

# Beszámoló a második Európai BIBFRAME Workshopról

Alaposan megtanultuk, hogyan tegyük közzé könyvtáraink tudásvagyonát a szemantikus webben, most már csak az a kérdés, mihez kezdjünk a publikált adatokkal: hogyan szolgáltatassuk, hogyan tegyük azokat saját intézményünk határain túl is használhatóvá? A második Európai BIBFRAME Workshop a jó gyakorlatok és előremutató kezdeményezések széles skáláját vonultatta fel, ugyanakkor a „hogyan tovább” legkínzóbb kérdéseire nem adott válaszokat. A közös munka jövőre Stockholmban folytatódik.

A festői szépségű Firenzében, az Európai Unió Történelmi Levéltárának otthont adó Salviati Villában találkoztak az új bibliográfiai keretrendszer, a BIBFRAME iránt elkötelezett és érdeklődő közgyűteményi szakemberek. A téma aktualitását és egyre növekvő jelentőségét mutatja, hogy a tavalyi évhez képest megduplázódott a résztvevők száma: míg a 2017-ben, Frankfurtban tartott workshopra negyvenegy fő regisztrált, az olaszországi rendezvényre húsz ország intézményeiből csaknem kilencvenen érkeztek, s azon szerencsések, akik mindkét tanácskozáson jelen lehettek, megállapíthatták, hogy öröndetes mértékben növekedett a fiatalabb generáció képviselőinek száma is.

A szakmai program szeptember 17-én, egy egész napos, nyolc előadásból álló tutoriállal vette kezdetét, amelynek célja a BIBFRAME, továbbá az implementációs projektek jelenlegi helyzetének áttekintő bemutatása volt. A bevezető, köszöntő előadást *Mauro Guerrini*, a Firenzei Egyetem oktatója tartotta, aki hangsúlyozta: korszakalkotó változások tanúi vagyunk a könyvtári munkát érintő szabványok és szabályzatok, valamint az integrált könyvtári rendszerek fejlődésének területén. Ez a változás azonban, amely valamikor az RDA tesztelése idején, 2008-ban kezdődött, új együttműködési lehetőséget is kínál a közgyűteményi szféra (az ún. GLAM-szektor) intézményeinek: galériáknak, könyvtáraknak, levéltáraknak és múzeumoknak egyaránt. A transzformáció főszereplője pedig nem más, mint a BIBFRAME, amely a források leírásának többentitáshoz szemléletén, valamint az adatok közötti minősített kapcsolatokon, azaz a linked data-technológián alapuló bibliográfiai adatmodellek egyike. Kidolgozásának fő célja az volt, hogy egy minden szempontból megfelelő, napjaink informatikai trendjeit követő, korszerű formátum jöjjön létre, amely a világháló szerves részévé tudja tenni a közgyűteményi katalógusokat, valamint képes az új feldolgozási szabályzat, az RDA szerint elkészített forrásleírások rögzítésére: ezekkel a MARC sajátos struktúrája miatt már nem tud kielégítően megbirkózni. A BIBFRAME – fűzte tovább a gondolatot az olasz szakember – teljesen új bibliográfiai ökoszisztémát

hozott létre, amelynek elfogadottsága és ismertsége, a nemzetközi együttműködésnek és az egyre szaporodó implementációknak köszönhetően, egyre nő a világban.

Sally McCallum, a Library of Congress munkatársa alapos, áttekintő előadásában megerősítette: a BIBFRAME megalkotásának célja az volt, hogy a MARC-formátumot („a” formátumot – jegyezte meg szellemesen) egy olyan adatsere-keretrendszerrel (*data interchange framework*) váltsák fel, amely a bibliográfiai adatokat könnyebben hozzáférhetővé teszi a világhálón. Emellett követelmény volt vele szemben, hogy terjedjen ki a 20. században megjelent új forrástípusokra (pl. elektronikus és online elérésű forrásokra), foglalja egységes rendszerbe a legkülönbélebb típusú metaadatokat (formai és tárgyi leíró metaadatok, példányadatok, megőrzési, technikai, adminisztratív információk, stb.), valamint szolgálja ki a legkülönbélebb méretű, típusú könyvtárak igényeit is. Technikai szempontból pedig biztosítsa az adatelemek URI-k segítségével történő, minősíthető összekapcsolását, valamint legyen átjárható a MARC-formátum felől, azaz tegye lehetővé a korábban létrehozott bibliográfiai adatvagyon újrafelhasználását.

