, která by projekt automatizace mohla vést. Tak do knihovny nastoupila Henriette D. Avramová, tvůrkyně [formátu MARC](#) [11]. Nejprve provedla analýzu katalogizačních záznamů a pokračovala návrhem možnosti jejich distribuce v strojem čitelné podobě. Knihovna si od formátu slibovala možnost vytvářet bibliografická data ve strojem čitelné podobě a na magnetických páskách je distribuovat do amerických knihoven, které si je měly samy tisknout. Díky tomu se pro Kongresovou knihovnu mělo zjednodušit plnění výše zmíněné zákonné povinnosti.

Tak roku 1968 vznikl formát MARC, standardizovaný formát sloužící pro výměnu dat mezi systémy. Spoluprací s Britskou národní bibliografií následně vznikl formát MARC II (původní formát MARC byl přejmenován na MARC I), resp. UK MARC. Do roku 1995 vzniklo 24 různých podob původního formátu, jehož struktura se roku 1973 stala [mezinárodním standardem ISO 2709 Dokumentace – Formát pro výměnu bibliografických informací na magnetické pásce](#) [12]. V souladu se strukturou vznikl i univerzální MARC – UNIMARC, a to díky spolupráci [Mezinárodní federace knihovnických](#)

[asociací a institucí](#) [13] (International Federation of Library Associations and Institutions, IFLA) s Henriette D. Avramovou.²

Od roku 1973 byl ke stejnému účelu vytvořen i německý formát MAB (později ještě MAB2). Od roku 1997 však začaly knihovny po celém světě přecházet na formát MARC 21, který vznikl a je udržován díky spolupráci Kongresové knihovny, Kanadské národní knihovny, Britské knihovny a později i Německé národní knihovny. Knihovny tak začaly distribuovat své záznamy v podobě sice obměněné, ale stále vycházející ze struktury formátu, který vznikl v 60. letech 20. století v Kongresové knihovně z důvodu zjednodušení distribuce záznamů na tehdy používaném médiu – magnetické pásce. Začala se tak nabízet otázka, jestli není čas formát dat přizpůsobit dnešnímu světu, kdy magnetické pásky většina současných knihovníků nikdy neviděla, diskety si někteří matně vybavují a na CD a DVD nám již v noteboocích chybí mechaniky. Je vůbec struktura MARC nutná?

A vadí MARC někomu?

Otázku, zda je struktura MARC nutná, lze položit i obráceně: a komu vadí? Formát byl vytvořen tak, aby záznamy zabraly co nejméně místa na nosiči, který byl v době vzniku formátu používán, tedy na magnetické pásce. Je to formát výměnný, který byl vytvořen pro výstup ze systému, resp. vstup do systému. Nebyl vytvořen jako formát pro uložení záznamů v systému (interní formát) ani jako formát určený pro katalogizaci. Pokud by byl používán výhradně jako vstupní nebo výstupní (tedy výměnný), knihovníci katalogizátoři by se s ním ani nesetkali a celý tento článek by byl určen spíše pro systémové knihovníky. Mimochodem díky striktnímu rozlišování funkcí formátu byl přechod z formátu MAB na formát MARC 21 v německých knihovnách velmi jednoduchý. Systémoví knihovníci v Německé národní knihovně pouze namapovali interní formát PICA+ na nový formát. Tamější knihovníci katalogizují ve formátu PICA 3 a to, že se skupina výstupních formátů rozšířila nebo změnila, pro ně ani pro jejich práci nemuselo být důležité.

Je tedy nutné, aby formát MARC 21 byl nahrazen? Bude nahrazen jedním formátem? Nebo bude nahrazen podle účelu několika formáty: formátem pro výměnu a prezentaci dat, formátem pro uložení dat v systémech nebo také formátem výměny mezi knihovnami?

Zveřejňování dat

Již od roku 2001 se díky [Timu Bernersovi-Lee](#) [14] setkáváme s výrazem [sémantický web](#) [15]. Takový web je srozumitelný strojům, a to díky vhodné struktuře dat. Na jejím základě stroje dokáží rozpoznat jednotlivé informace. Publikační model pro zveřejňování strukturovaných dat na webu, který je založen na [protokolu HTTP](#) [16], [identifikátorech URI](#) [17] (Uniform Resource Identifiers, jednotné identifikátory zdrojů) a [datovém modelu RDF](#) [18] (Resource Description Framework, systém popisu zdrojů), se nazývá model [propojených dat](#) [19] (Linked Data).³ Zveřejňování bibliografických záznamů v této podobě přináší řadu výhod – čtenář například záznam dokumentu najde přímo z webu. Pokud vše funguje ideálně, najde i řadu relevantních informací navíc.⁴

