

Univerzita Karlova

Filozofická fakulta

Ústav informačních studií a knihovnictví

Informační studia a knihovnictví – Informační věda

Teze disertační práce

Historie, současnost a budoucnost výměnných formátů bibliografických

dat

History, Presence and Future of Exchangeable Formats of Bibliographic Data

Školitel: Doc. PhDr. Rudolf Vlasák

2016

PhDr. Klára Rösslerová

Obsah

1. Předmluva.....	3
2. Úvod.....	4
3. Struktura práce	5
4. Teze disertační práce	6
4.1. Typologie formátů.....	6
4.2. Analýza formátů pro výměnu bibliografických dat.....	6
4.3. Nejpoužívanější formáty ve světovém měřítku	6
4.4. Uplatnění propojených dat v knihovnách.....	7
4.5. Doporučený postup převodu do struktury propojených dat v knihovnách.....	8
4.6. Doporučený postup převodu do struktury propojených dat v českých knihovnách	9
4.7. Experimentální převod a analýza existujících záznamů.....	9
4.8. Připravenost knihovních softwarů.....	10
4.9. Postoje klíčových hráčů při tvorbě standardů.....	11
5. Závěr.....	13
6. Seznam pramenů citovaných v tezích	17

1. Předmluva

Předkládaná disertace se zabývá historií, současností a možnou budoucností výměnných formátů bibliografických dat. Mou původní vizí při zadávání tématu bylo analyzovat používané výměnné formáty, najít na nich to dobré nebo špatné a navrhnout formát ideální. K tématu mě vedla častá kritika formátu MARC 21 a negativní hodnocení přechodu z formátu UNIMARC v České republice. V průběhu psaní práce však došlo k vývoji: v zahraničních knihovnách se tématem stala propojená data. Obsah disertace se tak aktuálnímu trendu zcela přirozeně přizpůsobil.

K tématu disertace mě přivedl dlouhodobý zájem o výměnné formáty z doby pregraduálního studia, ale i pozdější pracovní zkušenosti. Ve své diplomové a rigorózní práci jsem se zabývala historií vzniku výměnných formátů. Během svého zaměstnání ve firmě Comdat spol. s r.o. jsem jako projektová manažerka vedla projekty retrokonverze bibliografických záznamů českých knihoven. Zabývala jsem se retrokonverzí z lístkových katalogů i převody dat v nejrůznějších formátech do formátů typu MARC. V průběhu psaní této disertace jsem se navíc zúčastnila projektu mapování bibliografických údajů do struktury propojených dat ve Finské národní knihovně.

2. Úvod

Cílem předkládané disertace je nastínění budoucího možného vývoje na poli výměnných formátů bibliografických dat včetně doporučení pro knihovny, jakým způsobem se změně přizpůsobit. Prostředkem je nejprve vysvětlení historických souvislostí v oblasti zavádění informačních technologií do knihoven a podrobná analýza výměnných formátů bibliografických dat používaných ve světových knihovnách a zcela samostatně v českém (československém prostředí). Výměnné formáty byly nutnou podmínkou pro počátek sdílené katalogizace. Po etablování formátu MARC v angloamerickém prostředí došlo k takovému rozvoji, který vedl až k nutnosti vytvoření výměnného formátu mezi výměnnými formáty – formátu UNIMARC. V důsledku malé finanční podpory však došlo k postupnému útlumu tohoto formátu a naopak díky silnému finančnímu zázemí Kongresové knihovny se postupně stal vedoucím formátem v celosvětovém měřítku formát MARC 21. **Formáty typu MARC i přes svoje stáří stále plní funkci výměnného formátu, přestože jeho původní využití bylo pro výměnu bibliografických záznamů na magnetických páskách. Stále častěji tak knihovníci začali hovořit o nutnosti náhrady zastaralého výměnného formátu něčím modernějším.** Pro zjištění názoru odborné knihovnické obce jsem knihovníky oslovila: oslovila jsem několik odborných komunit, jejíž členové se prostřednictvím vytvořených dotazníků vyjadřovali k nutnosti náhrady formátů typu MARC jiným typem formátu, a zároveň jsem zjišťovala jejich názory na budoucí vývoj. Z dotazníků, ale hlavně z aktivit odborných uskupení nebo knihoven, je zřejmé, že **budoucím vývojem budou propojená data.** Zároveň jsem analyzovala knihovnické softwary, abych zjistila, zda jsou na případnou změnu připravené.

