

Standards a technologie propojených otevřených dat: Mohou nám propojená otevřená data a jejich standardy pomoci řešit problémy současné katalogizační praxe v České republice?

Open Linked Data Standards and Technologies: Could Linked Open Data Standards Help to Solve Problems of Current Cataloguing Practice in the Czech Republic?

PhDr. Barbora Drobíková, Ph.D. / Ústřední knihovna, Univerzita Karlova (Charles University in Prague),
Josef Martího 407/2, 162 00 Praha 6; ORCID 0000-0002-6340-5940
Mgr. Michal Denár / Městská knihovna Česká Třebová (Municipal library Ceska Trebova),
Smetanova 173, 560 02 Česká Třebová
Bc. Josef Moravec / Moravská zemská knihovna (Moravian Library), Kounicova 65a, 601 87 Brno
Mgr. Lenka Maixnerová, Mgr. Marie Balíková, PhDr. Petra Štastná / Národní knihovna
České republiky (National Library of the Czech Republic, Mariánské náměstí 190/5, 110 00 Praha 1

RESUMÉ: Studie se zabývá možnostmi implementace propojených dat (linked data) v českých knihovnách. Analyzuje současný stav zpracování bibliografických a autoritních dat v České republice a navrhuje postupy pro přechod na nové formáty. Zdůrazňuje nutnost modernizace, která by vedla ke zlepšení přístupu k informacím a efektivnější kooperaci mezi knihovnami a dalšími institucemi. Studie představuje výzvy a přínosy této změny pro české knihovnictví.

KLÍČOVÁ SLOVA: propojená data, BIBFRAME, RDF, IFLA LRM, metadata, katalogizace, formát MARC, spolupráce knihoven

SUMMARY: The study examines the possibilities of implementing linked data in Czech libraries. It analyzes the current state of processing bibliographic and authority data in the Czech Republic and proposes procedures for transitioning to new formats. The study emphasizes the need for modernization, which would lead to improved access to information and more efficient cooperation between libraries and other institutions. It outlines the challenges and benefits of this change for Czech librarianship.

KEYWORDS: linked data, BIBFRAME, RDF, IFLA LRM, metadata, cataloguing, MARC format, library cooperation

Tato studie vznikla na základě institucionální podpory dlouhodobého koncepčního rozvoje Národní knihovny ČR jako výzkumné organizace poskytované Ministerstvem kultury ČR (DKRVO 2024-2028), oblast 11: Propojená otevřená data.

Úvod

V prosinci roku 2006 byla ustanovena v Kongresové knihovně (Library of Congress, Washington, D.C.) pracovní skupina pro budoucnost bibliografické kontroly (*Working Group on the Future for Bibliographic Control*) pod vedením José-Marie Griffiths (Univerzita v Severní Karolině v Chapel Hill). Skupina měla mj. za úkol shromáždit nové poznatky o vlivu standardů pro zpracování bibliografických a autoritních záznamů a katalogizačních postupů na správu informačních zdrojů v knihovnách a v přístupu k nim v novém informačním a technologickém prostředí (Library of Congress, 2006).

Skupina již v lednu 2008 zveřejnila významnou zprávu s názvem: *O záznamu (On the Record)* (Library of Congress. Working Group on the Future of Bibliographic Control, 2008).¹ V kapitole 3.1 *Web jako infrastruktura* konstatovala, že formát MARC je postaven na čtyřicet let starých technikách programování a není v souladu s aktuálními styly programování. Formát MARC je používán výhradně v prostředí knihoven a není kompatibilní s jinými systémy pracujícími s bibliografickými údaji. Širší použití bibliografických dat vyžaduje formát, který bude akceptovat a rozlišovat metadata vytvořená odborníky, generovaná automatizovaně a vytvořená uživateli, včetně anotací (recenzí, komentářů) a údajů o použití zdroje.

Na základě doporučení formulovaných v této zprávě Kongresová knihovna vyhlásila 31. 10. 2011 iniciativu: Bibliografický rámec pro digitální věk (*Bibliographic Framework for the Digital Age*) – zkráceně **BIBFRAME** (Library of Congress, 2011).² Vyhlášení iniciativy BIBFRAME bylo jedním z důležitých impulsů v rozvoji nových formátů bibliografických dat na bázi modelu RDF³ (*Resource Description Framework* = Rámec pro popis zdrojů) – v českém jazyce propojená data, v angličtině obecně známé jako „linked data“. Argumentem pro volbu modelu RDF bylo mj. to, že je to metoda doporučená konsorciem W3C (*World Wide Web Consortium*) pro koncepční popis nebo modelování dat ve webovém prostředí.

Použití RDF a dalších technik podpořených konsorciem W3C obecně umožňuje lepší integraci dat z knihovních systémů a dalších systémů z oblasti kulturního dědictví v prostředí internetu s cílem pokročilejšího a širšího přístupu uživatelů k informacím (Library of Congress, 2011). Jedním z hlavních výsledků této iniciativy je vytvoření ekosystému modelu, ontologie a dalších nástrojů pro tvorbu a správu propojených dat se stejnojmenným názvem BIBFRAME, který je postupně implementován ve vybraných knihovních databázích po celém světě.

Vedle formátu BIBFRAME je dalším významným počinem v této oblasti rozvoj ontologie RDA na bázi RDF vycházející z katalogizačních pravidel RDA: Resource Description and Access – verze Official⁴. *RDA Official* spolu s BIBFRAME reprezentují dvě navzájem se významně ovlivňující a v současnosti též postupně se doplňující iniciativy, které je

1 V anglickém jazyce název: *On the Record*.

2 V této souvislosti nelze též opomenout legendární článek Roye Tennanta: „MARC must die“ [MARC musí zemřít], který publikoval již v roce 2002 a v němž se zamýšlí nad zastaralostí struktury formátu a technologií s ním souvisejících.

3 RDF – Resource Description Framework

4 Katalogizační pravidla RDA jsou v úplnosti dostupná prostřednictvím nástroje *RDA Toolkit*. Nejdříve byla sepsána ve struktuře modelů FRBR a FRAD jako běžné paragrafy, které známe již z předchozích katalogizačních standardů, např. AACR2R. Tato verze je označována jako *Original*, poslední aktualizace proběhla v dubnu 2017. Verzi těchto pravidel používáme též v české katalogizační praxi. V současnosti je ovšem rozvíjena verze *Official*, která je postavena na modelu IFLA LRM (viz dále v textu) a zejména je úzce spjata s ontologií RDA, tzv. RDA Registry a slovníky hodnot. Jednotlivé entity, atributy a vztahy mají přiděleno jednoznačné IRI (Internationalized Resource Identifier – Mezinárodní identifikátor zdroje; je založen na principech URI). Plánuje se, že tato verze bude od roku 2026 jediná platná.

nutné vnímat jako nástupce formátu MARC 21 a katalogizačních pravidel RDA ve verzi *Original* (využívaná v současnosti v ČR) pro bibliografická i autoritní data.

Jak na tento vývoj budeme reagovat v České republice? Je možné i v našem prostředí postupně změnit formáty dat v knihovnách? Co vše by bylo nutné připravit, aby k takové změně mohlo dojít?

Cíl studie a použité metody

Cílem studie je analyzovat možnosti implementace a využití formátů propojených dat v prostředí českých knihoven. Na základě rešerše a analýzy zahraničních zdrojů představíme problematiku formátů propojených dat a možné výhody jejich implementace. Zhodnotíme současný stav zpracování a kooperace v oblasti bibliografických a autoritních dat v ČR se zaměřením na přípravu dat k možné konverzi do formátů propojených dat. Pro výzkum jsme využili zejména analýzu bibliografických a autoritních dat České národní bibliografie. Rámcově popisujeme procesy zpracování bibliografických a autoritních dat v České republice. Identifikujeme oblasti pro zlepšení a optimalizaci výměny a sdílení dat v knihovní síti i pro kooperaci mezi knihovnami a okolními systémy, zejména nakladatelskou sférou. Uvádíme výhody implementace formátů propojených dat pro kooperaci v knihovnách i mezi knihovnami a okolními systémy ve webovém prostředí. Přířínosem studie je nástin řešení implementace ekosystému propojených otevřených dat v České republice.

Od formátu MARC k propojeným datům

Zpracování bibliografických a autoritních dat je v České republice ovlivněno několika standardy. Od roku 2004 je používán jako hlavní výměnný formát MARC 21, od roku 2015 jsou používána – v kombinaci s formátem MARC 21 – katalogizační pravidla RDA ve verzi *Original*. Tyto mezinárodní standardy jsou doplňovány metodikami a interpretacemi, které jsou zveřejňovány na stránkách Národní knihovny ČR věnovaných katalogizační politice.⁵ V průběhu předchozích téměř dvou století se však na našem území používala řada jiných standardů a metod. Mnoho fondů bylo historicky zpracováváno na katalogizačních lístcích dle různých standardů. Lístky byly do strojem čitelné podoby převáděny teprve od 90. let 20. století v rámci retrokonverzních (strojově čtení lístků) nebo retrokatalogizačních (s dokumentem v ruce) projektů. V dnešních databázích proto najdeme opravdu směsici mnoha různých přístupů a pravidel.

Protože jsme v ČR přijali MARC 21 jako standard v roce 2004, mohlo by se zdát, že se jedná o poměrně moderní formát. Opak je ovšem pravdou. MARC 21 úzce navazuje na své předchůdce, jejichž vznik spadá už do 60. let 20. století. Formát má tedy již téměř šedesátiletou historii, a jak se hovořilo ve zprávě *O záznamu* (Library of Congress. Working Group on the Future of Bibliographic Control, 2008), byl vytvořen z dnešního pohledu pomocí poměrně starých technik programování a správy dat. Formát MARC 21 je úzce spjat s anglo-americkými katalogizačními postupy, jeho podoba byla ovlivněna podobou papírového katalogizačního lístku. Pro strukturu dat v některých polích a podpolích je nutné stále používat interpunkci podle standardu ISBD (*International Standard Bibliographic Description* = Mezinárodní standard pro bibliografický popis). Celková struktura je velmi zastaralá a neflexibilní. Neumí dobře reagovat na současné datové modely

5 Bliže viz webová prezentace Národní knihovny ČR: Katalogizace, zpracování: <https://ipk.nkp.cz/odborne-cinnosti/knihovni-procesy/katalogizace-zpracovani>.

(např. *IFLA Library Reference Model*, dále jen IFLA LRM, Riva, 2017). Formát MARC 21 se často používá v knihovních systémech nejen jako výměnný formát dat. Knihovní systémy mnohdy nabízejí pro katalogizaci formuláře postavené na jednotlivých polích formátu MARC, pole jsou označena příslušnými tagy, knihovník zapisuje indikátory, označuje jednotlivá podpole, musí využívat předepsanou interpunkci.

MARC 21 (nebo i jiné formáty typu MARC) je používán výhradně knihovní komunitou a je v podstatě nesrozumitelný pro jiné systémy používané v paměťových institucích (archivech, muzeích) nebo v systémech nakladatelského a knižního trhu. Z výše uvedených důvodů i pro lepší komunikaci dat s širším okolím knihoven v síti internet, a zejména pro snazší komunikaci dat ve webovém prostředí by bylo vhodnější formáty MARC⁶ opustit a začít používat formáty postavené na obecnějších standardech, které jsou využívány i mimo knihovnictví.

S formáty MARC a s technologiemi s nimi souvisejícími (např. Z39.50) pracují knihovny po celém světě již opravdu dlouhou dobu, pomocí nich jsou sdíleny stamiliony bibliografických a autoritních záznamů. „Na formátu MARC 21 je postaveno vše od systémové integrace až po veškeré katalogizační práce“, jak konstatují odborníci ze Švédské národní knihovny (National Library of Sweden, 2019). Je proto značně obtížné tuto praxi ukončit a začít používat zcela nové postupy a techniky. Abychom mohli zhodnotit možnosti konverze stávajících bibliografických a autoritních dat do formátů propojených dat, je nutné podrobněji prozkoumat procesy tvorby bibliografických a autoritních záznamů v ČR a zejména kooperaci při jejich vytváření.

Především blíže vysvětlíme, jak se pracuje s propojenými daty v zahraničních knihovnách a jaký vliv by mohla mít implementace formátů propojených dat na celý proces zpracování dat v knihovnách.

Téma propojených dat v knihovnách

Využití propojených dat v knihovních systémech není v ČR novým tématem. Již v roce 2010 Jindřich Mynarz a Jan Zemánek (Mynarz a Zemánek, 2010) publikovali článek *Úvod do linked data* v periodiku *Knihovna plus*, kde charakterizují principy formátů propojených dat i jejich (jak sami uvádějí) technologický profil. Velkou část článku tvoří využití propojených dat v knihovnictví, jako příklad z ČR uvádějí převedení Polytematického strukturovaného hesláře do formátu SKOS⁷ (Národní technická knihovna, 2016–2024).

V dalších letech pak následují v českém jazyce články, které tato témata zmiňují, a to od autorek Barbory Drobíkové (2013, 2014), Kláry Rösslerové (2016, 2017a, 2017b, 2018) či Heleny Kučerové (2018, 2019)⁸. Významným počinem v této oblasti jsou projekty Národní knihovny ČR týkající se propojení terminologické databáze TDKIV a databáze jmenných autorit s Wikidaty, které jsou popisovány zejména v pracích Lindy Jansové (2019, 2020) a Zdeňka Bartla (2019). Možnosti využití propojených dat v databázi *Knihovny.cz* popisují autoři Michal Denár a Josef Moravec (2023).