Nyilvánvaló, hogy ezek az igények az adatokkal összefüggő valamennyi tevékenység (bevitel, csere, tárolás) újragondolását, hovatovább az évtizedes szakmai tudás és szokások megkérdőjelezését jelentik. Az előadó azonban kiemelte, hogy a BIBFRAME használatával gépi következtetés végezhető az információkeresési folyamatban, valamint hogy a linked data-technológia segítségével a közgyűjteményi szektoron kívülre, akár közvetlenül a keresőszolgáltatásokba is eljuttathatóvá válnak az adataink.

Sally McCallum ezt követően a BIBFRAME alapstruktúráját ismertette. Az FRBR, az RDA, a CIDOC CRM, az indecs, valamint a FIAF felhasználásával megalkotott bibliográfiai keretrendszer az FRBR első entitás csoportjának mindössze három elemével dolgozik (mű, megjelenési forma, példány), mivel a BIBFRAME-mű a definíció szerint magába foglalja a kifejezési forma jellemzőit is. A MARC-alapú leírásokkal ellentétben élesen elkülönülnek egymástól a mű és a megjelenési forma adatai (például a közreműködők, tárgyszavak, műfajmegnevezések, vagy épp a célközönség adatai a műhöz kapcsolódnak), továbbá ugyancsak elkülönülnek egyazon mű különféle hordozókon közreadott megjelenési formái. A BIBFRAME szótárelmek elnevezései, ahol lehetséges, az RDA-ban leírtakat tükrözik.

A Kongresszusi Könyvtárban két fázisban tesztelték, tesztelik az új keretrendszert. Az első fázis végrehajtásához – amelyben 40 feldolgozó könyvtáros vett részt, természetesen képzés után – összesen 14 millió bibliográfiai rekordot konvertáltak, valamint egységesített cím-rekordokat is; az ezekből előállított mű-rekordokhoz lehetett kapcsolódni a megjelenési formák leírásával. A második fázis bővebb munkatársi gárdával (60+ fő), és jóval nagyobb adathalmazzal (19 millió rekord és több mint egymillió egységesített cím-rekord, amelyeken a konvertálást követően utófeldolgozást is végeztek) indult el 2017-ben, s ebben már jóval nagyobb szerepet kapott a Library of Congress linked data-szolgáltatása is<sup>1</sup>. A feldolgozó könyvtárosok BIBFRAME-szerkesztőt használva, majd hagyományos módon is leírták a kezükbe kerülő forrásokat.

Sally McCallum a tutorial-napon elhangzott másik előadásában a BIBFRAME-konverzió néhány sajátosságáról és a művelethez kapcsolódó kihívásokról számolt be. A folyamat során tekintettel kell lenni a kiinduló MARC-rekord vegyes mivoltára (azaz arra, hogy több entitásról is tartalmaz információkat), és a különösen technikás mű-konverzióra, ahol a bibliográfiai rekordok címinformációkat közlő valamennyi mezőjét, és

az egységesített-cím-authority rekordokat is fel kell dolgozni, majd az esetleges ismétlődéseket utófeldolgozás során megszüntetni. A konverziót számos szoftver segítségével el lehet végezni; az ún. aktív konverterek, mint pl. a YAZ vagy a Metaproxy, az átalakítás során adatgazdagítást is végeznek az ismert névterekből, míg a Kongresszusi Könyvtár által készített, ingyenesen letölthető szoftver ilyen kiegészítést nem végez.

Az előadásban elhangzott, hogy mivel valószínűleg még sokáig szükség lesz a MARC-ra, ki kell dolgozni a BIBFRAME-ről visszafelé történő konverziós folyamatot is – részben azért, mert egyes automatizált alrendszerek, pl. a kölcsönzés ma még nem állíthatók át teljesen linked data-alapokra. A MARC-ra való visszaalakításnak három fontos problémakörét különítette el az előadó: a művek retrokonverzióját, a nem latin betűs leírások kezelését, valamint az URI-k megőrzését, főleg azokét, amelyek rögzített hosszúságú mezőkből származnak. Felmerült a gondolat, hogy cím-authority helyett bibliográfiai formátumú mű-rekordokat kellene létrehozni, ebben az esetben azonban le kell mondani az utalások készítéséről, az időszaki kiadványok kezeléséhez szükséges mezőkről (640-646 a MARC21-ben), ha pedig az authority-formátumnál maradunk, nem tehetünk minden megjelenési formára érvényes megjegyzéseket (összefoglaló, díjak, stb.), és nem rögzíthetünk tárgyszavakat mű-szinten.