V praxi to znamená poskládat jednotlivé údaje ze záznamu do podoby trojic (podmět – predikát – předmět, přičemž podmět a předmět jsou vždy nějakou „věcí“ a predikát mezi nimi určuje vztah). Dále je potřeba pro všechny podmínky a pro co nejvíce předmětů najít nebo vytvořit URI, aby byla zajištěna jejich vzájemná propojitelnost. Pro predikáty je třeba vyhledat vhodný slovník tvořený jednotlivými výrazy s vlastními URI. Podle průzkumu OCLC z roku 2015⁵ mezi nejpoužívanější slovníky patří:

- [Simple Knowledge Organization System](#) [20] (skos),
- [Friend of a Friend](#) [21] (foaf),
-

[DCMI Metadata Terms](#) [22] (dcterms),

- [Dublin Core Metadata Element Set](#) [23] (dce),
- [Schema.org vocabulary](#) [24] (schema),
- [The Bibliographic Ontology](#) [25] (bibo),
- [Local Vocabulary VOCABS](#) [26] (rda),
- [Europeana Data Model Vocabulary](#) [27] (edm),
- [ISBD Elements](#) [28] (isbd),
- [WGS84 Geo Positioning](#) [29] (geo),
- [BIBFRAME Vocabulary](#) [30] (bf).

Ideálním výstupem pak je trojice URI – URI – URI, přípustný je i výstup v podobě trojice URI – URI – slovo (literál).

Aby využití propojených dat mělo smysl, je vhodné najít pokud možno ke všem „věcem“ již existující URI a nově vytvořená URI s nimi propojit. Jenom tak lze následně těžit z propojitelnosti, protože pokud ostatní použijí pro stejnou „věc“ stejné URI, trojice se propojí (např. dojde k propojení autora v katalogu s informacemi o autorovi ve Wikipedii nebo dojde k propojení názvu knihy s weby věnujícími se recenzím knih). Vyhledání existujících URI může být časově náročné. Dalším krokem je tyto trojice zpřístupnit tak, aby jim dokázaly porozumět stroje, tedy aby byly vůbec využitelné. Právě pro stroje (či lépe aplikace) je určen rámec RDF.

Vytvořená data ve struktuře propojených dat je nutné uložit na bezpečné místo serveru a nakonec je zpřístupnit uživatelům nebo třetím stranám.

Propojená data místo formátu MARC 21?

Zveřejňováním dat ve formě propojených dat se knihovny otevírají světu. Informace uložené v hlubinách knihovních katalogů vyplouvají na volný web. Můžeme si tedy položit otázku, zda tedy není čas formát MARC nahradit. Odpověď je jednoduchá – není. Formát MARC, je-li používán tak, jak byl navržen, je pouze jedním z možných výstupů. Propojená data mohou (a mnohde již jsou) jen dalším možným výstupem. Jsou to jen jinak poskládaná stávající data obohacená o URI. Práce katalogizátorů může být pořád stejná, jen s hledáním URI navíc (tato činnost však může vycházet z postupů při práci s autoritními záznamy a být alespoň zčásti automatizována) a s daleko bohatším výsledkem.

Teorie nebo už praxe?

Průkopníky v oblasti propojených dat v knihovnách jsou [Kongresová knihovna](#) [31] a [OCLC](#) [32]. Obě instituce se propojenými daty zabývají od roku 2011. Kongresová knihovna vytvořila vlastní publikační model BIBFRAME, OCLC použila jako základ prvky ze slovníku Schema.org, které rozšířila

o další prvky z vlastního jmenného prostoru. Své záznamy ve formě otevřených dat [zpřístupňuje OCLC](#) [33] v [katalogu WorldCat](#) [34].

Problematice propojených dat se věnuje i [Švédská národní knihovna](#) [35], která spravuje švédský [souborný katalog LIBRIS](#) [36]. V tomto katalogu jsou dostupné záznamy 175 švédských knihoven. Celkem se jedná o více než šest milionů bibliografických záznamů. Knihovníci Švédské národní knihovny se při převádění dat z formátu MARC 21 do podoby propojených dat řídili velmi jednoduchou myšlenkou – je lepší přijít s něčím teď než lpět na detailech a čekat na dokonalost.

Další skandinávskou zemí, která chce své katalogy otevřít světu, je Finsko. [Finská národní knihovna](#) [37] s převodem svých bibliografických záznamů do podoby propojených dat začala v roce 2015. Převod dat je finančně podpořen [Programem pro otevřená data](#) [38] (Open Data Programme). Ten byl vyhlášen v návaznosti na rozhodnutí finské vlády z roku 2011, které nařizuje institucím zpřístupnit své veřejné informační zdroje v podobě propojených dat. V rámci projektu vzniká také úplná dokumentace použitých postupů, kterou budou moci využít i další finské knihovny.