6 Ani rámec vytvořený pro práci s formátem MARC 21 ve struktuře XML (tzv. MARCXML) nepřinesl výraznou změnu v komunikaci s okolními informačními systémy. I když byly díky tomu vyvinuty nové flexibilnější konverzní nástroje, vnitřní struktura a granularita formátu zůstala nedotčena, otevřenost struktury XML zde využita nebyla – <https://www.loc.gov/standards/marcxml/>.

7 Simple Knowledge Organization System – <https://www.w3.org/2004/02/skos/>

8 Řešitelský tým Heleny Kučerové a Evy Bratkové vytvořil Znalostní bázi pro obor organizace informací a znalostí. Báze je zpřístupněna ve formě propojených otevřených dat. Díky propojení obsahu a struktury znalostní báze s externími zdroji mají uživatelé zajištěn přístup i ke znalostem, jejichž zdroje se nacházejí vně znalostní báze (citace z webu báze – Kučerová, Bratková, 2015–2016 – <http://ko.cuni.cz/node/3253>).

Pro bližší seznámení s propojenými daty v knihovnách lze využít rozcestník na portálu *Informace pro knihovny* (Národní knihovna ČR, 2024)⁹. V roce 2023 se též uskutečnil webinář *Katalogizace a propojená data* pořádaný Svazem knihovníků a informačních pracovníků ČR. Vídea z webináře jsou dostupná na webu SKIP ČR.¹⁰

V zahraničí bylo na téma propojených dat v knihovnách v posledních dvaceti letech vydáno a napsáno již velké množství prací. Systematický přehled publikovaných zdrojů na téma propojených dat v knihovnách uvedl tým pod vedením Panorea Gaitanou v prvním čísle časopisu *Journal of Information Science* v roce 2024 (Gaitanou, 2024). Článek shrnuje práce vydané mezi lety 2008 až 2019. Zabývá se výhradně články publikovanými v odborných periodikách v anglickém jazyce, kapitolami z knih a statěmi ze sborníků. Vynechává obhájené diplomové či dizertační práce, bílé knihy (white papers) či podobné zdroje.

Výsledky systematického přehledu jsou rozděleny do několika kapitol:

- Implementace propojených dat v oblasti kulturního dědictví s podkapitolami: Implementace propojených dat v knihovnách a bibliografická kontrola, Implementace propojených dat ve specifických projektech, Specifické přístupy k propojeným datům a metodologie.
- Popis specifických bibliografických modelů s podkapitolami: Model FRBR¹¹, BIBFRAME a RDA.
- Problematika interoperability – mapování a převodní tabulky s podkapitolou Mapování a převodní tabulky používající model BIBFRAME.
- Další otázky s podkapitolami: KOS (*Knowledge Organization Systems* = systémy pořádní informací), Propojená data a kvalita metadat, Soukromí v knihovnách, Pozice knihovníka v prostředí propojených dat a Zdroje pro vzdělávání.

Autoři zpracovali celkem 239 zdrojů. Výše uvedené názvy kapitol a podkapitol jasně ukazují, kterými tématy se publikované zdroje v období let 2008–2019 nejčastěji zabývaly. Jsou to zejména témata týkající se bibliografické a autoritní kontroly, modelů IFLA LRM a BIBFRAME, interoperability metadat v knihovních systémech s důrazem na přechod na nové formáty propojených dat. Systematický přehled potvrdil fakt, jak sami konstatují jeho autoři, „že propojená data se stávají hlavním trendem v katalogizaci knihoven, zejména ve velkých knihovnách po celém světě, a také v nejdůležitějších výzkumných projektech iniciovaných knihovnami ve snaze učinit bibliografická data a knihovní sbírky dostupnější ve webovém prostředí a smysluplnější a znovupoužitelná pro své uživatele“ (Gaitanou, 2024, s. 218). Díky takto podrobnému přehledu je ale též zřejmé, že existují témata, kterých se zatím mnoho autorů ve svých výzkumných pracích z období let 2008–2019 nedotýká. Gaitanou a kolektiv konstatují, že jsou to zejména témata týkající se kontroly kvality metadat, chybí pravidla pro zacházení s (meta)daty a ke sdílení (meta)dat ve formátu RDF a další témata.

9 <https://ipk.npk.cz/odborne-cinnosti/propojena-data>

10 <https://www.skipcr.cz/knihovnicke-akce/katalogizace-propojena-data>

11 FRBR – Functional Requirements for Bibliographic Record – Funkční požadavky na bibliografické záznamy, předchůdce modelu IFLA LRM – <https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frbr/frbr-cs.pdf>.

Z novějších prací bychom rádi zmínili dizertaci Sophie Zapounidou z roku 2020 s názvem *Study of library data models in the Semantic Web environment*, v níž srovnává modely FRBR, BIBFRAME, RDA a EDM¹². Velmi inspirativní je též diplomová práce Julie Unterstrasser z roku 2023, v níž je ukázáno, jak přechod na formát propojených dat ovlivnil práci a praxi knihovníků ve Švédské národní knihovně. Autorka zdůrazňuje významný posun v práci katalogizátorů „**od cataloging ke catalinking**“, neboli **od katalogizace k vytváření propojení**, jako zcela zásadní změnu paradigmatu bibliografického a autoritního zpracování informačních zdrojů v knihovnách. Důležitý je též aspekt nutného dalšího vzdělávání knihovníků v oblasti propojených dat.

Význam tématu propojených dat neukazují jen publikované články či dizertace, ale též konkrétní probíhající projekty implementace formátů propojených dat do procesů bibliografické kontroly v mnoha velkých knihovnách po celém světě. Od roku 2017 se pravidelně každoročně koná evropská konference k formátu BIBFRAME – *BIBFRAME Workshop in Europe* (<https://www.bfwe.eu/>), která přináší aktuální zprávy o stavu implementace formátů propojených dat z jednotlivých zemí v Evropě s velkým množstvím aktuálních příspěvků od významných autorů v tomto oboru. Inspirativní jsou studie a konferenční příspěvky Iana Bigelowa se spoluautorkami z Univerzity v Albertě (např. 2020, 2022, 2023), Tiziany Possemato se spoluautory z knihovny Casalini Libri (např. 2020, 2022, 2023), která je organizátorem workshopu, nebo Nancy Lorimer ze Stanfordovy univerzity (např. 2022, 2023) či Sally McCallum z Kongresové knihovny (např. 2022, 2023) a mnoha dalších odborníků.

Množství již realizovaných projektů převodu bibliografických a autoritních dat do formátů propojených otevřených dat ukazuje též studie Proposal for the Publication of Linked Open Bibliographic Data z roku 2023 od autorů F. A. de Jesuse a F. F. de Castra, kteří identifikovali celkem 58 projektů z celého světa, projekty národních či univerzitních a specializovaných knihoven a sítí ve Španělsku, Finsku, Švédsku, Německu, Maďarsku a zejména Spojených státech amerických.

Potřebu přechodu na propojená data též ukazuje *3R Project*, jehož cílem bylo kompletní přepracování pravidel RDA ve verzi *Original* na RDA ve verzi *Official*. Nedílnou součástí RDA Official jsou RDA Registry – ontologie pro formát propojených dat RDA/RDF. Cílem bylo také kompletní přepracování nástroje RDA Toolkit, jednotlivých instrukcí a paragrafů s ohledem na využití pravidel v kombinaci s formáty propojených dat (např. Alemu, 2022, s. 197; Oliver, 2021). Je předpoklad, že od roku 2026 se budou využívat RDA již pouze jako pravidla ve verzi Official a verze Original bude zrušena.¹³

Entitně orientovaná katalogizace: jak se může proměnit správa bibliografických a autoritních dat

Již v roce 1995 publikoval Michael Heaney důležitou studii zabývající se objektově-orientovanou katalogizací (Heaney, 1995). Tehdy studii zasadil ještě do rámce katalogizačních pravidel AACR2R¹⁴. Heaney vyzýval k většímu důrazu na přesnou identifikaci různých typů autorit, které lze vzájemně propojovat. Síť propojených autorit by pak reprezentovaly

12 EDM – Europeana Data Model – https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation/EDM_Factsheet.pdf.

13 Více viz: <https://www.rdatoolkit.org/news/rda/countdown-clock-original-rda-toolkit-set-may-2026>.

14 V té době nebylo příliš mnoho známo o modelu FRBR, protože ten byl zveřejněn až v roce 1998. Formát RDF – Resource Description Framework se stal doporučením W3C v roce 1999 a s myšlenkou propojených dat přišel Tim Berners Lee v roce 2006.

jednotlivé záznamy. Lze říci, že toto vizionářské dílo v současnosti může dojít jistého naplnění. V kontextu formátů propojených dat se nově často objevují termíny jako *správa (management) identit* nebo *entit* či *entitně založená katalogizace* (např. Durocher aj., 2020; Stalberg aj., 2020; MacEwan, 2022; Zapounidou aj., 2024).

Technologie propojených dat je založena na přesné identifikaci strukturovaných dat, která reprezentují instance entit a vztahů mezi nimi pomocí jednoznačných identifikátorů URI (instance entit, vztahy a k nim příslušné URI jsou registrovány v řízených slovnících a ontologiích). Strukturovaná data reprezentující určité typy entit a kontrolované slovníky, jak píše Zapounidou aj. (2024), jsou vlastním srdcem katalogizačního procesu známého jako autoritní kontrola. V tomto bodě tedy procesy autoritní kontroly, jak je známe z knihovních bází, koincidují s technologiemi správy propojených dat. Správa propojených dat ovšem vyžaduje do značné míry automatizované zpracování založené na identifikátorech URI. V procesech autoritní kontroly se často spoléháme na lidskou interpretaci a využití textových řetězců, např. v rovině volby autorizovaných vstupních prvků podle různých kulturních a jazykových zvyklostí (viz podrobněji Zapounidou aj., 2024).

Pokud hovoříme o entitách, pak formáty propojených dat, ať už máme na mysli např. BIBFRAME nebo formáty založené na RDA/RDF a IFLA LRM, předpokládají správu veškerých entit vyskytujících se v bibliografických databázích. Nejedná se jen o entity, které jsou dnes v oblastech zájmu autoritní kontroly, jako jsou jména (personální, korporativní), názvy děl, geografická jména či předmět. V jazyce modelu IFLA-LRM (Riva aj., 2017) jsou to navíc entity týkající se vyjádření, provedení, jednotky, aktéra, jména či časového rozpětí. Pro instance veškerých entit je nutné spravovat slovníky hodnot s jednoznačnými URI. Identifikace dokumentu (dnes reprezentována bibliografickým záznamem) bude pak tvořena jako síť vzájemných vztahů mezi instancemi entit (jednotlivé výskyty entit, např. konkrétní osoba, konkrétní místo), přičemž jak vztahy, tak i jednotlivé instance entit budou reprezentovány identifikátorem URI (vzpomeňme výše citovanou myšlenku J. Unterstrasser (2023): „od *cataloging* ke *catalinking*“.

Výhody implementace propojených dat nejen pro katalogizaci

Doposud jsme se v článku zabývali pouze implementací formátů propojených dat do oblasti správy knihovních dat či pro katalogizaci. Výhoda nasazení formátů propojených dat se ovšem nejlépe projeví v propojení knihovních bází a externích zdrojů ve webovém prostředí, v lepším zviditelnění knihoven na webu a tím též v lepších službách uživatelům knihoven. Formáty typu MARC nejsou snadno srozumitelné pro jiné systémy mimo knihovnickou komunitu. Formáty propojených dat mohou umožnit lepší interoperabilitu dat mezi různými komunitami – nakladatelským prostředím, sektorem označovaným zkratkou GLAM – galeriemi, knihovnami, archivy, muzei. Publikování dat ve formátech propojených dat umožní webovým vyhledávačům snazší indexování dat z knihovních databází a zviditelnění dat při běžném vyhledávání ve webovém prostředí. Tato data umožní propojování knihovních databází s externími zdroji informací, jako jsou např. GeoNames nebo Wikidata, a obohacení uživatelských rozhraní knihovních katalogů a *discovery systémů* z externích zdrojů.

První výstupy projektů obohacování dat z externích zdrojů je možné vyzkoušet i v České republice např. v portálu *Knihovny.cz*, jak je popisují autoři Denár a Moravec (2023). Dalším příkladem dobré praxe může být nástroj *NKlink* (Jonáčková & Dostál, 2020), který obohacuje autoritní záznamy o externí identifikátory včetně identifikátorů

Wikidat. Možnosti obohacovat knihovní data o údaje z externích zdrojů patří k nejčastějším argumentům pro nasazení formátů propojených dat v knihovnách a nahrazení zastaralých formátů typu MARC. Protože zavedení propojených dat pomůže pracovat s daty výrazně efektivněji, lze tak nabídnout uživatelům nové možnosti při hledání, které jsou nyní dostupné jen komplikovaným způsobem nebo nejsou dostupné vůbec.

Situace v ČR

Kooperativní tvorba bibliografických záznamů

Kooperativní tvorba bibliografických záznamů je v ČR postavena na několika pilířích. Hlavním pilířem jsou centrálně definované standardy, jako jsou katalogizační pravidla (v ČR jsou to nyní RDA: *Resource Description and Access*, verze *Original*) a výměnný formát (MARC 21 pro bibliografické i autoritní záznamy). Další důležité pilíře tvoří budování České národní bibliografie (včetně souboru národních autorit) a tvorba souborných katalogů, do nichž mohou české knihovny jednak přispívat novými bibliografickými záznamy, jednak z nich mohou zpětně záznamy stahovat a využívat je v lokálních bázích.