*Philip Evan Schreur*, a Stanford Egyetem könyvtári hálózatának igazgatóhelyettese *Evolúció, forradalom, átalakulás* címmel tartotta előadását, melynek lényege annak illusztrálása volt, hogy egy linked data-alapú architektúrára való áttérés az intézmény valamennyi szolgáltatási ágát jelentősen átalakítja; így például a beszerzést is. Az egyetemi könyvtár számos beszállítóval EDI-kapcsolatban áll a Symphony könyvtári rendszer modulján keresztül – ilyen partnerek pl. a Coutts, a Casalini, a Harrassowitz, vagy a francia AMALIVRE –, és többektől nem csupán a dokumentumot, hanem annak alapszintű bibliográfiai leírását is megkapja. A beszerzés, mivel több lépésből álló folyamat, nagyon összetett rendszerfunkciókat igényel, amelyek közül nem mindent lehet, de nem is mindent kell BIBFRAME-alapokra helyezni – mondta az előadó.

Az egyetemen a 2010-es évek közepe óta futó nagy projekt, a *Linked Data for Production* második fázisa (LD4P2) keretében megindult egy teljesen RDF-alapú feldolgozói környezet, a *Sinopia* kialakítása. A Philip Schreur által bemutatott rendszertervek alapján a katalógizálási munkafolyamat a már felvitt művek közül történő választással kezdődik majd, amelyhez aztán helyi profilokban meghatározott, további adatelemek vihetők fel, miközben létrehozhatók a más entitások felé mutató kapcsolatok is. Az új típusú feldolgozás folyamata és az alkalmazandó metaadatformátumok természetesen más kompetenciákat, képességeket, tudást igényelnek, ezért az előadó hangsúlyozta a képzések, valamint az új munkakörök definiálásának szükségességét.

Philip Schreur a workshop egy későbbi előadásában visszatért az LD4P2-projekt részleteire, amely négy amerikai egyetem, a Cornell, a Harvard, a Stanford, valamint az Iowai Egyetem, továbbá az angol-amerikai közös katalógizálási program, a PCC együttműködésére épül. A fentebb említett feldolgozói környezet fejlesztésén kívül a projekt célja egy szakkönyvtáraktól származó, BIBFRAME-re konvertált adatállomány létrehozása, megfelelő technológia és workflow kidolgozása a MARC-rekordok URI-kkal történő automatizált gazdagítására, a *Blacklight* discovery-felület funkcionalitásának kibővítése linked data-alapú szolgáltatásokkal, továbbá a könyvtári metaadatok erőteljesebb integ-

rációja a világháló szövetébe a Wikidata, valamint a Wikipedian-in-Residence program segítségével.

Az @CULT nevű olasz információtechnológiai cég képviselője, *Tiziana Possemato* izgalmasan részletes prezentációból kiderült, hogyan gazdagítják entitás-azonosítókkal a MARC21-rekordokat, egyszerűen hogyan készítik őket elő a minél pontosabb és teljesebb RDF-konverzióra, később pedig a SHARE-VDE portálon történő megjelenítésre. A bibliográfiai és authorityrekordokból kinyert, személyekre és művekre vonatkozó adatokat relációs adattáblába (az ún. Cluster Knowledge Base-be) töltik, s az azonos entításra utaló, azaz magas hasonlósági pontszámot (*similarity score*) kapott névváltozatok ugyanazt az azonosítót, az ún. klaszterazonosítót kapják. A személynevek esetében először a 100-as és 400-as mező értékeit dolgozzák fel, a bibliográfiai rekordok 245\$c almezőjéből érkező információt már az ezekből nyert csoportokhoz lehet sorolni. A művek esetében a feldolgozás az egységesített címek klaszterezésével kezdődik, melyhez a második lépésben kerülnek a 245\$a-ban tárolt adatok. A csoportosítás végeztével az egyes klaszterek külső névtárból (VIAF, LCNAF, ISNI, Wikidata, stb.) származó URI-kkal gazdagodnak, ezt a munkafázist az ún. Authify-programmodul végzi. Ezután következhet a tripletekbe történő konverzió, amelyet a Lodify modul a BIBFRAME szótárát használva, és a Library of Congress által kidolgozott specifikációt követve végez. Ezt követően az adatokat, mint említettük, a SHARE-VDE portálon teszik közzé: erről a projektről Tiziana Possemato egy másik előadásban árult el további részleteket. A munkában olasz és amerikai partnerek (a Casalini Libri, 16 észak-amerikai könyvtár és természetesen az @CULT) vesznek részt.