Schéma propojených dat pravděpodobně nahradí formáty typu MARC v oblasti publikování dat na webu, čímž se knihovníky vytvořené bibliografické záznamy jednoduše dostanou k širší uživatelské základně, a to včetně přidaných informací, neboť propojená data fungují na principu propojování (a tedy obohacování) informací z různých zdrojů. Zároveň je zřejmé, že pro samotnou výměnu bibliografických záznamů formáty typu MARC nevhodné nejsou – je-li tento formát používán skutečně pouze jako výměnný, a ne vstupní pro katalogizaci, jak se často stává.

3. Struktura práce

Text práce je rozdělen do deseti hlavních kapitol. Kapitoly 2 a 3 jsou věnovány terminologii a typologii výměnných formátů bibliografických dat. Ve čtvrté kapitole jsou analyzovány výměnné formáty používané v knihovnictví. Výměnné formáty jsou analyzovány chronologicky podle doby svého vzniku. Podkladem pro analýzy byly dobové výzkumné zprávy, články, vzpomínky i soukromé dokumenty. V páté kapitole jsou analyzovány výměnné formáty vyvíjené a používané v České republice. Pro zjištění současného stavu v oblasti používaných výměnných formátů bylo provedeno několik analýz, jejich výsledky jsou představeny v šesté kapitole. Postupně tak byly analyzovány formáty, ve kterých jsou poskytovány národní bibliografie nebo obecně bibliografické záznamy. Cílem těchto analýz bylo zjistit a doložit hypotézu, že nejpoužívanějším výměnným formátem v současnosti je formát MARC 21. Sedmá kapitola disertace je věnována teoretickému úvodu do oblasti propojených dat a vysvětlení výhod. Detailní popis převodu do této struktury (kapitola 8) včetně doporučení pro české prostředí a praktické ukázky (kapitola 9). Teoretický základ mi poskytly jak technické zprávy, tak i články popisující zkušenost převodu bibliografických záznamů do struktury propojených dat ve významných knihovnách. Analýzy dávají představu o připravenosti knihovních softwarů na publikování ve formě propojených dat (kapitola 10), ale i názorů odborné veřejnosti na jednotnost v oblasti propojených dat (kapitola 11). Názory tří rozdílných skupin odborné veřejnosti byly zjišťovány formou kvalitativního výzkumu (dotazníkového šetření). Pro první skupinu respondentů byla navíc použita tzv. delfská metoda, která prostřednictvím opakovaného dotazování a sdělení výsledku respondentům dává možnost predikce vývoje.

4. Teze disertační práce

4.1. Typologie formátů

Existují různé typologie formátů. Typologie formátů dle jejich struktury ani dle lokality není nijak problematická, je pouhým rozdělením existujícího. Přelévání kategorií může být problematické u rozdělení formátů dle životního cyklu. Striktní dodržování rozlišení formátů na vstupní, interní a výstupní zjednodušuje knihovnám přechod na jiný národní formát a zároveň jim výrazně ulehčuje život při vytváření nových výstupních (publikačních) formátů. Zcela se také mohou oprostít od otázky, zda je struktura formátu MARC (MARC 21) zastaralá, protože je při jejich způsobu práce irelevantní.