Česká národní bibliografie (ČNB)

Národní knihovna ČR podle zákona 257/2001 Sb. (dále jen knihovní zákon) § 9, odst. 2b „zpracovává národní bibliografii a zabezpečuje koordinaci národního bibliografického systému“. Na této úloze dle následujících paragrafů téhož zákona spolupracují všechny krajské knihovny (§ 11, 2a) a specializované knihovny (§ 13, 2a).

Základ ČNB tvoří knihovny v rámci tzv. „clusteru“: Národní knihovna ČR, Moravská zemská knihovna v Brně (také v roli krajské knihovny Jihomoravského kraje), Vědecká knihovna v Olomouci (také v roli krajské knihovny Olomouckého kraje). Tyto knihovny pracují ve společné databázi. Dále jsou do ČNB přebírány záznamy krajských knihoven prostřednictvím Souborného katalogu ČR. Bibliografickými záznamy pro beletrii přispívá Městská knihovna v Praze v rámci projektu *Central* (Lichtenbergová, 2023).

Tvorba ČNB představuje jednu z tzv. národních funkcí Národní knihovny ČR. Databáze ČNB je velmi významným a svým zaměřením nenahraditelným zdrojem informací o publikovaném kulturním dědictví v České republice.

ČNB zpřístupňuje v současnosti 1,2 mil. záznamů. Podíl na tvorbě bibliografie může ilustrovat statistika původců záznamů¹⁵ tvořených podle nových pravidel RDA platných od roku 2015. Od tohoto roku bylo vytvořeno 209 tis. záznamů dle RDA, z toho vytvořila:

- 64 tis. – Národní knihovna ČR – sigla ABA001
- 19 tis. – Moravská zemská knihovna v Brně – sigla BOA001
- 18 tis. – Vědecká knihovna v Olomouci – sigla OLA001

Dohromady tedy tyto tři knihovny založily od roku 2015 zhruba polovinu bibliografických záznamů v ČNB. Na dalších sto tisících záznamech se podílely ostatní krajské

15 Sigla knihovny uvedená v poli 040, podpoli „a“ ukazuje na knihovnu, která jako první vytvořila záznam, neukazuje na jiné knihovny, které záznam mohly aktualizovat a obohacovat.

a další specializované knihovny, několik jednotek záznamů bylo původně založeno dokonce v zahraniční knihovně. Městská knihovna v Praze se podílela téměř 13 tisíci záznamy¹⁶.

Souborné databáze

Souborný katalog České republiky, centralizovaný heterogenní souborný katalog, obsahuje 8,4 mil. bibliografických záznamů¹⁷ včetně záznamů ČNB. Elektronicky je dostupný od roku 1995 (Svobodová, 2003). V současnosti na jeho budování spolupracuje 530 knihoven (Souborný katalog České republiky, 2023). Producentem Souborného katalogu ČR je dle zákona 257/2001 Sb., § 9, odst. 2a Národní knihovna ČR. Vedle ČNB je i budování Souborného katalogu součástí tzv. národních funkcí Národní knihovny ČR.

Do Souborného katalogu lze přispívat přes protokol OAI-PMH výhradně záznamy monografií. U seriálů je možné aktualizovat odběr jednotlivých titulů s využitím online formuláře. Frekvence přispívání je různá. Na vstupu do Souborného katalogu je kontrolována duplicitnost i kvalita záznamu. V případě, že záznam v Souborném katalogu ČR (dále jen SKC) již existuje, je připsána pouze sigla nové (zasílající) knihovny a vytvořena vazba do jejího lokálního katalogu. V případě, že je záznam nový, je celý vložen do SKC včetně vazby na lokální bázi. Pokud nemá záznam žádoucí kvalitu, je vrácen zasílající knihovně k opravě. Některé knihovny neposkytují pro sklizení do báze SKC celé své fondy, ale jen určitou zvolenou část, např. podle druhu dokumentu.

Jako dalšího zástupce souborných databází jmenujme **portál Knihovny.cz**, do kterého je nyní zapojeno 100 knihoven. Portál navíc zpřístupňuje další zdroje, které sklízí, včetně Souborného katalogu ČR (Knihovny.cz, 2024b). S využitím protokolu Z39.50 si mohou knihovny nechat definovat jednotlivé profily, které jim umožňují prohledávat a přebírat záznamy z různých českých i zahraničních zdrojů (Knihovny.cz, 2024a). Výhodou báze *Knihovny.cz* je, že často přebírá z knihoven záznamy větší části fondů. Je tedy poměrně zajímavým zdrojem například záznamů audioknih nebo speciálních dokumentů, jako jsou mluvicí knihy nebo deskové hry.

Vzhledem k roli portálu *Knihovny.cz* zde nemají tituly skutečně souborný záznam (SKC). Pro potřeby vyhledávání se pracuje s každým záznamem, jenž systém obdržel ze všech knihoven, které daný titul vlastní. Záznamy jsou propojené za pomoci deduplikace (Kurfürstová aj., 2023). V různých momentech se používá jiný záznam, například na základě domovské knihovny přihlášeného uživatele nebo výběru knihovny při vyhledávání. Přes Z39.50 jsou pak nabízeny všechny záznamy v indexu *Knihovny.cz*. Jejich kvalita je značně rozdílná. Zároveň ale často portál poskytuje záznamy, které nejsou dostupné v jiných zdrojích.

Současné workflow kooperace

Vedle nesporných výhod, které z budování ČNB i souborných databází pro kooperativní katalogizaci plynou, lze sledovat i jisté slabiny, které doprovázejí stávající modely kooperace po celou dobu existence knihovnických databází, a to nejen v elektronickém prostředí. Mohli bychom je shrnout pod dva pojmy, a těmi jsou *duplicitnost* a *asynchronnost*. To se v praxi projevuje například následující situací: Knihovna poptává hotový záznam právě doručeného titulu, například v Souborném katalogu ČR. Pokud ho tam nenalezne, musí

16 Data analyzována dle: <http://ddb.qa-catalogue.eu/nkp/?tab=data&lang=en>; aktuální k říjnu 2023; původce záznamu zpracován dle pole 040 podpole „a“.

17 https://aleph.nkp.cz/web/skc/_stat/skc_tp

ho vytvořit. Toto se může stát v rozpětí hodin či dní v několika knihovnách. Pokud knihovna spolupracuje se SKC, bude jí vytvořený záznam sklizen prostřednictvím protokolu OAI-PMH, a to někdy i s týdenní (či dokonce měsíční) periodicitou. V SKC pak vznikají multiplicitní záznamy o různé kvalitě, které se musejí složitým způsobem deduplikovat. Navíc nelze v deduplikovaných záznamech jednoduše aplikovat dílčí změny na základě záznamů zaslaných z knihoven. V určitém momentu systém už jen registruje, že titul existuje v dané knihovně.

Část práce, kterou generuje tato paralelní činnost, tak nemá pro kooperační systém význam. V jednotlivých knihovnách pak tímto způsobem vznikají záznamy s odlišnou úplností a kvalitou zpracování. To neznamená, že je systém špatně navržený. Všechny negativní vlastnosti vycházejí z vícerychlostní dynamiky distribuce záznamů v systému. Ta je způsobena částečně tím, že sklizení záznamů ze stovek knihoven je samo o sobě časově náročné, stejně jako následná deduplikace a další zpracování na straně souborných bází. Svůj podíl má i formát MARC, jehož struktura často neumožňuje efektivně kvalitativně hodnotit jednotlivé záznamy s ohledem na tvorbu kvalitního souborného záznamu. Používají se hodnotící váhy, postavené na algoritmech, ale ty dokáží pracovat spíše s formální kontrolou záznamů. Je to dáno také tím, že množství informací ve struktuře MARC je zapsáno pouze jako text, bez zasazení do širšího kontextu či přiřazení konkrétního významu k informaci. Celý proces dělá ještě složitějším komplikovaná syntaxe a pravidla pro tvorbu záznamů, které v některých ohledech umožňují různý přístup k popisu či zápisu.

ČNB je tvořena, jak jsme výše uvedli, zejména třemi hlavními knihovnami v clusteru. Tyto knihovny získávají vydané dokumenty na území ČR především formou tzv. povinného výtisku. To samo o sobě přináší často značné zpoždění ve zpracování dokumentů, které jsou již nějakou dobu k dispozici na trhu. Regionální či odborné knihovny mohou tyto dokumenty získat reálně časově dříve než „clusterové“ knihovny a jsou pak nuceny dokumenty originálně zpracovat. Vzhledem k požadavkům svých uživatelů často nemohou čekat až vznikne kvalitní záznam v ČNB, či než se dostane záznam pro stažení do souborných bází. I když ČNB tvoří základ pro SKC zejména v oblasti české produkce, může se stát, že se záznamy v ČNB od záznamů stejných titulů liší. Obsahují některé praktické informace, které ocení jak pracovníci ve výpůjčních službách, tak samotní čtenáři. Příkladem může být příslušnost k cyklu nebo kvalitní anotace.

Největším problémem současného kooperativního systému je značný počet duplicit a multiplicit záznamů. Ty vznikají tak, že do centrálních souborných databází přicházejí záznamy z lokálních bází, a to v rozdílném čase. To je dáno především akviziční politikou jednotlivých knihoven ve spojitosti s distribucí. Navíc knihovny do souborných databází posílají své záznamy v různé frekvenci. V případě beletrie s rychlostí katalogizace nových titulů výrazně pomohl projekt *Central*, řízený Městskou knihovnou v Praze (Projekt Central, 2024b). Beletrii nakupují 3× týdně a zpracují kolem 16 titulů denně. Uvádějí, že do Souborného katalogu odesílají záznamy novinek po 3–7 dnech od vydání. I zde však vzniká asynchronnost. Dle statistik projekt pokryje přibližně 80 % beletrie vydávané na našem území v největších českých nakladatelstvích (Projekt Central, 2024a).

Nástin budoucího řešení

Při modelu kooperace, kdy záznam vzniká až po vydání dokumentu a jeho uvedení na trh, se těmito slabínám nelze vyhnout. Tuto situaci by bylo možné pro mnoho případů

řešit vytvořením záznamu ještě před samotným vydáním dokumentu nebo – nejspíše – souběžně s jeho uvedením na trh.

Významným hráčem by mohla být nově budovaná společná databáze knihoven a nakladatelů s názvem *Registr českých knih*, někdy též zkráceně nazývaná ReČeK (Maixnerová, 2023). Díky ní by knihovny mohly získat základní metadata ještě před vydáním titulu. Projekt by mohl nabídnout použitelnou datovou alternativu „katalogizace v knize“, která se v našem prostředí rozšířila jen částečně, spíše u odborné literatury. Díky přímému zápisu hodnot nakladatelem by záznam mohl být úplnější, než nyní poskytují báze CIP¹⁸ a ISN (ohlášené knihy). Metadata budou navíc dostupná pro stažení ještě před vydáním knihy i v síti knihoven (např. pro akviziční účely), počínaje vytvořením záznamu a na jednom místě. Trvalý identifikátor entity zůstane stejný a veřejný, metadata budou ale dále doplňována v průběhu procesu katalogizace. Dokud titul skutečně nevyjde, lze údaje upravovat a reagovat tak například na změny názvu nebo počtu stran.

Pokud by každý takový titul získal i jednoznačný identifikátor (dříve než ISBN a ČČNB¹⁹), bylo by možné vytvářet záznamy odkazující se na tento identifikátor. Pokud zároveň vznikne v Národní knihovně metadatové centrální úložiště, bylo by technicky řešitelné distribuovat všechny změny v záznamech do celého ekosystému. Vznikal by tak jeden záznam, který by systém distribuoval do lokálních bází. Tímto řešením by se dalo zamezit vzniku rozdílných verzí záznamů mezi centrální bází a lokálním záznamem v zapojené knihovně. Zároveň by centrální záznam mohl být kooperativně doplňován a zkvalitňován. Každá změna by se záhy projevila ve všech lokálních kopiích. Navíc by bylo možné určité údaje vkládat pouze pro místní použití, nebo naopak některé informace pro použití v konkrétní knihovně skrývat.

Národní autority

Kvalitní bibliografické záznamy si dnes již neumíme představit bez selekčních údajů opřenejších o autoritní záznamy. Autoritní záznamy jsou významným stavebním kamenem bibliografických databází. Umožňují jednoznačně identifikovat konkrétní instance entit, propojují bibliografické záznamy a hrají významnou roli při vyhledávání v databázích a metadatach. V současnosti se v knihovních bázích zejména využívají autoritní záznamy pro osoby (personální jména), korporace, předmětové termíny/deskriptory, formální deskriptory, geografická jména, názvy děl a vyjádření (formou autoritních záznamů pro anonymní díla nebo formou autoritních záznamů typu jméno-název pro díla uvedená pod jménem autora). Zastoupení autoritních forem výše uvedených typů entit ovšem není zdaleka v bibliografických záznamech stoprocentní.

Jako příklad je možné uvést databázi ČNB²⁰, v níž 866 tis. záznamů obsahuje pole 100 (Hlavní záhlaví – osobní jméno). 797 tis. výskytů obsahuje podpole 7, v němž je uveden identifikátor autoritního záznamu. Ze 731 tis. výskytů pole 700 (Vedlejší záhlaví – osobní jméno) obsahuje podpole 7 s identifikátorem autoritního záznamu 628 tis. výskytů pole (viz tabulka 1).