A svéd nemzeti könyvtár, a Kungliga Biblioteket volt az első közgyűjtemény a világon, amely tíz évvel ezelőtt a szemantikus weben is közzétette az általa üzemeltetett, és vagy 500 könyvtár által gyarapított svédországi közös katalógusban (LIBRIS) összegyűlt bibliográfiai adatokat. A mostani workshopon *Niklas Lindström* arról számolt be két előadásában, hogy a szolgáltatás új platformja, amely az XL nevet viseli, és amely 2018 nyarán leváltotta a Voyagert, teljes egészében a linked datára épül, adatformátumként pedig a JSON-LD-t használja. Az RDA rendelkezései alapján, entításonként feldolgozott adatokat a közös adatbázisba mentés után MARC-ban exportálni lehet az egyes intézmények lokális rendszereibe (sőt, a feldolgozás közben a felület állandó MARC21-előnézetet is kínál). A katalógizáláshoz egy saját építésű ontológiát (a KBV-t) is használnak, amely jelentős részben támaszkodik a BIBFRAME 2.0-s változatára, és kisebb részben más szótárakra, pl. a MADS-ra vagy az RDA-ra. Ezt egészíti ki egy sajátos MARC-szótár azoknak a relációknak az esetében, amelyeket nem lehetett a fenti szótárak használatával kifejezni. Az adatok konverziójában a MARCframe-nek nevezett MARC-JSON-LD megfeleltetési specifikáció nyújt segítséget.

Igen érdekes előadást hallhattunk *Nancy Lorimertől*, aki a Stanford Egyetemen kifejlesztett CEDAR metaadatszerkesztő felület<sup>2</sup> működését mutatta be. A rövid regisztráció után bárki számára hozzáférhető interfész lehetővé teszi leíró űrlapok (template-ek) készítését a rendszerben megtalálható ontológiákra, így pl. a BIBFRAME-re vagy a bibliotek-o-ra építve, majd az ezek segítségével történő metaadatfelvitelt és az adatok több formátumban történő megosztását, illetve repozitóriumba (például a Stanford Digital Repository-ba (SDR)) történő feltöltését.<sup>3</sup> A linked data-alapokon megvalósuló adatrögzítés részlet-

gazdagabb feltárást, ellenőrzöttebb metaadatokat, a MARC-tól való függetlenedést és hatékonyabb, magasabb szintű együttműködést tesz lehetővé a világhálón keresztül.

*Jodi Williamschen*, a Kongresszusi Könyvtár munkatársa a könyvtár „hivatalos” BIBFRAME-szerkesztőfelületét, valamint annak funkcionalitásbeli újdonságait mutatta be. Az előre definiált, grafikus űrlapok (az ún. profilok) segítségével bárki könnyedén kipróbálhatja, milyen „érzés” az egyes bibliográfiai forrástípusokat – könyvet, kottát, hangfelvételt, periodikát, stb. – az RDA szabályai szerint entitásokra bontva, a BIBFRAME elemkészletét használva feldolgozni. A beviteli mezők – ahogyan azt egy MARC-űrlapon megszokhattuk – szabad szöveget vagy szótárértéket (néha profilspecifikus alapértelmezéssel) tartalmazhatnak, az authority-adatokat a BIBFRAME Editor olyan állományokból emeli be, mint a LCNAF, az LCSH, a LCGFT vagy a LCMPT, más helyeken pedig természetesen az RDA szótáraiban meghatározott értékekkel dolgozhatunk. A jelenlegi verzióban<sup>4</sup> már lehetőségünk van az elkészített leírások mentésére, másolására, sőt a mások által felvitt adatokat is szabadon módosíthatjuk, sőt, tetszés szerint bővíthetjük az űrlapot olyan adatelemekkel, amelyek a BIBFRAME szótárában definiálva vannak.