4.2. Analýza formátů pro výměnu bibliografických dat

V předkládané disertační práci byly analyzovány formáty MARC, MARC II, UNIMARC, Společný komunikativní formát, MAB, MAB2 a MARC21, samostatně v kapitole zaměřené na Českou republiku navíc Obecná datová struktura a Výměnný formát. Cílem bylo zjištění rozdílů mezi jednotlivými formáty, nalezení shod, důvody případného zániku a zjištění praxí ověřené ideální minimální struktury a obsahu nutné či vhodné pro využití bibliografických záznamů. Analýzou a následným srovnáním formátů byly zjištěny některé shody, s ohledem na cíl disertace je důležité zjištění existence adresářů, které byly přínosné pro ušetření místa při přenosu na magnetických páskách, z dnešního pohledu jsou však již zbytečné. Ve formátech je dále pracováno s kódovanými poli, které usnadňují a rozšiřují vyhledávání záznamů a práci s nimi (např. při vytváření citací). Kódovaná pole tak nesou velké množství informací a při jakékoliv konverzi nebo vytváření nové struktury dat by neměla být opomenuta.

Z vývoje je však hlavně zřejmé, že přes snahy o modernizaci formátů nemá klíčovou roli samotná kvalita formátu, ale síla instituce, která za jeho vývojem a udržováním stojí. Historie již ukázala, že světovým vůdcem v této oblasti je Kongresová knihovna a její aktivity je vhodné nejenom sledovat, ale případně je i následovat.

4.3. Nejpoužívanější formáty ve světovém měřítku

Pro zhodnocení současného stavu a zjištění nejčastěji používaných výměnných formátů bibliografických dat byly provedeny tři různé analýzy: zjištění formátu dat, ve kterém jsou poskytovány národní bibliografie, analýza formátu dat

poskytovaných prostřednictvím protokolu Z39.50 a analýza znakových dat dle seznamu datahub.io. **Z výsledků je zřejmé, že vedoucím formátem současného knihovnictví je formát MARC 21.** Je v něm poskytováno 81 % národních bibliografií. Naopak formát UNIMARC je na ústupu, je v něm poskytováno pouze 19 % národních bibliografií. Vedoucí úloha formátu MARC 21 je zřejmá i z formátů dat poskytovaných přes protokol Z39.50: data v něm poskytuje 75 % knihoven, ve formátu UNIMARC pouze 12 % knihoven. Jednotlivé národní varianty formátu MARC využívá od 2 do 11 % knihoven, které však své záznamy zároveň poskytují i ve formátu MARC 21 nebo UNIMARC. Z analýzy je opět zřejmé, že formát, podporovaný Kongresovou knihovnou, má vedoucí postavení.

Analýza datových sad dle datahub.io nevypovídá nic o zastoupení formátu MARC 21 nebo UNIMARC, je však zajímavá z hlediska propojených dat. Z analyzovaných knihoven své záznamy formou propojených dat zpřístupňuje 68 % knihoven, nejvyšší zastoupení lze sledovat u knihoven národních.

4.4. Uplatnění propojených dat v knihovnách

Pokud knihovny chtějí sledovat trendy a přizpůsobovat se jim (což činí např. v oblasti propagace a obecně komunikace na sociálních sítích), měly by svá cenná data zpřístupňovat takovým způsobem, aby je uživatelé mohli pohodlně vyhledávat za pomoci běžných vyhledávačů. S pojmem, resp. s myšlenkou propojených dat (*linked data*) se poprvé setkáváme opět u Timothyho Barners-Leeho roku 2006. Propojená data jsou publikačním modelem pro zveřejňování strukturovaných dat na webu, který je založen na webových standardech jako HTTP a URI a technologiích sémantického webu jako je datový model RDF. (MYNARZ, 2010) Uplatňovat propojená data znamená vytvářet vazby mezi údaji z různých zdrojů. Může docházet k propojování velmi různorodých dat vytvářených spolu nesouvisejícími organizacemi v různých částech světa, nebo naopak mohou být součástí heterogenních systémů jediné organizace. (BIZER, 2009) Bizer, Heath a Barners-Lee roku 2009 rozšiřují myšlenku webu dat na „web věcí na světě popsané podle údajů na webu“. Námi vytvořená data tak pak můžou libovolně používat i další uživatelé pro své vlastní aplikace, a tím naše data obohacovat (např. propojení bibliografických záznamů s informacemi z Wikipedie, zpravodajskými servery (BBC, New York Times).