Databázi ČNB lze pokládat za velmi dobře spravovanou databázi s nejvyšší možnou kvalitou katalogizace. Personální jména jsou nejčastěji a nejběžněji tvořenými autoritními záznamy vůbec. Přesto ani v polích 100 či 700 pro personální jména v ČNB

18 <https://www.nkp.cz/sluzby/sluzby-pro/cip-katalogizace-v-knize/co-je-cip>

19 Číslo České národní bibliografie

20 Data analyzována dle: <http://ddb.qa-catalogue.eu/nkp/?tab=data&lang=en>; aktuální k říjnu 2023.

nejsou všechny tvary jmen opřeny o autoritní záznamy. Tento stav je zcela logický vzhledem k tomu, že ČNB obsahuje různé vrstvy záznamů pocházející z různých období.

Tab. 1 Poměr vyplněných identifikátorů u personálních a korporativních autorit v záznamech ČNB

Pole MARC	Celkový počet výskytů	Přítomnost identifikátoru národní autority (podpole 7)	Podíl identifikátorů
Hlavní záhlaví – osobní jméno (100)	866 295	797 667	92,08 %
Vedlejší záhlaví – osobní jméno (700)	731 573	628 432	85,90 %
Hlavní záhlaví – korporativní jméno (110)	27 360	25 304	92,49 %
Vedlejší záhlaví – korporativní jméno (710)	206 444	191 854	92,93 %

Vedle údajů, které je v současnosti možné propojit na autoritními záznamy, navíc existují v bibliografických záznamech údaje, pro něž by bylo vhodné v zájmu jednoznačnosti vyhledávání a propojení autoritní záznamy či kontrolované slovníky vytvořit, ale kvůli katalogizačním tradicím se tak neděje. Jsou to zejména údaje o nakladatelích (tématu se věnovala např. Drobíková aj., 2016), místech vydání či výroby dokumentu, můžeme sem zařadit i problematiku zápisu času/data nebo časového rozpětí (časové rozpětí se vyskytuje v mnoha polích bibliografických i autoritních záznamů – např. datum vydání pole 264, 008; datum vytvoření záznamu pole 008/00-06, datum aktualizace záznamu 005, data související s věcným obsahem dokumentu uvedené ve věcných polích – 648, 045, u autoritních záznamů jsou to data spojená s jednotlivými osobami, s korporacemi, s časově omezenou existencí správních celků či se vznikem nebo aktualizací díla, případně jeho vyjádření, vytvoření nahrávky apod.).

Kooperace na tvorbě autorit

V případě autoritních záznamů je kooperativní tvorba v ČR využívána především pro autoritní záznamy personálních, korporativních a geografických jmen. Kooperativní tvorba jiných typů autoritních záznamů (věcné termíny, názvové autority) je pro náročnější správu, komplexnost struktury a závislost na terminologických soustavách jednotlivých oborů omezená na kooperaci příslušných oddělení Národní knihovny ČR a clusterových knihoven. V praxi proto často narazíme na situaci, kdy knihovny používají neautoritní tvary jmen a názvů v příslušných polích (názvových, předmětových). Možnost vzájemného propojení záznamů může být z těchto důvodů významně omezena.

Stejně jako v případě kooperace při tvorbě bibliografických záznamů i zde dochází ke zpoždění a jisté asynchronitě tvorby autoritních záznamů. Autoritní záznamy jmen autorů vznikají obvykle až po vydání vesměs tištěného dokumentu, a to zejména knihy. V menší míře je zohledňována článková tvorba, brání v potaz jsou autoři článků v českých periodikách nebo periodikách vycházejících na našem území. Vědecká komunita publikující články v zahraničí nebo autority pro jejich výzkumná pracoviště, na kterých působí, mnohdy ani nespádají do záběru českých národních autorit. Podobná situace

se objevuje běžně i v případě jiných typů dokumentů, jako jsou elektronické učebnice nebo výuková videa.

Nezpracováni nyní zůstávají také autoři (a další původci) děl, která vycházejí pouze elektronicky. V případě přijetí elektronického povinného výtisku vznikne potřeba vytvářet autority i pro dosud nezpracované autory. Pro Národní knihovnu ČR to může znamenat další zatížení. Do budoucna je řešením co největší decentralizace tvorby autorit mezi větší množství spolupracujících subjektů, které budou tvořit společně jednu metadatovou bázi. Tento systém by měl být koncipován tak, aby umožňoval různé úrovně práv pro zápis a návrh úprav. Řešení musí být nastaveno tak, aby i přes účast širší komunity bylo i nadále minimalizováno vytváření duplicitních záznamů²¹. Úplné eliminaci duplicit se pravděpodobně ani v budoucnu nevyhneme, ale bude pokračovat snaha o minimalizaci jejich výskytu, a to jak na procesní úrovni, tak i technickými nástroji a postupy.

Ve světě i v ČR se nezávisle na tvorbě autoritních záznamů rozšiřuje jednoznačná identifikace autorů pomocí jiných typů identifikátorů, např. identifikátory ORCID (*Open Researcher and Contributor ID* = otevřené identifikační číslo pro výzkumníky a přispěvatele)²² pro publikování v oblasti vědy a výzkumu nebo identifikátor ISNI (*International Standard Name Identifier* = Mezinárodní standardní identifikátor jména)²³. O identifikátor ORCID mají možnost odborníci z univerzit a výzkumných pracovišť zažádat ještě před publikační aktivitou. Přidělený identifikátor je pak provází při publikování různých zdrojů (např. studií, učebnic, článků v elektronické podobě), ať už v ČR či zahraničí.

Propojená data jako prostředek pro efektivnější spolupráci knihoven

Jak jsme již zmínili, formát MARC nyní ovlivňuje celý proces tvorby a distribuce metadat. Na časové ose procesu vytvoření jednoho konkrétního záznamu v Souborném katalogu ČR (SKC) můžeme demonstrovat, jak současný model funguje.

Jako vzorový záznam jsme zvolili dílo Šikmý kostel, třetí díl, od Karin Lednické. Titul vyšel 15. 4. 2024²⁴ a jeho záznam²⁵ se v SKC objevil následující den, tedy 16. 4. 2024. Bohužel v něm bylo uvedeno nesprávné číslo ČNB (pole 015). Protože šlo o očekávaný titul, byl zájem o záznam ze strany knihoven značný. Záznam se začal prostřednictvím protokolu Z39.50 šířit do knihoven spolu s uvedenou chybou. Počet stažení nelze nijak spočítat. Jen část knihoven, která záznam využila, je zapojená do kooperace v rámci SKC. Můžeme tedy jen sledovat počty importů z jednotlivých knihoven. Jednotlivé importy můžeme vztáhnout ke konkrétním dnům a časům, ale ty často nejsou shodné s datací vložení do lokálníchází.

Jak jsme již zmínili výše, importy do SKC probíhají většinou v týdenních cyklech, ale v praxi se setkáme s kratší i výrazně delší frekvencí. Chyba byla opravena 24. 4. 2024 pracovníkem SKC. Automatická úprava záznamů v lokálních bázích po změně v SKC je problematická, proto se běžně záznamy aktualizují ručně nebo se neaktualizují vůbec. Chyby tak mohou v bázích knihoven zůstávat. Na vzorku záznamů jsme zjistili, že čísla ČNB vůbec neobsahují. To může být problém v budoucnu, protože se jedná o klíčový identifikátor, který může sehrát důležitou roli při budoucí migraci do formátů propo-

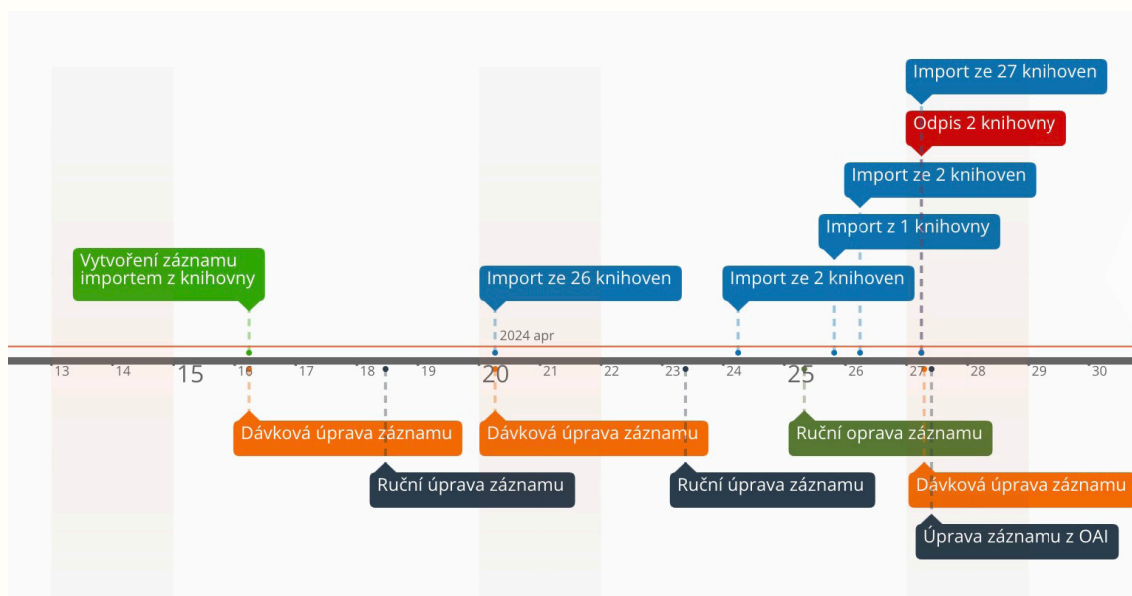
21 Duplicity mohou vznikat například kvůli různým zápisům osobních jmen, pseudonymům atd.

22 Popis identifikátoru viz: <https://info.orcid.org/researchers/>.

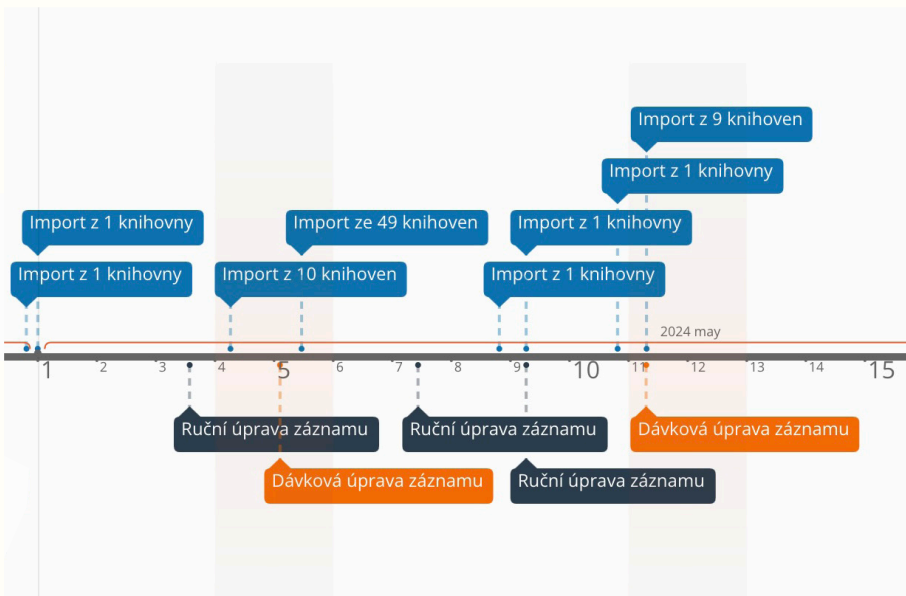
23 ISNI je mezinárodní standard schválený organizací ISO jako ISO 27729 pro identifikaci osob, více viz: <https://isni.org/>.

24 <https://www.knihydobrovsky.cz/kniha/sikmy-kostel-3-645233890>

25 https://aleph.nkp.cz/F/?func=direct&doc_number=010153154&local_base=SKC



Obr. 1 Šikmý kostel, třetí díl (2024): časová osa změn záznamu



jených dat a zároveň může pomoci i při další práci s metadaty. Bez něj je jednoznačná identifikace jen těžko představitelná²⁶.

Součástí přechodu na propojená data by měl být i souhrn strategických rozhodnutí, jež nám pomohou při změně vyřešit většinu problémů, které přináší stávající systém kooperace. Měli bychom se přitom zaměřit na dva klíčové aspekty. Ty jsme identifikovali jako časovou asynchronnost mezi potřebou záznamu a jeho dodáním do centrálního úložiště a neexistenci jednoznačného identifikátoru titulu, který je k dispozici před vydáním. Další klíčovou vlastností nového řešení je vlastnost distribuce veškerých změn z úložiště do zapojených institucí, a to velmi rychle, téměř v reálném čase.

Zásadní místo v zamýšleném systému by měl mít Registr českých knih (ReČeK). Tato databáze by měla vznikat jako společné dílo nakladatelů, knihoven a České národní agentury pro ISBN a ISMN. Bude obsahovat informace o knihách od okamžiku zařazení do edičních plánů nakladatelů. Požadavkem nakladatelů, kterým podmínili spolupráci, je schopnost ReČeK poskytovat strukturované informace o titulech ve formátu ONIX²⁷. Ten se úspěšně mezinárodně používá pro potřeby nakladatelů, distributorů i prodejců knižní produkce. ReČeK by pak měl informace od nakladatelů obohatit o kontextové informace, které jsou důležité pro použití v knihovnách (napojení autorů na personální autority, věcné autority namapované na věcný popis nakladatele, báze nakladatelů atd.). Plánuje se, že při importu záznamu do ReČeK bude okamžitě přiděleno ISBN. „Knihovnickou verzi“ metadat pak musí být systém schopen poskytovat knihovnám, a to minimálně po přechodnou dobu ve formátu MARC, zároveň však ve struktuře vhodné pro tvorbu propojených dat.