Bizonyos szempontból kilógott a sorból a nap utolsó előadása, amelyet *Isabel Quintana*, a Harvard Egyetem képviselője tartott, mégpedig az ISNI nemzetközi azonosító használatának lehetőségeiről a számos névtípust tartalmazó NACO nemzetközi névauthority-állományban, azaz, talán ismertebb nevén, a LCNAF-ban. Az ISNI (*International Standard Name Identifier*) az ISBN-hez hasonlóan egy egyedi, állandó számsort alkalmaz azonosításra, ezt a számsort személyekhez, de akár testületi nevekhez is hozzárendelhetjük.<sup>5</sup> Az ISNI adatbázisába – amely a <http://www.isni.org/search> oldalon kipróbálható, de van letölthető kezelőfelülete is, a WinIBW – tömegesen kerülnek be az adatok. Azt, hogy melyik névalaknak kell új ISNI-t kiadni, s melyiket lehet már egy meglévőhöz hozzárendelni, egy algoritmus dönti el, amely rendkívül szigorúan „ítél”, abból az alapelvből kiindulva, hogy könnyebb két, összetartozó nevet egységesíteni egy azonosító alá, mint szétválogatni azokat, amelyek hibásan kerültek össze. A szoftveres betöltést természetesen humán ellenőrzési-egységesítési munka követi. Ezután az ISNI-azonosítót már fel lehet használni pl. MARC-rekordok adatgazdagítására és a rögzített adatok egyértelműsítésére (a MARC21-ben pl. a 024-es mezőben tárolják az ISNI értékét.)

A szeptember 18-án kezdődő workshop bevezető előadását közösen tartotta a Kongresszusi Könyvtár két munkatársa, akik újra összefoglalták a MARC-adathalmaz konverziójával, az eredmények utólagos feldolgozásával, valamint a tesztprojekt második fázisával kapcsolatos legfontosabb tudnivalókat. Új információ volt, hogy a jelen pillanatban „hivatalos kiegészítésként” létező ontológiákat (ArtFrame, PMO, stb.), amelyeket a Stanford Egyetemen fejlesztettek a LD4P projekt részeként, esetleg ténylegesen integrálnák a BIBFRAME-be, de ez még csak lehetőségként merült fel, s nem tudni biztosan, valóban megtörténik-e az egyesítés, vagy a kiegészítések, ahogy eddig, úgy a jövőben is különálló szótárakként funkcionálnak majd. (Bár érdemes megjegyezni, hogy a BIBFRAME Editor hangzó dokumentumokra vonatkozó profiljai már most tartalmaznak olyan elemeket, amelyek a PMO-ból származnak.) A konverzióval kapcsolatban *Jodi Williamschen* elmondta, hogy sok esetben problémás volt az authorityrekordokból létrejött mű-leírásokkal összeolvasztani a bibliográfiai rekordokból kiválogatott mű-információkat: téves összerendelések történtek, illetve adatvesztés lépett fel – ezért ma már kevesebb utólagos egyesítést végeznek a folyamatban, inkább linkekkel kötik össze az egyes entitásokat.

Az Országos Széchényi Könyvtár BIBFRAME-mel kapcsolatos terveiről e sorok írója számolhatott be a nemzetközi közösségnek. A nemzeti könyvtár 2010 óta tesz közzé kisebb-nagyobb dataseteket a szemantikus weben, s legújabb törekvése, hogy a Digitális Képtár, valamint a Magyar Elektronikus Könyvtár folyamatosan bővülő adathalmazát, amelyek szabadon letölthetők a két gyűjtemény weboldaláról, illetve SPARQL-végpontokon keresztül kereshetők, a BIBFRAME használatával újrakonvertálja – ezek jelenleg főként Dublin Core-relációkat tartalmaznak. Az előadás hipotetikus kérdések feltevésével zárult, melyek főleg arra irányultak, vajon mennyi közzétett adatra van szükség, hogy a BIBFRAME valóban ismert és elismert legyen a világhálón, illetve mennyire használhatók az adataink a külső partnerek számára, azaz mennyire teljesítik a Linked Open Usable Data (LOUD) Rob Sanderson által megfogalmazott követelményeit.

Az Alberta Egyetemről érkezett *Ian Bigelow* rövid prezentációjában a kanadai linked data-intézményrendszert vázolta fel. Előadása alapján a legfontosabb szervezetek, amelyek meghatározzák a legfontosabb stratégiai irányvonalakat, illetve a könyvtárak metaadatkezelési munkájában, valamint a linked data technológiájának megismertetésében segítséget nyújtanak, a jobbra egyetemi könyvtárakat tömörítő Canadian Linked Data Initiative (CLDI), valamint a kanadai könyvtáros szövetség, a CFLA metaadatokkal és adatcserével foglalkozó szakbizottságai. Az Alberta Egyetem könyvtára mindemellett aktívan dolgozik az olaszok által koordinált SHARE-VDE-projektben, és a Stanford Egyetem LD4P2 kezdeményezésében, továbbá a *Sinopia* metaadatkezelő rendszert kidolgozó nemzet- és intézményközi együttműködésben (LD4P Cohort).