Díky tlaku vlády na zpřístupnění dat začali vyvíjet aktivitu i britští knihovníci, kteří se propojenými daty zabývali již od roku 2009. V roce 2011 v podobě propojených dat [zpřístupnili část Britské národní bibliografie](#) [39], konkrétně přes tři miliony záznamů. Od roku 2010 jsou ve formě propojených dat zpřístupněny [všechny bibliografické záznamy Maďarské národní knihovny](#) [40]. Německá národní knihovna začala ve formě propojených dat [poskytovat národní bibliografii](#) [41]. Na záznamy však platí časové embargo, současné záznamy si je možné stáhnout za poplatek. Přesto bylo zpřístupněno již 11,5 milionů bibliografických záznamů. Francouzská národní knihovna zpřístupňuje portál [data.bnf.fr](#) [42] s více než dvěma miliony stránek o osobách a institucích a taktéž pracuje s propojenými daty. Tato služba eviduje přes 100 tisíc unikátních přístupů měsíčně, přičemž v 80 % se jedná o přístupy prostřednictvím běžného vyhledávače, nikoliv přímým přístupem na webovou stránku služby. [Digitální knihovna Europeana](#) [43], která vznikla v roce 2008, zpřístupňuje přes 53 milionů digitálních děl (obrazů, hudby, knih apod.), veškerá data jsou od roku 2012 k dispozici ve formě otevřených propojených dat.⁶

Jako užitečná inspirace pro ostatní knihovny může posloužit např. článek *A Beginner's Guide to Creating Library Linked Data: Lessons from NCSU's Organization Name Linked Data Project* [7](#).

Závěr

Díky struktuře propojených dat se cenné informace vytvářené knihovníky konečně dostávají z uzavřených knihovních katalogů a databází (v angličtině je často používán výraz *silos*) i na volný web, kde je naleznou koncoví uživatelé – stroje, aby je zprostředkovali dalším koncovým uživatelům – čtenářům. Je nutné si uvědomit, pro jaké účely jsou výměnné formáty bibliografických dat v knihovnictví používány: měly by být používány pro zápis a přenos (výměnu) mezi bibliografickými agenturami a jinými institucemi.

V tuto chvíli je však zřejmé, že v budoucnu se odštěpí další funkce pro publikování bibliografických dat na webu. Formát MARC 21 tedy může existovat dál.

- [1](#). LIBRARY OF CONGRESS. *Cataloging Distribution Service* [online]. Washington: Library of Congress, 2002 [cit. 2010-06-01]. Videozáznam o historii distribuce záznamů. Dostupné z: <http://www.loc.gov/cds/who.html> [44]
- [2](#). EHRlichová, Klára. *Osobnost a dílo Henrietty D. Avramové (1919–2006) [Henriette D. Avram (1919–2006: her Personality and Work)]*. Praha, 2009-03-14. 116 s., 22 s. příl. Rigorózní práce. Univerzita Karlova v Praze, Filozofická fakulta, Ústav informačních studií a knihovnictví. Konzultant rigorózní práce PhDr. Eva Bratková.
- [3](#). BIZER, Christian, HEALTH, Tom, BERNERS-LEE, Tim. *Linked Data: the story so far* [online]. 2009 [cit. 2014-03-18]. This is a preprint of a paper to appear in: Heath, T., Hepp, M., and Bizer, C. (eds.). Special Issue on Linked Data, International Journal on Semantic Web and Information Systems (IJSWIS). Dostupné z: <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf> [45]

- 4. PAPADAKIS, Ioannis, KYPRIANOS, Konstantinos, STEFANIDAKIS, Michalis. Linked Data URIs and Libraries. *D-Lib Magazine* [online]. 2015, May/June, vol. 21, no. 5/6 [cit. 2017-07-24]. Dostupné z: <https://dx.doi.org/10.1045/may2015-papadakis> [46]
- 5. SMITH-YOSHIMURA, Karen. Analysis of International Linked Data Survey for Implementers. *D-Lib Magazine* [online]. 2016, July/August, vol. 22, no. 7/8 [cit. 2017-07-24]. Dostupné z: <https://dx.doi.org/10.1045/july2016-smith-yoshimura> [47]
- 6. RÖSSLEROVÁ, Klára. *Historie, současnost a budoucnost výměnných formátů bibliografických dat [History, Presence and Future of Exchangeable Formats of Bibliographic Data]*. Praha, 2016. 188 s., xlvii s. příl. Disertační práce. Univerzita Karlova, Filozofická fakulta, Ústav informačních studií a knihovnictví. Školitel doc. PhDr. Rudolf Vlasák. Dostupné také z: https://is.cuni.cz/webapps/zzp/detail/105313/24537045/?q=%7B%22____searchform____search%22%3A%22historie+po%5Cu010d%5Cu00edta%5Cu010dov%5Cu00fdch+her%22%2C%22____searchform____butsearch%22%3A%22Vyhledat%22%2C%22PNzzpSearchListbasic%22%3A%2250%22%7D&lang=cs [48]
- 7. HANSON, Eric M. A Beginner's Guide to Creating Library Linked Data: Lessons from NCSU's Organization Name Linked Data Project. *Serials Review* [online]. 2014, vol. 40, issue 4, s. 251–258 [cit. 2017-07-24]. Dostupné z: <https://dx.doi.org/10.1080/00987913.2014.975887> [49]