Při vytvoření záznamu v databázi ReČeK by měl být v ideálním případě zároveň vytvořen i záznam v centrálním metadatovém úložišti (se zpětnou vazbou do systému ReČeK). Struktura propojených dat umožní lépe popisovat vzájemné vazby entit. Bude nutné přehodnotit způsoby vytváření některých druhů záznamů.

Dle současné instrukce je možné u vícesvazkových monografií vytvářet buď souborný záznam pro všechny svazky (tzv. shora) nebo záznamy jednotlivých dílů (tzv. zdola). Instrukce říká, v kterých případech se má zvolit jedna z uvedených variant. Bohužel kvůli různému výkladu instrukcí se stává, že knihovny vytvářejí pro tytéž tituly záznamy jak shora, tak některé knihovny zdola. Tyto záznamy není vždy jednoduché od sebe na první pohled rozlišit. Jejich struktura může být velmi podobná. Speciálně při hledání přes Z39.50 není v rozhraní nikde jasně uvedeno, jedná-li se o záznam jednoho titulu nebo souboru titulů. Rozpoznání je možné pouze za pomoci některých (v podstatě dílčích) aspektů záznamu – například rok vydání může být u souboru zapsán jako rozsah, fyzický popis bude obsahovat počet svazků místo počtu stran a podobně, u souborných záznamů by měl být na pozici 19 v LDR uveden kód „a“²⁸. Někdy se tak stává, že je do lokální báze stažen souborný záznam s číslem ČČNB, ten je pak následně upraven formou zdola a nedojde k odstranění ČČNB, což způsobuje komplikace při identifikaci dokumentu.

Ve struktuře propojených dat máme výrazně více možností, jak spojovat jednotlivé entity do logických celků, a to včetně hierarchických vazeb. Tyto entity snadno můžeme

26 Jsme si vědomi toho, že tento příklad ukazuje nejen na asynchronnost v komunikaci záznamů mezi knihovnami, ale též ukazuje na různou interpretaci současných platných standardů v ČR ze strany knihoven.

27 Blíže k formátu ONIX viz: <https://www.editeur.org/83/Overview/>.

28 Kvůli častým opravám a přelévání záznamů mezi knihovními bázemi se dle našeho názoru na kódy v LDR nelze úplně spolehnout.

rozlišit tak, že budou mít odlišnou třídu. Vazby se mohou v čase dynamicky měnit. Pro lepší pochopení se uvádíme konkrétní příklad opět s využitím titulu Šikmý kostel.

V bázi ČNB existuje tzv. souborný záznam pro cyklus Šikmý kostel. Ten má přidělen vlastní identifikátor čČNB. Záznam obsahuje také ISBN jednotlivých dílů. Zároveň v souborném katalogu ČR i v lokálních katalozích knihoven mohou existovat samostatné záznamy pro jednotlivé díly. Při popisu pomocí propojených dat můžeme záznamy jednotlivých dílů vzájemně propojit.

Podobně například u záznamů personálních autorit lze spojovat různé identity osob. Opět nemusí vzniknout zastřešující záznam, který bude obsahovat různé odkazové formy jména, pseudonymy nebo jazykové varianty jmen. Vše mohou definovat vzájemné vazby entit. Souborná entita ovšem také existovat může, záleží na použité/navržené ontologii.

Při návrhu formátů a vazeb bychom neměli zůstat v zajetí zažitých přístupů ke způsobu popisu, které za desítky let ovlivnil formát UNIMARC, později MARC 21 a používaná katalogizační pravidla. Měli bychom spíše vycházet z obecnějších principů s cílem vytvořit takové struktury, které nám umožní popisovat realitu přímočaře. Záznamy propojených dat se mohou postupně rozrůstat a získávat na komplexnosti. Na začátku může existovat jednoduchá vertikální struktura, která s rostoucím počtem vazeb může vytvářet složité stromové struktury. Propojená data, na rozdíl od formátu MARC, nám ale umožní na tyto struktury nahlížet z mnoha úhlů a směrů, podle toho, co nás bude zajímat. Tvorba metadat může probíhat decentralizovaně, výsledky však budou dostupné na jednom místě.

Decentralizace tvorby metadat umožní zapojení různých hráčů a významnou roli pak bude mít specializace na určitou oblast. Část práce zastane automatizace v podobě botů²⁹ specializovaných na konkrétní úkony. Například v projektu Wikidata boti doplňují identifikátory do existujících položek z externích zdrojů, případně přidávají vzájemné vazby.

Předpoklady pro přechod na propojená data

Datový model RDF jako základní kámen pro technologie propojených dat rozlišuje (kromě prázdného uzlu) dva základní typy uzlů, které se v trojicích dat mohou vyskytovat na místě objektu – URI (respektive IRI) a tzv. *literály*. Naším cílem by mělo být maximální využití URI odkazujících na objekt. To nás vede k tomu, abychom co nejvíce informací získávali z autoritních bází, ontologií, tezaurů a řízených slovníků. Než tedy začneme tvořit záznamy v propojených datech, musíme mít připravené tyto stavební kameny. Prvním rozhodnutím, které nás čeká, je výběr vhodných zdrojů, které budou základem budoucích metadat.

Jedním z důležitých zdrojů budou autoritní záznamy. Neobejdeme se bez jmenných a věcných autorit, převedených do formátů propojených dat. Dále bychom měli zmapovat vhodné zdroje na národní úrovni, ale nesmíme zapomínat ani na zdroje zahraniční. Nyní například nemáme k dispozici aktuální řízený slovník s údaji o nakladatelích. Ten by v budoucnu mohl poskytovat zmíněný ReČeK. Dalším krokem bude příprava vhodných zdrojů pro použití v praxi. To bude obnášet převody do potřebných formátů a struktur. Vzniklá mapa ontologií by měla být neustále udržována kurátorským týmem, který by měl vyhledávat nové zdroje a dohlížet na relevanci stávajících.

²⁹ Za bota se zde považuje kód, který legálně a cíleně provádí specifické úlohy za účelem zkvalitnění metadat.

Předpokladem je též práce v jiných typech editorů záznamů³⁰. Tvorba propojených metadat bude výrazně jiná než v současných katalogizačních editorech, které běžně používáme pro MARC (katalogizátor bude spíše „propojovat“ než „katalogizovat“). Některá pole budou napojena na rozsáhlé seznamy hodnot, editor bude muset dobře fungovat jako průvodce. Načítání hodnot by nemělo výrazně zpomalovat práci.

Tab 2 – SWOT analýza

Silné stránky	Slabé stránky
Struktura dat umožňující jejich snadnější strojové zpracování	Propojená data jsou uložena ve složitější struktuře než MARC (který je lineární) a proto mohou být složitější na pochopení ze strany tvůrců metadat
Napojení dat bibliografického univerza na okolní svět	Změna systému práce nemusí být přijata širší odbornou komunitou
Využití již existujících dat mimo knihovni svět pro vylepšení dat bibliografického univerza	
Kooperace probíhající ve výrazně větším měřítku než dosud	
Příležitosti	Hrozby
Modernizace správy metadat v knihovnách	Složitost převodu metadat do nového formátu
Vyřešení sdílení metadat v systému knihoven ČR	Náročnost vytvoření centrálního úložiště a celého ekosystému sdílení dat
Poskytování kvalitních metadat pro využití mimo knihovny	Neschopnost prosadit a realizovat přechod na propojená data na národní úrovni
Poskytování nástrojů pro práci s knihovnickými daty pod otevřenou licenci	

Kam směřujeme

Přechod na propojená data není jen náhrada jednoho metadatového typu za jiný. Propojená data přináší značně odlišný přístup při zaznamenávání reality do metadat. Pokud tuto změnu přijmeme, otevřou se knihovnám nové možnosti, především v oblasti kooperativního vytváření metadat. S tímto bychom se neměli spokojit a musíme usilovat také o to, aby nově navržený systém přinesl nástroje, které nám umožní lépe řídit kvalitu metadat a proces jejich tvorby učinit transparentnějším.

Opuštění formátu MARC, který je srozumitelný pouze knihovnám, nám poskytuje jedinečnou možnost, jak ještě více otevřít data pro další využití i mimo náš ekosystém. Nestačí jen vystavit balíky dat pod vhodnou otevřenou licenci. Musíme nabídnout i odpovídající nástroje, které umožní s miliony záznamů efektivně pracovat. Takovými nástroji jsou API (*Application Programming Interface* = programovací rozhraní pro aplikace), která jsou nutným základem při zajištění integrace do informačních systémů či aplikací. Pro analytické potřeby bychom měli nabízet určitou formu dotazovací služby (*query service*), která umožňuje formou dotazů získávat různé druhy strukturovaných

30 V současnosti jsou známy zejména následující editory: Sinopia (<https://sinopia.io/>) vytvořená ve spolupráci následujících institucí: Cornell University, Harvard University, Library of Congress, Stanford University, School of Library and Information Science at the University of Iowa, Program for Cooperative Cataloging (PCC). Dalším editorem je Marva (<https://bibframe.org/marva/editor>), který je vyvíjen pro potřeby Kongresové knihovny. Editor JCricket je nyní testován pro správu veškerých typů entit ve formátu propojených dat v ekosystému mnoha knihoven v rámci konsorcia Share-VDE (<https://www.svde.org/>).

informací. V neposlední řadě musíme odborné i laické veřejnosti nabídnout přívětivé rozhraní pro vyhledávání.

Přechod od formátu MARC bude postupný. Musíme počítat s tím, že po přechodnou dobu budeme vytvářet metadata ve formátu MARC i ve struktuře propojených dat. Máme-li pracovat efektivně, nemůžeme vytvářet všechna data obojím způsobem. Pravděpodobně nejlepším způsobem je vytvoření propojených dat s obsahem specifických polí MARC. Z takových hybridních záznamů lze generovat dostatečně kvalitní záznamy ve formátu MARC bez zbytečné dvojí práce. Takovou cestou jdou některé zahraniční knihovny v současnosti. V návrhu budoucího systému by část pracující s formátem MARC měla existovat jako dočasný modul, který bude možné bezpečně vypnout ve chvíli, kdy již nebude potřebný.

Při designu systému bychom měli mít neustále v patrnosti omezení, která existují v současném řešení. Za ně považujeme především:

- vznik multiplicit v ekosystému metadat,
- různý přístup při popisu různých typů dokumentů,
- centrální úložiště postavené na proprietární technologii, která neumožňuje přístup z různých systémů,
- velmi omezené možnosti distribuce změn v záznamech do lokálních bází,
- omezené možnosti logování činností v repozitáři,
- nemožnost provádět složitější dotazování nad datovými sadami, při výpadku centrálního úložiště nebo distribuční služby nejsou k dispozici potřebná data,
- nemožnost provádět exporty z repozitáře v jiném formátu než je MARC,
- omezené možnosti provádět složitější analytické nebo statistické dotazy nad daty v repozitáři.

Na obrázku níže je návrh schématu jednotlivých modulů systému pro sdílenou tvorbu a distribuci metadat. Jde zatím o obecnější koncept, který vychází ze dvou požadavků: systém bude používat propojená data a funkcionality a pokusí se o řešení stávajících limitů a problémů kooperativní tvorby metadat, a to především na půdorysu Souborného katalogu ČR a jeho současných služeb.

Středobodem řešení je metadatové úložiště. To je zamýšleno jako centralizované s decentralizovanou distribucí metadat do lokálních bází v zapojených knihovnách. Toto úložiště by mělo být nezávislé na knihovním systému, který v budoucnu bude využívat Národní knihovna ČR nebo který bude použit pro správu Souborného katalogu ČR. Stejně tak by mělo být možné s úložištěm spolupracovat z knihovních, ale i jiných systémů. Záznamy v lokálních bázích budou přesnou kopií centrálního záznamu, včetně identifikátorů. Díky tomu, že identifikátory jsou typu URI a vždy spadají do konkrétního jmenného prostoru, je možné toto dodržet – na rozdíl od záznamů ve formátu MARC 21, kde jsou identifikátory zpravidla platné pouze lokálně a vazba na původní záznam se tak může ztratit. Každou změnu v centrálním záznamu bude možné snadno synchronizovat s lokálními databázemi.

Lokální báze zároveň slouží jako redundantní záloha, ze které je možné případně obnovit záznamy v centrálním úložišti. Každé lokální úložiště má možnost připojit k záznamu lokálně používané informace, které se s centrálním záznamem nebudou synchronizovat (budou existovat v jiném jmenném prostoru). Lokální systém a vyhledávací rozhraní umožní skrývat část informací z centrálního záznamu, které v konkrétní knihovně nemají využití, nebo naopak zobrazovat další informace z externích zdrojů podle aktuální potřeby, případně k lokálnímu záznamu připojit informace významné pouze pro konkrétní lokální knihovnu a do centrálního úložiště se nedistribuuji.

V úložišti budou uloženy všechny stavební prvky potřebné pro tvorbu záznamů. Jsou to nejen instance entit, které využijeme k označení autorů nebo k věcnému popisu, ale i hodnoty z různých seznamů nebo slovníků. Ty získáme konverzí stávajících autoritních záznamů, ale bude možné získávat je i z různých externích zdrojů. Takové zdroje bude ovšem nutné nejprve vybrat, zhodnotit jejich kvalitu a dlouhodobou udržitelnost.