*Lendvay Miklós*, az Országos Széchényi Könyvtár informatikai igazgatója egy olyan metaadatszerkesztő platform fejlesztéséről számolt be, amely sok tekintetben hasonlít a Nancy Lorimer által korábban bemutatott CEDAR rendszerre. A magyar nemzeti könyvtár külső partnerek bevonásával olyan tudástár-felületet fejleszt, amely több, tematikus gyűjtemény metaadatainak befogadására is képes. A flexibilis, testreszabható leíró űrlapokat (sémákat) alkotó beviteli mezők (attribútumok) egyenként paramétrezhetők, s a nekik megfelelő RDF property-k több szótárból, így a BIBFRAME-ből is választhatók.

*Jackie Shieh* a George Washington Egyetemen végzett, linked datával kapcsolatos munkákat ismertette. Főként a MARC-rekordban tárolt, nyelvre és írásrendszerre vonatkozó metaadatok konverziós sajátosságaira helyezte a hangsúlyt: a BIBFRAME-konverter a 008-as MARC21-mező három karakterpozíciójából veszi a nyelvre, a 880-as mező \$6-os almezőjéből pedig az alkalmazott eredeti írásrendszerre vonatkozó információkat, és ebből a BCP-47 szabályozásnak megfelelő nyelvi kódokat állít elő. Az előadásban szó esett még e nyelvi kódok szerkezeti felépítéséről, valamint egyes schema.org-típusok (*communicate action*, *transfer action*) alkalmazásáról a párhuzamos adatok kezelésében.

*Osma Suominen*, a finn nemzeti könyvtár munkatársa a BIBFRAME és a schema.org szótára közötti konverzió gyakorlati vonatkozásairól, továbbá egy konverziós segédeszköz létrehozásáról és a működtetés során szerzett tapasztalatokról számolt be. A schema.org-ra történő átalakítás két lépésben történik: a SPARQL CONSTRUCT átalakító lekérdezés segítségével előállított adatokat a második fázisban egy újabb lekérdezés futtatásával finomítják (*reconciliation*) – erre részben azért van szükség, mert a Kongresszusi Könyvtár marc2bibframe2 konvertere egyenként dolgozza fel a leíró MARC-rekordokat, s mindegyiknek egy-egy külön művet hoz létre, ami igen redundáns kimenetet eredményez. A konverzió során a két szótár eltérő értelmezése miatt a BIBFRAME-Work-ök



és a BIBFRAME Instance-ek is Creative Work-ké alakulnak, amelyek a workExample – exampleOfWork tulajdonságpár segítségével kapcsolódnak össze.

*Richard Wallis* online elhangzott előadásában a könyvtárak előtt álló három legfontosabb választási lehetőséget ismertette, amikor azok a linked data-technológia implementálása mellett döntenek. A MARC21 ma már lehetőséget biztosít arra, hogy bizonyos meghatározott almezőkben az egyes authority-adatokhoz URI-t is eltároljunk, ez azonban természetesen nem jelent igazi adatösszekapcsolást, és csak ideiglenes, előkészítő lépésként javasolt az alkalmazása. Ott van természetesen a BIBFRAME, amely kellő mélységben és granularitással írja le a bibliográfiai forrásokat, és megfelelő hordozóformátuma lehet az intézmények közötti adatcserének. Wallis saját, humoros – és sajnos félreértett – kifejezésével élve azonban a weben való megtalálhatóság növelésére a BIBFRAME körülbelül annyira használható, mint egy csokoládéból készült teáskanna. Erre a célra a schema.org a Google általi támogatottsága, valamint könnyű html-be ágyazhatósága miatt sokkal alkalmasabb, hátránya azonban, hogy könyvtári tematikájú kiegészítése (az ún. Schema BibEx) ellenére sem eléggé specifikus. A két ontológia együttes alkalmazása azonban megfelelő stratégia lehet. Az átjárhatóság biztosítása érdekében hívták életre a BIBFRAME 2 Schema-projektet (bibframe2schema.org), amelynek célja egy megfelelő mapping kidolgozása, valamint szoftvereszközök fejlesztése a BIBFRAME-adatok schema.org-kifejezésekkel történő gazdagításának, továbbá a BIBFRAME-Schema.org-irányú konverzió megvalósításának céljaira.