Možností využití propojených dat v knihovnách se aktivně zabývá hlavně Kongresová knihovna a OCLC. Prakticky je ale využívá už velké množství knihoven, např. Švédská národní knihovna, Finská národní knihovna, Německá národní knihovna, Maďarská národní knihovna nebo jsou využívány v digitální knihovně Europeana.

Životní cyklus bibliografických dat se proměňuje nebo již proměnil. Přestože knihovníci stále katalogizují převážně ve formátu MARC 21 a knihovny data v tomto formátu mezi sebou distribuují, přibyl na konci článku převod do struktury propojených dat. Díky této struktuře se cenné informace vytvářené knihovníky konečně dostávají z uzavřených knihovních katalogů a databází (v angličtině je často používaný výraz *silos*) i na volný web, kde je naleznou koncoví uživatelé – stroje, aby je zprostředkovali dalším koncovým uživatelům – čtenářům. Ti je nejenom jednoduše naleznou již při prvotním vyhledávání informací z vyhledávačů typu Google, ale díky propojování naleznou data obohacená, propojující na další zdroje mimo katalogy. Knihovny tak mají šanci lépe zpřístupnit své záznamy.

4.5. Doporučený postup převodu do struktury propojených dat v knihovnách

Z praxe zahraničních knihoven je zřejmé, že na projektech převodu bibliografických záznamů do struktury propojených dat se obvykle podílí tým několika lidí (do 10) a celý projekt trvá dny až týdny. Lze jít samozřejmě různými cestami: v americké praxi je stále častěji běžné, že knihovna zadá převod externí firmě, která vybrané záznamy převede tzv. na klíč. Tento způsob je jednoduchý, avšak vždy dojde k převodu pouze ucelené a neměnné skupiny dat. Dalším možným způsobem je katalogizace do struktury propojených dat např. za využití volně dostupného nástroje Kongresové knihovny. Třetím způsobem je samostatná práce přímo v knihovně.

Pokud se knihovna rozhodne převést data (záznamy) sama, postupuje v jednotlivých krocích. Převod dat do struktury propojených dat, a tím jejich zpřístupnění široké veřejnosti, by měl začít vybráním vhodných dat, v knihovnictví vybráním vhodné skupiny bibliografických záznamů. Instituce by měla nejlépe vybírat takové záznamy, které jsou unikátní. Dalším krokem je vybrání vhodného slovníku, resp. termínů z různých slovníků. Následuje transformace. Transformace neboli převod dat ze struktury formátu MARC do struktury RDF je procesem

mapování struktury bez ohledu na obsah. Jsou při něm automaticky vytvářeny URI, aniž by byly propojovány. Je vhodné, aby URI vycházelo z neměnné webové adresy, je třeba počítat s tím, že URI by se nemělo měnit. Aby využití propojených dat mělo smysl, je vhodné pokud možno najít ke všem „věcem“ již existující URI nebo nově vytvořená URI propojit. Jenom tak lze následně těžit z propojitelnosti, protože pokud ostatní použijí pro stejnou „věc“ stejné URI, trojice se propojí. Vytvořená data ve struktuře propojených dat je nutné uložit na bezpečné místo serveru. Data jsou ukládána do RDF databází nebo lze využít např. datahub.io. Vytvořená data jsou v posledním kroku zpřístupněna uživatelům nebo třetím stranám.

4.6. Doporučený postup převodu do struktury propojených dat v českých knihovnách

České knihovny (Národní knihovna České republiky) mohou vyčkávat (jak některé jiné země dělají), aktivně se do propojených dat pustit s cílem lepší dostupnosti bibliografických záznamů (tedy následně dokumentů) pro své uživatele nebo počkat na případné nařízení např. vlády (viz zkušenost z Finska, kde publikování dat ve formě propojených dat je pro data vytvářená ze státního rozpočtu povinná). Cesta samozřejmě může být i obráceně: Národní knihovna České republiky může začít sama a prostřednictvím lobby a úpravy legislativy přimět ostatní české knihovny ke spolupráci nebo knihovny motivovat prostřednictvím grantů Ministerstva kultury České republiky. Je třeba si však uvědomit navázanost na knihovní software a celkově ekonomickou stránku věci. Je vhodné do struktury převádět unikátní data. Vhodnými záznamy by tak byly záznamy z české národní bibliografie. Prvním krokem by mělo být vytvoření URI pro soubor českých národních autorit. Pro převod autorit by bylo možné využít identifikačního čísla autority. Jako základ lze doporučit BIBFRAME, který je právě pro bibliografické záznamy knihoven určen, je velmi rozšířený a knihovnická komunita v něm vidí případného budoucího leadera.