K dispozici musí být robustní nástroje, které umožní efektivně spravovat data uložená v repozitáři. Vzhledem k tomu, že řešení bude výrazně omezovat vznik duplicit, bude možné věnovat větší pozornost kvalitě. Správci úložiště tedy budou potřebovat nástroje, které jim umožní aktivně vyhledávat možné chyby a rychle je opravovat. Na této úrovni bychom také měli počítat s nasazením automatizovaných úloh založených na strojovém učení. Každá taková úloha by měla být trénovaná na datech dostupných v repozitáři a její funkčnost ověřena odborníky. Takové úlohy mohou pomáhat odstraňovat chyby při pořizování metadat. Mohou také metadata obohacovat nebo propojovat jednotlivé entity mezi sebou.

Propojením jednotlivých entit jako stavebních prvků budou vytvářeni nástupci dnešních záznamů, a to k identifikaci dokumentů, jejich obsahů, osob, ale i knihoven samotných.

Řešení musí být připravena poskytovat po přechodnou dobu jak propojená data, tak data ve formátu MARC, ovšem bez nutnosti tvořit celé záznamy dvojmo. Toho lze dosáhnout například tak, že po dočasnou dobu bude součástí záznamů ve formátu propojených dat i část specifických polí formátu MARC. Jde o ta pole, která by bylo komplikované generovat z propojených dat v požadované syntaxi. V zahraničních knihovnách je to v současnosti zcela běžný postup. Takto je možné vytvářet záznamy ve formátu propojených dat a zároveň umožnit distribuci metadat ve formátu MARC. K distribuci budou nadále k dispozici tradiční protokoly jako OAI-PMH a Z39.50. Až nastane moment, kdy již záznamy v MARC nebudou třeba, bude možné vypnout celý modul, který distribuci MARC obstarával.

Řešení počítá s úzkým napojením na plánovanou databázi ReČeK. Ta má vznikat za přímé účasti nakladatelů. Těm má umožnit vkládat informace o plánovaných titulech k vydání. Data vzniklá po vydání budou k dispozici ve formátu ONIX pro potřeby knižní distribuce. Zároveň budou data sloužit i agentuře ISBN, která na jejich základě bude přidělovat ISBN jednotlivým titulům. Data budou čerpat i knihovny. Budou mít informace o teprve chystaných titulech např. v předprodeji, z čehož může výrazně těžit právě systém kooperativní tvorby metadat. Každý nově zadaný titul v ReČeK by měl automaticky získat unikátní ID titulu, prostřednictvím kterého jej bude možné identifikovat v celém procesu před i po vydání. V centrálním úložišti se tak bude moci objevit poměrně kvalitní záznam již v momentě, kdy bude titul oficiálně vydán. Pokud si ho knihovna propojí do své lokální báze a později dojde k jeho rozšíření nebo nějaké úpravě, systém tyto změn-

ny propíše do všech knihoven, kde se takový záznam vyskytuje. Změny nikdy nebudou probíhat v lokálním systému.

Toto by mělo být možné díky editoru, který bude zapisovat přímo do centrálního úložiště. Při pokusu vytvořit nový záznam dojde nejprve k ověření, zda již v úložišti záznam pro požadovaný dokument neexistuje. Pokud existovat bude, může ho tvůrce rozšířit nebo ponechat. Pokud záznam k dispozici nebude, dostane tvůrce možnost ho vytvořit. Technicky musí být tento proces nastaven tak, aby pečlivě řídil přístup do úložiště a aktivně bránil vzniku duplicitních záznamů. Součástí řízení přístupu pro zápis by měl být i systém práv. Někteří uživatelé mohou mít vyšší práva (zápis konkrétních „polí“, mazání, přepis hodnot atp.), jiní budou mít práva omezená. Práva mohou být přidělována správcem na základě splnění formálních podmínek, ale i na základě měření kvality zpracování záznamů.

Uživatelé, kteří budou tvořit kvalitní záznamy, pak mohou díky dobré reputaci získat vyšší práva. Naopak chybující uživatelé mohou o svá práva přijít. Systém by měl nabízet také možnost získávat i poskytovat zpětnou vazbu. Uživatelé by měli mít možnost označovat chyby či nepřesnosti. Zásah správce, například při opravě chyby, by mohl vyvolat notifikaci, která by upozornila uživatele, že udělal chybu a jak byla opravena.

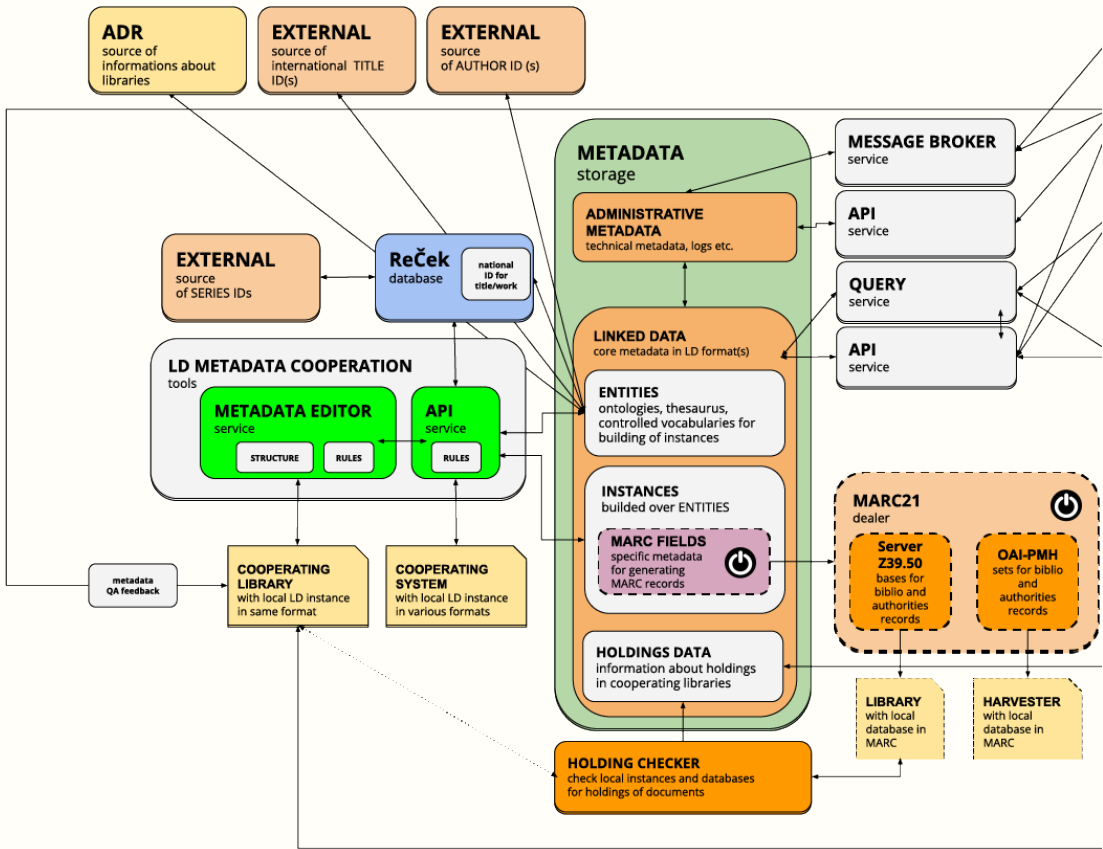
K dispozici by mělo být také API umožňující předávat data z jiných systémů než knihovnických. Jeho prostřednictvím by mělo být možné k existujícím záznamům doplňovat nejen různé identifikátory nebo vytvářet celé záznamy, ale i přidávat rozšiřující informace. Rozhraní API by pak mělo zajistit správné mapování údajů mezi externím systémem a úložištěm. Výstupem ze systému by mělo být několik nástrojů s veřejným i neveřejným režimem zpřístupnění. Příkladem veřejného zpřístupnění může být vyhledávací rozhraní, které umožní komukoli prohledávat záznamy v repozitáři. Můžeme si ho představit jako hledání v současném Souborném katalogu ČR nebo v bázi autorit. Veřejně budou dostupná také API a dotazovací služby. Jejich prostřednictvím by mělo být možné získávat data podle specifických potřeb uživatelů.

Využití funkcionalit veřejných služeb by mělo být možné limitovat licenčně i funkčně. Paralelně k nim budou existovat nástroje pro partnery. I jejich využití bude upraveno smlouvou a poskytované služby budou přizpůsobeny na míru potřebám partnerů. Jako partnery si můžeme představit jiné paměťové instituce, firmy, úřady nebo knihovny.

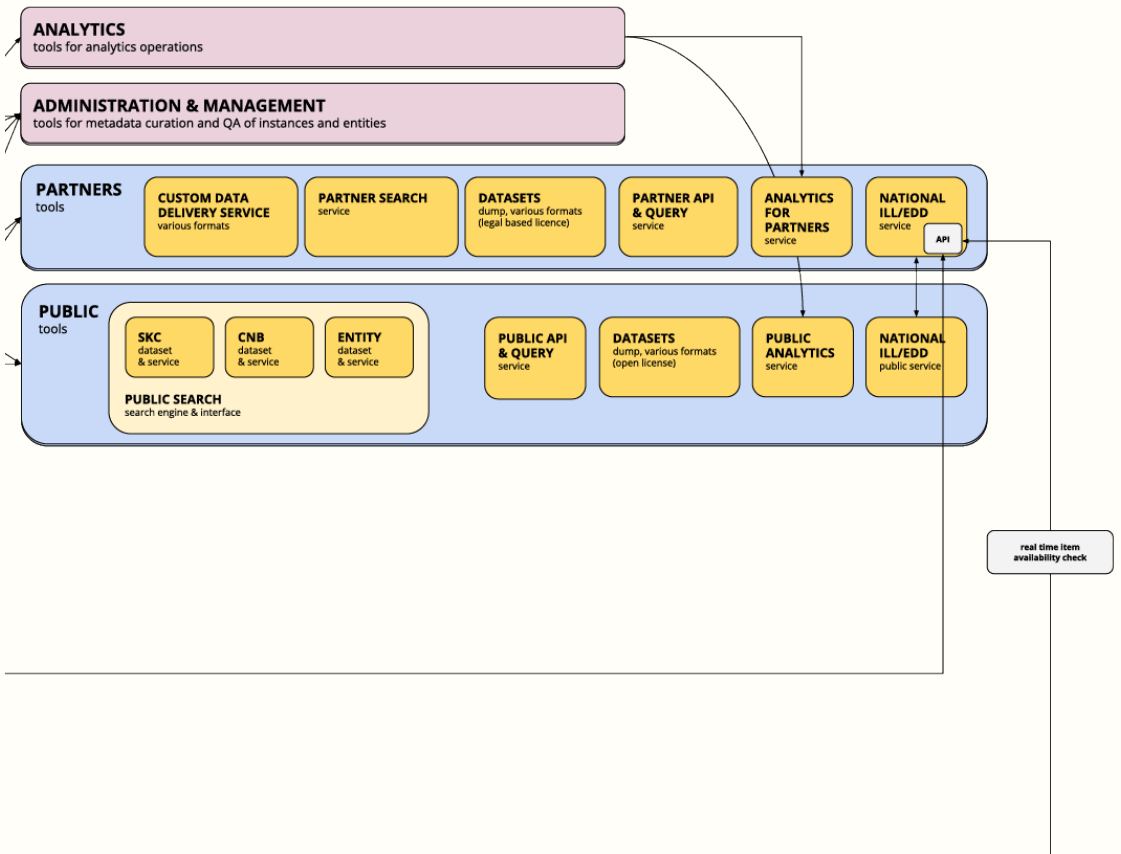
Tak, jak je tomu již nyní, by mělo být možné stahovat i celé balíky dat v různých formátech pod odpovídající licencí a s odpovídající dokumentací. Mělo by být možné upravovat a rozšiřovat funkcionality jednotlivých modulů zamýšleného řešení, a to podle toho, jak porostou potřeby uživatelů.

Modul vyhledávací služby (*query service*) by měl být vytvořen tak, aby umožnil uživatelům tvořit složitější dotazy nad celým úložištěm. Vyhledávání může poskytovat výrazně komplexnější výsledky hledání než standardní vyhledávání nad indexem. Takové služby se dají použít při různých specifických typech hledání, například při doménové analýze, mapování vědy nebo knižní produkce, a to na základě různých aspektů.

Důležitou součástí systému by měl být analytický modul. Ten by měl správcům, ale i uživatelům poskytovat v reálném čase informace o repozitáři. Každý z uživatelů by měl mít k dispozici údaje, díky kterým bude úložiště vnímat jako živý, ale transparentní organismus. Především správci potřebují znát mnoho informací o tom, co se v úložišti děje, jak se chovají uživatelé, a také o tom, jak celý systém funguje.



Obr. 2 Schéma systému sdílené kooperativní tvorby metadat



Vzhledem k tomu, že řešení by mělo poskytovat služby na národní úrovni, půjde o klíčový systém. Jeho vývoj by měl řídit tým v Národní knihovně ČR, kde musí být soustředěny všechny znalosti ohledně jeho fungování, aby bylo možné zajistit dlouhodobou udržitelnost systému, garantovat jeho maximální funkčnost a spolehlivost. Z ekonomických důvodů budou jednotlivé moduly, případně jejich části, sestavené z již existujících řešení. Vlastním vývojem by se pokrylo propojení jednotlivých částí do funkčního celku a případně některé dílčí součásti. Z důvodů udržitelnosti a flexibility by se mělo řešení vyhnout používání proprietárních produktů. Veškerý vývoj by měl být k dispozici veřejně pod vhodnou otevřenou licencí.

Použití způsobů otevřeného vývoje umožní zapojit do procesu větší počet vývojářů a práce tak decentralizovat. Vývojáři se také mohou v čase měnit bez toho, aby taková změna ohrozila další vývoj. Vedení vývoje a jeho směřování musí být v rukou Národní knihovny ČR. Ta může přizvat i další zapojené knihovny. Taková spolupráce by mohla být přínosná nejen při strategických rozhodnutích, ale zpětná vazba od zapojených knihoven by měla přispět k lépe fungujícímu řešení jako celku.