*Reinhold Heuvelmann* a MARC-tól a BIBFRAME-ig vezető út lényeges mérföldköveit tekintette át, amely út természetesen az ISBD és az AACR2 világtól indul, s az FRBR kidolgozását, a bibliográfiai kontroll jövőjéről és jelenlegi pozíciójáról való elgondolkodást (*On the Record*), az RDA megszületését, a MARC21 „feltüpirozását” és a könyvtári referenciamodell is érintve tart a szemantikus web ökoszisztémája és az ontológiák használatát, az adatok megosztása felé.

Ezt követte *Xiaoli Li* előadása, amelyben a linked data és a BIBFRAME szerepéről esett szó a PCC szervezetén belül. A PCC (*Program for Cooperative Cataloging*) egy körülbelül 500 aktív tagot tömörítő nemzetközi együttműködés a könyvtári feldolgozás területén; négy alprogramja közül kettőben (BIBCO, CONSER) bibliográfiai rekordokat, a másik kettőben (NACO, SACO) pedig authority-rekordokat hoznak létre. A PCC 2018-2021 között érvényes stratégiájában több linked datával kapcsolatos pontot is találunk: az adat-összekapcsolás technológiájának minél alaposabb és széleskörűbb megismertetése a könyvtári dolgozókkal, a különféle értékszótárak használati arányának növelése a feldolgozási munkában, és a különféle entitások hálózati szintű autorizálása – ez utóbbi cél tükröződött egyébként a PCC 2017-ben indított pilot projektjében is, amely az ISNI-adatbázis adatainak felhasználására és karbantartására irányult. Érdeemes megemlíteni továbbá, hogy 2018 végétől 2020-ig a PCC tizenhét tagintézménye is a Stanfordi Egyetem partnere lesz a Philip Schreur által korábban bemutatott, *Sinopia* nevű, felhőalapú metaadatszerkesztő felület elkészítésében.

A PCC több munkacsoportja foglalkozik a linked data technológiájával és a konverziós folyamat előkészítésével, ezen csoportok tevékenységéről *Jackie Shieh* tartott rövid ismertetőt. Az „URI-k a MARC-ban” munkacsoport javaslatokat tett egyebek mellett a \$0 és \$1 almezők széles körű használatára, ezekbe authority-, illetve ún. real world object (RWO) URI-k helyezhetők el, amelyek magát az entitást azonosítják. Ugyancsak

javasolták, hogy az eredetileg tárgyszavak forrását azonosító \$2-es almező más authority-mezőkben (pl. 1xx-7xx) is jelenjen meg, s az ORCID-hoz vagy a Wikidatához hasonló források jelölésére is alkalmas legyen. Az egyes bibliográfiai rekordokban leírt entitás azonosítójának (pl. az OCLC Work-ID) hordozására definiálták a 758-as mezőt. A „Best Practice” és a BIBFRAME-munkacsoport együttműködve javasolja jó gyakorlatok kidolgozását számos területen, ilyen pl. a géppel feldolgozható, URI-val és nem szöveges értéként megadott adatelemek körének növelése (dátumok megadása típusos literálként, az ISSN és más azonosítók használata linked dataként, amint ez lehetséges lesz), az RDA használatának egyes területei (tartalom típusa, médium típusa és hordozó típusa adatmezők, vagy a gyakoriságra vonatkozó értékszótár), vagy az ISSN-L nevű azonosító alkalmazása, amely egy számsor alá gyűjti a különféle hordozókon megjelent időszakikiadvány-változatokat.