4.7. Experimentální převod a analýza existujících záznamů

V disertaci bylo vyzkoušeno několik možných způsobů převodu bibliografického záznamu do struktury propojených dat nebo vytvoření záznamu přímo v editoru. Byl tak vytvořen nový záznam tzv. s knihou v ruce přímo v editoru BIBFRAME. Dále byl převeden existující záznam ve struktuře XML do struktury BIBFRAME.

Byly analyzovány záznamy ve struktuře BIBFRAME, ve struktuře vlastních propojených dat Německé národní knihovny, záznam z katalogu WorldCat ve struktuře propojených dat a záznam Britské knihovny. U záznamů byly sledovány jak vybrané slovníky, použití termínů slovníků pro jednotlivé údaje, tak i samotné vyjádření hodnot.

Záznamy vytvořené analyzovanými institucemi používají různé slovníky a jejich využití se neshoduje ani pro jednotlivé výrazy použité pro vyjádření bibliografických údajů. Nelze tedy doporučit žádný konkrétní slovník jako nejpoužívanější, ani žádné konkrétní slovníky pro vyjádření určitých výrazů. Zkoumáním bylo dále zjištěno, že při vytváření nebylo maximálně využito potenciálu propojených dat a velké množství údajů bylo vyjádřeno slovně (literály). Toto platí v nejvyšší míře pro konverzi bibliografického záznamu z XML do BIBFRAME, která sice záznam vyjádřila ve formě propojených dat, údaje však ve formě URI dle předpokladu sama nevyjádřila.

Zároveň bylo zjištěno, že do struktury propojených dat jsou převáděny pouze základní údaje. Např. kódované informace uváděné ve struktuře formátu MARC 21 nejsou ve struktuře propojených dat (vyjma jazyka) vyjádřeny. Tímto zjištěním však nelze říci, že by záznamy byly nedostatečné. Pro účely publikování záznamu a jeho zpřístupnění jsou informace dostačující.

4.8. Přípravenost knihovních softwarů

Přípravenost knihovních softwarů je klíčová pro budoucí vývoj. Vývoj knihovních softwarů je časově i finančně nákladná činnost a jejich modernizace úzce souvisí s tím, jaké knihovny mohou používat technologie. Lze samozřejmě polemizovat nad tím, zda pokud knihovny změni formáty dat, producenti se přizpůsobí, či nakolik změny na straně producentů uspíší rozvoj na straně knihoven. Část disertace je proto věnována analýze knihovních softwarů a jejich následujícímu srovnání. Při jejich zkoumání byl využit seznam producentů knihovních softwarů evidovaných v databázi vytvořené Marshalllem Breedingem *Library Technology Guides*¹, který byl doplněn o producenty z České republiky.

¹ Na webové stránce Library Technology Guides (librarytechnology.org) tento americký knihovních uveřejňuje informace týkající se produktů a služeb používaných knihovnami. Kromě toho také vytváří a zveřejňuje elektronický adresář knihoven libraries.org.

Bylo analyzováno 45 knihovních softwarů s celkovým počtem 20 688 zjištěných instalací od 22 producentů. Všechny uvedené knihovní softwary umožňují export ve struktuře formátu MARC. Celkem 12 knihovních softwarů (odpovídá 27 %) s 6 485 instalacemi (odpovídá 31 %) podporuje práci s propojenými daty. Zbýlých 33 knihovních softwarů (odpovídá 73 %) s 14 203 instalacemi (odpovídá 69 %) s propojenými daty s jistotou nepracuje nebo se informaci nepodařilo dohledat. S propojenými daty tedy pracuje nebo může pracovat necelá třetina knihoven. O uvedený počet se zasloužila hlavně OCLC.