Shrnutí

Přechod na propojená data v knihovnách je největší změnou za poslední tři dekády. Nejde však pouze o změnu metadatového formátu. Změna je komplexní přeměnou celého ekosystému tvorby a využití metadat. Základní stavební kameny pro propojená data mají původ ve filozofii a logice. Struktura propojených dat závisí na použitých ontologiích. Jejich cílem je popis reálného světa (SOWA, 1995). To poskytuje robustní základ pro celý ekosystém propojených dat a je to zcela zásadní posun od jednoúčelového formátu MARC 21.

Schopností jednotlivých entit se navzájem propojovat se setřou dosud patrné hranice mezi typy záznamů. Přestane tak vlastně i existovat bibliografický záznam v pojetí, jak ho známe dnes. Míra podrobností zobrazených uživateli může být různá podle požadavků konkrétní aplikace a potřeb uživatele. Protože v propojených datech popisujeme entity reálného světa, je tedy jedno, zda popisujeme knihu, osobu nebo například instituci (knihovnu). Nad propojenými daty lze vytvořit systémy, které jsou schopné nabídnout hledání nejen nad tituly a jejich autory, ale mohou propojit informace o knihovnách, ve kterých jsou hledané entity dostupné. Je tak možné položit dotaz typu: *Najdi mi knihu „1984“ od autora „George Orwell“ nacházející se v knihovnách s bezbariérovým přístupem, které mají otevřeno v sobotu.* Jestliže se budeme intenzivně věnovat také strojovému zpracování plných textů při klasifikaci obsahu, můžeme nabízet vyhledávací nástroje, které umožní podrobné mapování literatury s ohledem na čas, formu, místa nebo témata. Ta největší změna oproti použití MARC není v tom, kolik informací v záznamech popisuje objekty, ale jak jsou tyto informace zapsané a jak je možné je dále využívat. Samotná struktura MARC umožňuje dobře využívat jednotlivé informace v něm zapsané, ale není možné tyto informace mezi sebou provazovat. Pokud ano, jsou tyto vazby často uloženy v lokálních systémech a v ekosystému se nešíří. V každém z lokálních systémů se pak tyto vazby musí vytvářet znovu.

Abychom se na přechod na ekosystém propojených dat mohli dobře připravit, je důležité již v současnosti otevřít diskusi o otázkách, jakým způsobem popisujeme a identifikujeme entity (instance entit) a zda bychom mohli v této oblasti některé aspekty optimalizovat již dnes, i když stále pracujeme s původními formáty v zažité katalogizační

praxi. Témata, na něž jsme při našich analýzách pro potřeby této studie naráželi, se týkají zejména oblastí:

- způsob zpracování vícesvazkových monografií;
- tzv. popis „shora a zdola“;
- rozlišování dotisků a vydání jednotlivých titulů;
- identifikace seriálů;
- způsob zpracování autorit vs. management identit.

Jejich výčet není zdaleka vyčerpávající. Ovšem problémy identifikované v těchto oblastech jsou poměrně výrazné a již v současnosti komplikují komunikaci záznamů jak uvnitř knihovni sítě, tak i komunikaci dat mezi knihovny a okolními systémy.

Závěr

Cesta k propojeným datům bude náročná. Bude vyžadovat změnu přístupu a myšlení, které po několik generací ovlivňovalo knihovníky i knihovní systémy. V zahraničí se této problematice věnují již několik let. Oproti tomu máme jisté zpoždění. Ovšem výhodou může být možnost využít existujících znalostí a poučit se z chyb, které se průkopníkům logicky nevyhýbají.

Od kolegů v zahraničí jsou k dispozici cenné informace, jak celým procesem projít z hlediska personálního managementu v knihovnách. Bude třeba vysvětlovat na všech úrovních, proč změnu potřebujeme, co přinese a co bude vyžadovat po každém ze zapojených. Zásadním krokem bude včas nastartovat podpůrný proces rekvalifikační pracovníků, kteří tvoří metadata. Například zkušenosti ze Švédska hovoří o klíčové roli celoživotního vzdělávání jako základního pilíře pro úspěšný přechod na propojená data.

Je velmi důležité, aby se témata propojených dat postupně etablovala v odborné diskuzi v knihovnách tak, aby knihovní komunita získala širší povědomí o problematice a mohla se na tuto změnu dobře připravit. Za přínos této práce považujeme návrh postupu pro kooperativní tvorbu metadat, která za pomoci propojených dat řeší většinu slabých míst, kterými trpí stávající systém sdílené katalogizace. Tento návrh vychází z pečlivého zkoumání limitů a nedostatků. Jsme si vědomi toho, že návrh je zatím obecný a nemůže tak nabízet odpovědi na všechny otázky. Vznikl právě proto, abychom otevřeli diskuzi a dále se mohli věnovat jednotlivým částem systému, a také abychom postupně zpřesňovali jeho budoucí podobu. Přesto si myslíme, že změna je v našem prostředí realizovatelná.

Použité zdroje

ALEMU, Getaneh, 2022. *The future of enriched, linked, open and filtered metadata: making sense of IFLA LRM, RDA, linked data and BIBFRAME*. London: Facet Publishing, 2022. ISBN 9781783304943.

BARTL, Zdeněk, 2019. Soubory národních jmenných autorit a propojená data (linked data). In: *Knihovny současnosti 2019*. Online. Praha: Sdružení knihoven České republiky; V Brně: Moravská zemská knihovna, s. 66–70. ISBN 978-80-86249-89-6, 978-80-7051-278-4. Dostupné z: https://sdruc.cz/wp-content/uploads/2020/04/Sbornik_KKS19.pdf. [cit. 2024-05-25].

BERNERS-LEE, Tim, 2006. *Linked Data*. Online. W3, 2006-07-27, last change 2009/06/18 18:24:33. Dostupné z: <https://www.w3.org/DesignIssues/LinkedData.html>. [cit. 2024-06-28].

BIGELOW, Ian and HEATHER, Pretty, 2020. BIBFRAME Readiness: A Canadian Perspective. In: *BIBFRAME in Europe Workshop*. Online. September 22, 2020, Dostupné z: https://www.bfwe.eu/virtual_2020. [cit. 2024-06-28].

- BIGELOW, Ian and SPARLING, Abigail, 2022. BIBFRAME Implementation at UAL: Planning for Success. Online. In: *BIBFRAME Workshop in Europe*. Online. September 20th, 2022, Dostupné z: https://www.bfwe.eu/budapest_2022. [cit. 2024-06-28].
- BIGELOW, Ian and SPARLING, Abigail, 2023. UAL LSP Migration Planning: BIBFRAME Needs and Requirements. In: *BIBFRAME Workshop in Europe*. Online. Brussels, September 20th 2023. Dostupné z: https://www.bfwe.eu/brussels_2023. [cit. 2024-06-28].
- ČESKO. Zákon č. 257/2001 Sb. Zákon o knihovnách a podmínkách provozování veřejných knihovnických a informačních služeb (knihovní zákon). Online. Dostupné z: <https://www.zakonyprolidi.cz/cs/2001-257>. [cit. 01.02.2024].
- DENÁR, Michal a MORAVEC, Josef, 2023. Využití propojených dat v portálu Knihovny.cz. *IT lib*. Online. Speciál 2, s. 26–46. ISSN 1335-793X. Dostupné z: <https://itlib.cvtisr.sk/clanky/vyuziti-propojenych-dat-v-portalu-knihovny-cz/>. [cit. 2024-06-28].
- DROBÍKOVÁ, Barbora, 2013. Standardy pro knihovní katalogy v sémantickém webu. *Knihovna: knihovnická revue*. Online. Roč. 24, č. 2, s. 72–83. ISSN 1801-3252. Dostupný z: <http://knihovna.nkp.cz/knihovna132/13272.htm>. [cit. 2024-05-25].
- DROBÍKOVÁ, Barbora, 2014. RDA a BIBFRAME: budoucí standardy bibliografické kontroly?, 2014. In: *Knihovny současnosti 2014*. Online. Ostrava: Sdružení knihoven. S. 109–118. Dostupné z: <https://ipk.nkp.cz/docs/knihovny-soucasnosti/knihovny-soucasnosti-2014>. [cit. 2024-05-26].
- DROBÍKOVÁ, B.; ODEHNALOVÁ, M.; JURANOVÁ, E.; KRÁLOVÁ, K. a SVATOŠ, L., 2016. FRBR and the publication statement: the problem of identification of relationships and attributes of the entity Manifestation. *Prolnflow*. Online. Roč. 8, č. 1. Dostupné z: <https://doi.org/10.5817/Proln2016-1-2>. [cit. 2024-05-25].
- DUROCHER, Michelle et al. 2020. The PCC ISNI Pilot: Exploring Identity Management on a Global, Collaborative Scale. *Cataloging & Classification Quarterly*. Roč. 58, č. 3–4. DOI: 10.1080/01639374.2020.1713952.
- EDITEUR, 2024. *Mapping from BISAC 2023 to Thema v.1.5*. Online. EDItEUR, akt. 2024-03-21. Dostupné z: <https://www.editeur.org/151/Thema/>. [cit. 2024-06-11].
- EUROPEANA Foundation. *Europeana Data Model (EDM)*. Online. Den Haag: Europeana Foundation. Dostupné z: <https://pro.europeana.eu/page/edm-documentation>. [cit. 2024-06-28].
- FORTIER, A.; PRETTY, H. J. & SCOTT, D. B., 2022. Assessing the Readiness for and Knowledge of BIBFRAME in Canadian Libraries. *Cataloging & Classification Quarterly*. Online. Roč. 60, č. 8, s. 708–735. Dostupné z: <https://doi.org/10.1080/01639374.2022.2119456>. [cit. 2024-05-25].
- Funkční požadavky na bibliografické záznamy*, 2001. Online. Překlad Ludmila Celbová. Praha: Národní knihovna ČR. Dostupné z: <https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frbr/frbr-cs.pdf>. [cit. 2024-06-11].
- GAITANAU, P., ANDREOU, I., SICILIA, M.-A., & GAROUFALLOU, E. 2024. Linked data for libraries: Creating a global knowledge space, a systematic literature review. *Journal of Information Science*. Online. Roč. 50, č. 1, s. 204–244. Dostupné z: <https://doi.org/10.1177/01655515221084645>. [cit. 2024-05-25].
- HEANEY, Michael, 1995. Object-Oriented Cataloging. *Information technology and libraries*. Roč. 14, č. 3, s. 135–153. ISSN 0730-9295.
- Identifikátory.cz: stránky o perzistentních identifikátorech*, 2022. Online. Národní technická knihovna. Dostupné z: <https://identifikatory.cz/cs/>. [cit. 2024-06-13].
- IFLA Study Group on the Functional Requirements for Bibliographic Records, 2021.
- ISO 3297:2022(en) Information and documentation — International standard serial number (ISSN)*. Online. ISO/TC 46/SC 9, Identification and description. Edition 7. Published 2022-06. Dostupné z: <https://www.iso.org/obp/ui/en/#iso:std:iso:3297:ed-7:v1:en>. [cit. 2024-06-13].
- JANSOVÁ, Linda, 2019a. Spolupráce terminologické databáze TDKIV s Wikidaty. In: *Konference Kokon*. Online prezentace. 201-05-16. Dostupné z: https://tdkiv.nkp.cz/docs/kokon_2019.pdf. [cit. 2024-06-13].
- JANSOVÁ, Linda, 2019b. Termíny v TDKIV ve Wikidatech: praktické zkušenosti. In: *Knihovny současnosti 2019*. Praha: Sdružení knihoven České republiky; V Brně: Moravská zemská knihovna. S. 71–74. ISBN 978-80-86249-89-6; 978-80-7051-278-4.