URL zdroje: <https://bulletinskip.skipcr.cz/vsechna-cisla/prohlizet-cisla/2017-rocnik-26-cislo-2/trendy-v-oblasti-bibliografickych-formatu-aneb>

Odkazy

- [1] <https://bulletinskip.skipcr.cz/vsechna-cisla/prohlizet-cisla/2017-rocnik-26-cislo-2/trendy-v-oblasti-bibliografickych-formatu-aneb>
- [2] <https://bulletinskip.skipcr.cz/vsechna-cisla/autori/rosslerova-klara>
- [3] <https://bulletinskip.skipcr.cz/vsechna-cisla/prohlizet-cisla/2017-rocnik-26-cislo-2>
- [4] <https://bulletinskip.skipcr.cz/vsechna-cisla/rubriky/trendy-v-knihovnach>
- [5] <https://bulletinskip.skipcr.cz/vsechna-cisla/klicova-slova/bibliograficke-formaty>
- [6] <https://bulletinskip.skipcr.cz/vsechna-cisla/klicova-slova/propojena-data>
- [7] <https://bulletinskip.skipcr.cz/vsechna-cisla/regiony/01-praha>
- [8] <http://www.skipcr.cz/regiony/01-praha>
- [9] <https://bulletinskip.skipcr.cz/vsechna-cisla/prohlizet-cisla/2017-rocnik-26-cislo-2/regionalni-valna-hromada-prazske-organizace>
- [10] <https://www.loc.gov/>
- [11] <http://www.loc.gov/marc/>
- [12] <https://www.iso.org/standard/41319.html>
- [13] <https://ifla.org/>
- [14] https://en.wikipedia.org/wiki/Tim_Berners-Lee
- [15] https://en.wikipedia.org/wiki/Semantic_Web
- [16] https://en.wikipedia.org/wiki/Hypertext_Transfer_Protocol
- [17] https://en.wikipedia.org/wiki/Uniform_Resource_Identifier
- [18] https://en.wikipedia.org/wiki/Resource_Description_Framework
- [19] https://en.wikipedia.org/wiki/Linked_data
- [20] <https://www.w3.org/TR/skos-reference/>
- [21] <http://xmlns.com/foaf/spec/>
- [22] <http://dublincore.org/documents/dcmi-terms/>
- [23] <http://dublincore.org/documents/dces/>
- [24] <http://schema.org/docs/schemas.html>
- [25] <http://bibliontology.com/>
- [26] <http://www.rdaregistry.info/>
- [27] <http://pro.europeana.eu/share-your-data/data-guidelines/edm-documentation>
- [28] <http://iflastandards.info/ns/isbd/elements/>
- [29] https://www.w3.org/2003/01/geo/wgs84_pos
- [30] <http://bibframe.org/vocab/>

- [31] <http://loc.gov/>
- [32] <http://oclc.org/>
- [33] <http://www.oclc.org/en/worldcat/data-strategy.html>
- [34] <http://www.worldcat.org/>
- [35] <http://www.kb.se/>
- [36] <http://libris.kb.se/?language=en>
- [37] <https://www.kansalliskirjasto.fi/en>
- [38] <http://vm.fi/en/open-data-programme>
- [39] <http://bnb.data.bl.uk/>
- [40] <https://datahub.io/dataset/hungarian-national-library-catalog>
- [41] http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkeddata_node.html
- [42] <http://data.bnf.fr/>
- [43] <http://www.europeana.eu/portal/en>
- [44] <http://www.loc.gov/cds/who.html>
- [45] <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>
- [46] <https://dx.doi.org/10.1045/may2015-papadakis>
- [47] <https://dx.doi.org/10.1045/july2016-smith-yoshimura>
- [48] https://is.cuni.cz/webapps/zzp/detail/105313/24537045/?q=%7B%22____searchform____search%22%3A%22historie+po%5Cu010d%5Cu00edta%5Cu010dov%5Cu00fdch+her%22%2C%22____searchform____butsearch%22%3A%22Vyhledat%22%2C%22PNzzpSearchListbasic%22%3A%2250%22%7D&lang=cs
- [49] <https://dx.doi.org/10.1080/00987913.2014.975887%5B/f>