Z výsledků analýzy lze odvodit, že knihovní softwary resp. knihovny nejsou na práci s propojenými daty připraveny.

4.9. Postoje klíčových hráčů při tvorbě standardů

Pro odhad budoucího vývoje byla vybrána jedna z kvantitativních metod, a to dotazníkové šetření formou elektronického dotazování. Výhodou této metody bylo oslovení zahraničních odborníků bez ohledu na místo jejich pobytu.

Toto dotazníkové šetření formou kladení otevřených otázek je základem tzv. Delfské metody, díky které jsou zjišťovány názory skupiny expertů, a to nezávisle na sobě.

Dotazníkový průzkum byl proveden na třech výzkumných vzorcích. Každý z těchto vzorků je tvořen skupinou odborníků: prvním byl Stálý výbor Sekce katalogizace Mezinárodní federace knihovnických sdružení a institucí (IFLA Cataloguing Section Standing Committee), druhým byli členové Poradního výboru MARC Kanceláře pro rozvoj sítí a standardů Kongresové knihovny (MARC Advisory Comitee, Network Development and Standard Office) a třetím členové e-mailové konference BIBFRAME. Dotazníkové šetření bylo provedeno během ledna a února 2016 a během července 2016.

Celkem 13 respondentů z počtu 43 (odpovídá 30 %) se domnívá, že proces nahrazení formátu MARC 21 jiným formátem již probíhá, dalších 13 se domnívá, že k nahrazení dojde do 5 let, dalších 9 do 10 let. Sledujeme-li pouze samostatné odpovědi, členové Stálého výboru Sekce katalogizace změnu očekávají nejčastěji do 5 let, za již probíhající ji nepovažuje nikdo a nahrazení v delším časovém horizontu taktéž nikdo neočekává. Z této odpovědi nelze usuzovat na tlak na změnu v oblasti výměnných formátů. Tomu neodpovídá postoj členů Poradního výboru

Kanceláře pro rozvoj formátu MARC: celých 78 % z nich považuje nahrazení formátu MARC 21 jiným formátem za nutnou, 56 % se domnívá, že ke změně již dokonce dochází. Z uvedeného je zřejmé, že nelze očekávat tlak „shora“, ale naopak že ke změně dojde tlakem uživatelů, chceme-li knihovníků, kterému se následně přizpůsobí standardy.

Další otázka byla směřována na postoj k propojeným datům: zda jsou to právě propojená data, která by v budoucnosti mohla formát MARC 21 nahradit, příp. zda ho nahradí pro sdílení nebo dokonce při tvorbě bibliografických záznamů.

Členové Stálého výboru Sekce katalogizace se shodují, že právě propojená data formát MARC 21 nahradí (bude-li tedy nahrazen), shodného názoru jsou i členové Poradního výboru MARC. Všichni odpovídající se pak shodují, že propojená data budou používána jak pro výměnu, tak i pro vytváření bibliografických dat.

Všichni respondenti se zároveň kloní k názoru, že **nebude existovat vedoucí struktura propojených dat, ale budou existovat různé varianty. Pokud by některá varianta měla mít převahu, byl by to BIBFRAME.** Z rozložení odpovědí je zřejmé, že nebude vyvíjen tlak na jednotnou strukturu ani asociací IFLA, ani samotnými knihovníky.

5. Závěr

Nejdříve s pokusy o automatizaci začaly Spojené státy americké, konkrétně Kongresová knihovna, a to v 50. letech 20. století, které vyústily vytvořením výměnného formátu MARC. Postupně začaly vznikat různé národní varianty, do roku 1986 jich vzniklo celkem 19. Tím vyvstala potřeba vzniku převaděčů, které umožňovaly konverzi dat a tím jejich směnitelnost. Různorodost formátů nebyla neřešitelným problémem, ale převody byly pracné a vyžadovala vytváření a permanentní údržbu konverzních tabulek. Náročnost konverze pak byla závislá na kvalitě převáděných záznamů, konkrétně na přísném dodržování katalogizačních pravidel a instrukcí pro formát. Jednotnost práci usnadňovala.