*Michalis Sfakakis* a bibliográfiai kapcsolatok megőrzésének fontosságáról beszélt rövid előadásában, amelyet a bibliográfiai család fogalmának tisztázásával kezdett. Ezen lényegében egy mű különféle kifejezési formáinak összességét értjük: fordításait, kivonatait, átdolgozásait, különféle formanyelvi (mozgóképek, zene, nyomtatott szöveg, stb.) változatait, stb. – a lényeg, hogy mindegyikben azonos szellemi tartalomegység fejlődik ki. A BIBFRAME sajátos entitáskezelésének eredményeként (FRBR-mű+kifejezési forma=BIBFRAME-mű), továbbá a Kongresszusi Könyvtár konverterének rekordonkénti, lineáris feldolgozásmódja miatt minden MARC-rekordból külön Work képződik más-más kifejezési forma-tulajdonságokkal, s elvész a közöttük lévő kapcsolat. A tesztelési munkák, melyek során szakértők által megállapított bibliográfiai családok MARC-rekordjait konvertálták BIBFRAME-re, azt igazolták, hogy a konverziós és az utólagos feldolgozási-egységesítési folyamaton még van finomítanivaló.

*Kristen Wilson*, az Index Data munkatársa a FOLIO (*The Future of Libraries is Open*) nevű open source platformot mutatta be. A FOLIO lényegében nem más, mint egy szerveren futó keretrendszer, egy szoftveres infrastruktúra, amely különféle appok, pluginek futtatására ad lehetőséget, ezek a bővítmények a FOLIO store-ból tölthetők le, s olyan feladatok végrehajtására használhatók, mint pl. a kölcsönzés, a katalógusban való keresés, vagy a pénzügyek intézése. A feldolgozást az ún. Inventory app támogatja, amelyhez export-import funkció társul, és természetesen új elemeket is lehet az adatbázishoz adni; a tervek szerint a FOLIO MARC-, és a jövőben akár BIBFRAME-alapú feldolgozást és triplestore-ban történő adattárolást és -visszakeresést is támogatni fog. A rendszer fejlesztés alatt áll, a korai tesztelők – javarészt egyetemi könyvtárak – 2019 elejétől kezdenek használni az intézményekben.

*Leif Andresen*, a konferencia utolsó előadója kifejtette: a bibliográfiai rekordok másolásán alapuló katalógizálás nem tud mit kezdeni azzal a helyzettel, ha az eredeti rekord új információkkal egészül ki, vagy elkészítője valamit módosít az adatokon – az adott leírást átvevő intézmények nemhogy nem kapják meg ezeket a módosításokat, de többnyire nem is értesülnek róluk. A linked data ökoszisztémájában azonban, állítja az előadó, a külső adatforrások használata még nagyobb arányban történik, mint eddig – még fontosabb tehát, hogy a megjelenési formák leírásakor a kapcsolódó művek adatai is dinamikusak, frissíthetők legyenek. A megoldás Andresen szerint a távoli entitások cache-elése, ennek a technológiának a részletei azonban még elég bizonytalanok.



S legvégül néhány záró gondolat. A BIBFRAME – mint a fenti példák talán bizonyítják – néhány területen, mint pl. az adatok keresőkbe juttatása, nem állja meg a helyét, s ugyancsak gyenge lábakon áll a más típusú közgyűjteményekkel folytatott kommunikációban, a könyvtárak közötti adatcserének, adatmegosztásnak azonban kiváló eszköze lehet. Érdemes odafigyelni az RDA és a BIBFRAME közötti kapcsolatra, és érdemes elgondolkodni a két ökoszisztéma egymáshoz igazításán a még pontosabb, még részletesebb leírások készítésének reményében. A BIBFRAME közösség ebben az ügyben az RDA Steering Committee véleményére, állásfoglalására is számít.

Hubay Miklós

## Jegyzetek

1. id.loc.gov
  2. <https://cedar.metadatacenter.org/>
  3. Érdemes megjegyezni, hogy a rendszert elsődlegesen a biológiai tudományok kísérleti adatainak metaadatolására fejlesztették ki, így alapvetően e területet leíró ontológiákkal találkozhatunk, azonban kisebb számban közgyűjteményi szótárakat is tartalmaz. A fejlesztők CedarOnDemand néven böngésző-kiegészítőt is készítettek, amely szintén a kutatási adatok egységes metaadatolását hivatott elősegíteni, növelve ezzel megtalálhatóságukat, újrafelhasználásukat.
  4. Itt kipróbálható: <http://bibframe.org/bibliomata/bfe/development.html>
  5. Az ISNI-ről bővebben: Dancs Szabolcs: Bemutatkozik az ISNI, a nemzetközi szabványos névazonosító. Tudományos és Műszaki Tájékoztatás, 64(4). 2017. pp. 199–207. Elérhető: <https://tmt.omikk.bme.hu/tmt/article/view/951/1018> (Utolsó megtekintés: 2018. október 6.)
- 



# European BIBFRAME Workshop