- JANSOVÁ, Linda, 2019c. Výsledky Pilotního projektu zpřístupnění TDKIV v podobě propojených dat. In: *Terminologický seminář*. Online. Národní knihovna ČR, 2019-04-04. Dostupné z: https://tdkiv.nkp.cz/docs/seminar-2019/jansova_pilotni_projekt_tdkiv.pdf. [cit. 2024-06-13].
- JANSOVÁ, Linda, 2020. Workshop Wikidata pro knihovníky. In: *Bulletin SKIP*. Online. Roč. 29, č. 1. ISSN 1213-5828. Dostupné z: <https://bulletinskip.skipcr.cz/vsechna-cisla/prohlizet-cisla/2020-rocnik-29-cislo-1/workshop-wikidata-pro-knihovniky>. [cit. 2024-06-13].
- JESUS, Ananda Fernanda de a CASTRO, Fabiano Ferreira de, 2023. Proposal for the Publication of Linked Open Bibliographic Data. *Cataloging & Classification Quarterly*. Roč. 61, č. 3–4, s. 358–379. DOI: 10.1080/01639374.2023.2234358.
- JONÁČKOVÁ, Lucie a DOSTÁL, Vojtěch. 2020. VYUŽITÍ NOVÝCH TECHNOLOGIÍ: NKlink – nový nástroj propojující Wikidata se světem knihoven. *Čtenář*. Roč. 72, č. 12, s. 433. ISSN 0011-2321.
- KNIHOVNY.cz, 2024a. *Přebírání záznamů přes Z39.50*. Online. Knihovny.cz, 2015–2024. Dostupné z: <https://www.knihovny.cz/Content/z-39-50>. [cit. 18. 5. 2024].
- KNIHOVNY.cz, 2024b. *Zapojené knihovny a zdroje*. Online. Knihovny.cz, 2015–2024. Dostupné z: <https://www.knihovny.cz/Content/zapojene-knihovny-a-zdroje>. [cit. 4. 6. 2024].
- KUČEROVÁ, Helena, 2018. Pojmový model bibliografických informací IFLA LRM. In: *Bulletin SKIP*. Online. Roč. 27, č. 2. ISSN 1213-5828. Dostupné z: <https://bulletinskip.skipcr.cz/vsechna-cisla/prohlizet-cisla/2018-rocnik-27-cislo-2/pojmovy-model-bibliografickych-informaci-ifla>. [cit. 4. 6. 2024].
- KUČEROVÁ, Helena, 2019. Bibliografická metadata v sémantickém webu. *Knihovna: knihovnická revue*. Online. Roč. 30, č. 2, s. 5–35. ISSN 1801-3252. Dostupné z: <https://knihovnav revue.nkp.cz/archiv/2019-2/recenzovane-prispevky/bibliograficka-metadata-v-semantickem-webu>. [cit. 2024-06-13].
- KUČEROVÁ, Helena a BRATKOVÁ, Eva, 2015–2016. *Znalostní báze pro obor organizace informací a znalostí*. Online. Univerzita Karlova v Praze – Filozofická fakulta, ©2015-2016, 20152016. Dostupný z: <http://ko.cuni.cz/>. [cit. 2024-06-13].
- KURFÜRSTOVÁ, Jana; ŽABIČKOVÁ, Petra a CEJPEK, Tomáš, 2023. Deduplikace na portálu Knihovny.cz. *Knihovna: knihovnická revue*. Online. Roč. 34, č. 1, s. 48–62, ISSN 1802-3250. Dostupné z: <https://knihovnav revue.nkp.cz/archiv/2023-1/recenzovane-prispevky/deduplikace-na-portalu>. [cit. 2024-05-25].
- LIBRARY of Congress, 2006. *Group Established to Discuss Bibliographic Control Future*. Online. 2006-12-01. Dostupné z: <https://www.loc.gov/item/prn-06-222/group-established-to-discuss-bibliographic-control-future/2006-12-01/> [cit. 2024-05-18].
- LIBRARY of Congress, 2011. *A Bibliographic Framework for the Digital Age*. Online. 2011-10-31. Dostupné z: <https://www.loc.gov/bibframe/news/framework-103111.html>. [cit. 2024-05-18].
- LIBRARY of Congress, 2022. *MARC 21 XML Schema: Official Web Site*. Online. Library of Congress, 2022-02-02. Dostupné z: <https://www.loc.gov/standards/marcxml/>. [cit. 2024-06-11].
- LIBRARY of Congress, 2023. *MARC standards. MARC Proposal no. 2023-06*. Online. Library of Congress, 2023-10-31. Dostupné z: <https://www.loc.gov/marc/mac/2023/2023-06.html>. [cit. 2024-06-28].
- LIBRARY of Congress, 2024. *MARC 21 Format for Bibliographic Data*. Online. Library of Congress, akt. 2024-06-07. Dostupné z: <https://www.loc.gov/marc/bibliographic/>. [cit. 2024-06-13].
- LIBRARY of Congress, 2008. Working Group on the Future of Bibliographic Control. *On the Record*. Online. January 2008. Dostupné z: <https://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>. [cit. 2024-05-18].
- LICHTENBERGOVÁ, Edita, 2023. *Česká národní bibliografie: základní informace, 2023-02-22*. Online. Dostupné z: <https://www.nkp.cz/o-knihovne/odborne-cinnosti/zpracovani-fondu/informativni-materialy/ceska-narodni-bibliografie-zakladni-informace-prezentace-z-workshopu-wikidat-22-2-2023>. [cit. 2024-04-12].
- LORIMER, Nancy, 2022. Closing loops: moving to production at Stanford. In: *BIBFRAME Workshop in Europe*. Online. September 20th, 2022. Dostupné z: https://www.bfwe.eu/budapest_2022. [cit. 2024-06-28].
- MACEWAN, Andrew, 2022. The International Standard Name Identifier: extending identity management across the global metadata supply chain. *JLIS*, č. 1. DOI: 10.4403/jlis.it-12728.
- McCALLUM, Sally, 2020. Developments at the Library of Congress. In: *BIBFRAME in Europe Workshop*. Online. September 22, 2020. Dostupné z: https://www.bfwe.eu/virtual_2020. [cit. 2024-06-28].

- McCALLUM, Sally, 2022. BIBFRAME Implementation Journey. In: *BIBFRAME Workshop in Europe*. Online. September 20th, 2022. Dostupné z: https://www.bfwe.eu/budapest_2022. [cit. 2024-06-28].
- McCALLUM, Sally; FORD, Kevin; LORIMER, Nancy a WILLIAMSCHEN, Jodi, 2023. MARC to BIBFRAME both ways and the developer's viewpoint. In: *BIBFRAME Workshop in Europe*. Online. Brussels, September 20th, 2023. Dostupné z: https://www.bfwe.eu/brussels_2023. [cit. 4. 6. 2024].
- MAIXNEROVÁ, Lenka. 2023. NOVÉ PROJEKTY: Registr českých knih. *Čtenář*. Online. Roč. 74, č. 3. Dostupné z: <https://svkkl.cz/en/ctenar/clanek/3863>. [cit. 2024-05-25].
- MYNARZ, Jindřich a ZEMÁNEK, Jan. 2010. Úvod k linked data. *Knihovna plus*. Online. Roč. 6, č. 1. ISSN 1801-5948. Dostupné z: <http://knihovna.nkp.cz/knihovnaplus101/myna.htm>. [cit. 2024-04-19].
- NÁRODNÍ knihovna ČR. Propojená data, 2024. In: *IPK: informace pro knihovny*. Online. Praha: Národní knihovna ČR, akt. 2024-02-21. Dostupné z: <https://ipk.nkp.cz/odborne-cinnosti/propojena-data>. [cit. 2024-05-25].
- NÁRODNÍ technická knihovna, 2016–2024. *Polytematický strukturovaný heslář*. Online. Dostupné z: <https://www.techlib.cz/cs/82897-polytematicky-strukturovany-heslar>. [cit. 2024-05-18].
- NATIONAL Library of Sweden, 2019. *National Platform Based On BIBFRAME*. Online. 2019. Dostupné z: <https://www.kb.se/download/18.d0e4d5b16cd18f600eafd/1569324736859/National%20Platform%20Based%20On%20BIBFRAME.pdf>. <https://www.kb.se/download/18.d0e4d5b16cd18f600eafd/1569324736859/National%20Platform%20Based%20On%20BIBFRAME.pdf>. [cit. 2024-05-18].
- OLIVER, Chris, 2021. *Introducing RDA: a guide to the basics after 3R*. Second edition. vyd. Chicago: ALA Editions, 2021. ALA editions special report. ISBN 978-0-8389-4888-0.
- POSSEMATO, Tiziana and CASALINI, Michele, 2020. Linked Open Data in Share-VDE: instructions for use. In: *BIBFRAME in Europe Workshop*. Online. September 22, 2020. Dostupné z: https://www.bfwe.eu/virtual_2020. [cit. 2024-06-28].
- POSSEMATO, Tiziana and LIONETTI, Anna, 2022. BIBFRAME-based SVDE ontology. In: *BIBFRAME Workshop in Europe*. Online. September 20th, 2022. Dostupné z: https://www.bfwe.eu/budapest_2022. [cit. 2024-06-28].
- PROJEKT CENTRAL, 2024a. *Central 2024: Microsoft Power BI*. Online. Dostupné z: <https://app.powerbi.com/view?r=eyJrIjojNjYyZyY1ZTctODg5MS00YjAyLW15NjctNDdhYTRjNTgxNjhlidwidiC16ImY5MWMmZjY2M5LTJkY2EtNGYwYy04MDk2LTUxNTZlZmZjg3NSlslmMiOj09>. [cit. 2024-05-18].
- PROJEKT CENTRAL, 2024b. *Rychlost katalogizace*. Online. Sdílená katalogizace. Dostupné z: <https://sdilenakatalogizace.cz/rychlost-katalogizace/>. [cit. 2024-05-28].
- RESOURCE Description Framework (RDF), 1999. *Model and Syntax Specification: W3C Recommendation 22 February 1999*. Online. Last updated: 2017/10/02 11:00:31. Dostupné z: <https://www.w3.org/TR/1999/REC-rdf-syntax-19990222/>. [cit. 2024-06-28].
- RDA: Resource, Description & Access Toolkit, 2010–2024. Online. American Library Association, Canadian Federation of Library Associations, and CILIP: Chartered Institute of Library and Information Professionals, ©2010-2024. ISSN 2167-3241. Dostupné z: <https://www.rdatoolkit.org/>. [cit. 2024-06-10].
- RIVA, Pat, Le BOEUF, Patrick and ŽUMER, Maja, 2017. *IFLA Library Reference Model: a Conceptual Model for Bibliographic Information*. Online. Consolidation Editorial Group of the IFLA FRBR Review Group. Den Haag: IFLA, as amended and corrected through December. Dostupné z: https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712.pdf. [cit. 2024-05-28].
- RÖSSLEROVÁ, Klára. 2016. Výměnné formáty bibliografických dat: jejich proměna v současnosti. *Knihovna: knihovnická revue*. Online. Roč. 27, č. 1, s. 43–60. ISSN 1801-3252. Dostupné z: <https://knihovnav revue.nkp.cz/archiv/2016-1/recenzovane-prispevky/vymenne-formaty-bibliograficky-dat-jejich-promena-v-soucasnosti>. [cit. 2024-05-28].
- RÖSSLEROVÁ, Klára, 2017a. Budoucnost výměnných formátů bibliografických dat: má MARC budoucnost? In: *Knihovny současnosti 2017*. Online. Praha: Sdružení knihoven ČR. S. 408–416. ISBN 978-80-86249-83-4.
- RÖSSLEROVÁ, Klára, 2017b. Trendy v oblasti bibliografických formátů, aneb, Má MARC budoucnost? *Bulletin SKIP*. Online. Roč. 26, č. 2. ISSN 1213-5828. Dostupné z: <https://bulletinskip.skipcr.cz/vsechna-cisla/prohlizet-cisla/2017-rocnik-26-cislo-2/trendy-v-oblasti-bibliograficky-formatu-aneb>. [cit. 2024-05-28].
- RÖSSLEROVÁ, Klára, 2018. European BIBFRAME Workshop 2017, aneb, První evropský workshop BIBFRAME. *Bulletin SKIP*. Online. Roč. 27, č. 1. ISSN 1213-5828. Dostupné z: <https://bulletinskip.skipcr.cz/vsechna-cisla/prohlizet-cisla/2018-rocnik-27-cislo-1/european-bibframe-workshop-2017-aneb-prvni>. [cit. 2024-05-28].

- SOUBORNÝ katalog ČR, 2023. *Seznam knihoven zasílajících do SK ČR záznamy v elektronické podobě*. Online. CASLIN. Dostupné z: <https://www.caslin.cz/caslin/spoluprace/spolupracujici-knihovny/seznam-knihoven-zasilajících-do-sk-cr-zaznamy-v-elektronicke-podobe>. [cit. 2024-05-25].
- SOWA, J.F., 1995. Top-level ontological categories. *International Journal of Human-Computer Studies*. Roč. 43, č. 5–6, s. 669–685.
- STALBERG, Erin et al., 2020. Exploring Models for Shared Identity Management at a Global Scale: The Work of the PCC Task Group on Identity Management in NACO. *Cataloging & Classification Quarterly*. Roč. 58, č. 3–4. DOI: 10.1080/01639374.2019.1699880.
- SVAZ knihovníků a informačních pracovníků ČR, 2023. *Katalogizace a propojená data*. Online. SKIP ČR. Dostupné z: <https://www.skipcr.cz/knihovnicke-akce/katalogizace-propojena-data>. [cit. 2024-06-18].
- SVOBODOVÁ, Eva, 2003. *Souborný katalog ČR jako informační zdroj*. Online. Národní knihovna ČR, 2003. Dostupné z: <https://www.caslin.cz/caslin/o-nas/prezentace-o-soubornem-katalogu-cr/rok-2003>. [cit. 2024-05-18].
- TENNANT, Roy, 2002. MARC must die. *Library Journal*. Online. Vol. 127, n. 17, p. 26. Dostupné z: <https://www.libraryjournal.com/story/marc-must-die>. [cit. 2024-06-18].
- UNTERSTRASSER, Julia, 2023. *Linked Data and Libraries : How the Switch to Linked Data Has Affected Work Practices at the National Library of Sweden*. Online. Dostupné z: <https://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-506075>. [cit. 01.02.2024].
- ZAPOUNIDOU, Sofia, 2000. *Study of library data models in the Semantic Web environment* Dizertační práce online. 2020. Uloženo v repozitáři Zenodo.org. DOI: 10.5281/ZENODO.4018523. [cit. 2024-05-25].
- ZAPOUNIDOU, Sofia et al., 2024. Entity Management Using RDA and Wikibase: A Case Study at the National Library of Greece. *Journal of Library Metadata*. Roč. 24, č. 2. DOI: 10.1080/19386389.2024.2307208.

DROBÍKOVÁ, Barbora et al. Standardy a technologie propojených otevřených dat: Mohou nám propojená otevřená data a jejich standardy pomoci řešit problémy současné katalogizační praxe v České republice? *Knihovna: knihovnická revue*. 2024, roč. 35, č. 2, s. 5–33. ISSN 1801-3252.