Nutnost vytváření mnoha převaděčů měl odstranit nově vzniklý mezinárodní výměnný formát Univerzální MARC – UNIMARC (Universal MARC), který navíc na rozdíl od amerického formátu MARC nebyl závislý na žádných katalogizačních pravidlech.

V devadesátých letech však začalo docházet k postupnému úpadku formátu v důsledku klesající podpory i uživatelské základny a malé flexibility a schopnosti reagovat na vznik nových dokumentů (zejména elektronické zdroje). V březnu 2003 převzala rozvoj formátu UNIMARC portugalská Národní knihovna v Lisabonu. Snížily se i finanční investice do jeho rozvoje a díky nedostatečným a pomalým modifikacím jeho význam postupně upadá a je nahrazován formátem MARC 21. (EHRLICHOVÁ, 2009)

Během této doby docházelo u amerického formátu MARC s názvem USMARC ke změnám. Byly tendence sjednotit americký, britský a kanadský MARC. Britská knihovna (British Library, BL) však s modifikacemi nesouhlasila, a tak v devadesátých letech vznikl nový formát ve spolupráci Kanadské národní knihovny (The National Library of Canada) a Kongresové knihovny ve Washingtonu – MARC 21, který se však od původního formátu USMARC lišil jen minimálně. Na harmonizaci obě knihovny pracovaly od roku 1994 a roku 1998 oficiálně ohlásily ukončení sjednocování obou formátů, což vyvrcholilo vydáním formátu pod společným jménem MARC 21 roku 1999.

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. (KTD a MARC 21 formats: background and principles, 1996) **V praxi jsou používány i jako formát při vytváření záznamů. V tuto chvíli je však zřejmé, že v budoucnu se odštěpí další funkce pro publikování bibliografických dat na webu.**

Pro zhodnocení současného stavu byly provedeny tři různé analýzy: zjištění formátu dat, ve kterém jsou poskytovány národní bibliografie, analýza formátu dat poskytovaných prostřednictvím protokolu Z39.50 a analýza znakových dat dle seznamu datahub.io. Z výsledků je zřejmé, že vedoucím formátem současného knihovnictví je formát MARC 21. V jeho struktuře je poskytováno 81 % národních bibliografií. Naopak formát UNIMARC je na ústupu. Vedoucí úloha formátu MARC 21 je zřejmá i z formátů dat poskytovaných přes protokol Z39.50: data v něm poskytuje tři čtvrtiny knihoven, ve formátu UNIMARC pouze desetina knihoven. Jednotlivé národní varianty formátu MARC využívá od 2 do 11 % knihoven, které však své záznamy zároveň poskytují i ve formátu MARC 21 nebo UNIMARC. Analýza datových sad dle datahub.io nevypovídá nic o zastoupení formátu MARC 21 nebo UNIMARC, je však zajímavá z hlediska propojených dat. Z analyzovaných knihoven, které poskytují záznamy ve formě propojených dat, lze nejvyšší zastoupení sledovat u knihoven národních.

V knihovnické obci se dlouho debatuje o zastaralosti, a tedy nutnosti náhrady, formátu MARC 21. Dotazníkovým šetřením mezi odbornou veřejností byly proto zjišťovány vize a doporučení cesty, kterou by se výměna bibliografických záznamů mohla nebo měla ubírat.

Třetina respondentů se domnívá, že proces nahrazení formátu MARC 21 propojenými daty již probíhá, další menší část se domnívá, že k nahrazení dojde do 5 let nebo nejpozději 10 let. Provedená analýza navíc ukazuje, že změnu brzy očekávají nebo po ní volají spíše knihovníci, kteří jsou zainteresovaní do vývoje formátu MARC 21 (tedy členové Poradního výboru MARC Kongresové knihovny), než členové Stálého výboru Sekce katalogizace IFLA.

Z uvedeného je zřejmé, že nelze očekávat tlak „shora“, ale naopak že ke změně dojde tlakem uživatelů, chceme-li knihovníků, kterému se následně přizpůsobí standardy v oblasti knihovnictví.

Z těchto analýz je dále zřejmé, že **knihovníci doporučují a předpokládají využití propojených dat, a to pro oblast publikování dat na webu, často ale i pro samotnou výměnu nebo i pro katalogizaci. Z propojených dat pak nejvíce naději vzbuzuje BIBFRAME, přestože odborná veřejnost nepředpokládá jednotnost v použitých schématech. Překážkou na cestě k propojeným datům mohou být používané knihovní softwary, které, jak ukázala jejich analýza, nejsou na propojená data připraveny, a nedostatek financí pro změnu.** Za tímto účelem již využívají konverzních programů, které jsou volně dostupné, volí komerční konverzi firmami, které se na vývoji podílely, případně se pouští do převodu samy.

Převod bibliografických záznamů do struktury propojených dat lze rozdělit do několika fází. První fází je výběr bibliografických záznamů pro převod. Měly by to být záznamy bezchybné, bez duplicit, zároveň by to měly být záznamy unikátní pro vytvářející instituci, tedy např. národní bibliografie dané země. Zároveň by měla být definována data ze záznamů, která budou určena pro převod, a to dle budoucích uživatelů. Bibliografické záznamy ve formátu MARC 21 jsou na informace velmi bohaté, je tedy možné (a pravděpodobné), že ne všechna pole formátu budou převedena. Druhým krokem je vybrání slovníku. Výběr slovníku by se měl vždy přizpůsobit instituci, která záznamy vlastní, i skupině budoucích uživatelů. Některé slovníky si vystačí s pojmem „titul“, slovníky vytvořené pro oblast knihovnictví jsou bohatší. Následujícím krokem je dohledání URI pro všechny „věci“ a vytvoření pravidla pro vytváření nových URI pro „věci“ unikátní, pro která ještě URI vytvořena nebyla. Vždy je však potřeba se maximálně snažit URI dohledat (nebo na něj odkázat), protože jenom tak lze maximálně vytěžit přidanou hodnotu propojených dat. Výběr způsobu zápisu je již prací spíše pro programátory.

Vytvořené záznamy jsou následně uloženy na servery, ze kterých je již přebírají aplikace vlastní i třetích stran.

Primárně jsou takto převáděny jednotlivé národní bibliografie, příp. ucelené celky bibliografických záznamů.

Reinhold Heuvelmann z Německé národní knihovny předpovídá zánik posledních knihovních systémů podporujících strukturu formátu MARC v roce 2060, kdy

i poslední záznamy v tomto formátu zmizí. Ale ještě před tím vyjde v roce 2047
článek. A jeho název? BIBFRAME must die². (HEUVELMANN)

² Název odkazuje na slavný článek R. Tennanta MARC must die.

6. Seznam pramenů citovaných v tezích

- BIZER, Christian; HEATH, Tom; BARNERS-LEE, Tim. *Linked Data: the story so far* [pdf]. 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 WWW: <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>
- EHRLICHOVÁ, Klára. *Osobnost a dílo Henrietty D. Avramové (1919-2006)*. 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.
- *KTD: Česká terminologická databáze knihovnictví a informační vědy (TDKIV)* [online databáze]. Praha: Národní knihovna České republiky, 2003-. Dostupné z WWW: <http://sigma.nkp.cz/F/>
- *MARC 21 formats: backround and principles*. Revised November 1996 [cit. 2016-03-26]. Dostupné z WWW: <https://www.loc.gov/marc/96principl.html>
- MYNARZ, Jindřich a Jan ZEMÁNEK. Úvod k linked data. *Knihovna PLUS* [online]. 2010, (1), 1-11 [cit. 2016-02-29]. ISSN 18015948. Dostupné z: <http://search.ebscohost.com/login.aspx?direct=true&db=lls&an=57658628&scope=site>