

EÖTVÖS LORÁND TUDOMÁNYEGYETEM
BÖLCSÉSZETTUDOMÁNYI KAR

Doktori disszertáció

Hubay Miklós Péter

Szemantikusweb-technológiára épülő könyvtári forrásleírás

A BIBFRAME bibliográfiai keretrendszer használatának lehetőségei

DOI: 10.15476/ELTE.2022.082

IRODALOMTUDOMÁNYI DOKTORI ISKOLA

Dr. Kulcsár-Szabó Zoltán DSc, egyetemi tanár, a Doktori Iskola vezetője

KÖNYVTÁRTUDOMÁNY DOKTORI PROGRAM

Dr. habil. Kiszl Péter, intézetigazgató és tanszékvezető egyetemi docens, a doktori program vezetője

A BIZOTTSÁG TAGJAI ÉS TUDOMÁNYOS FOKOZATUK:

A bizottság elnöke:	Barátné Hajdu Ágnes PhD, egyetemi tanár
Hivatalosan felkért bírálók:	Boda Gáborné Köntös Nelli PhD, egyetemi adjunktus Dudás Anikó PhD, könyvtárvezető
A bizottság további tagjai:	Dippold Péter PhD, igazgató, a bizottság titkára Kiszl Péter PhD, habilitált egyetemi docens
A bizottság póttagjai:	Kerekes Pál PhD, c. egyetemi docens Németh Katalin PhD, egyetemi adjunktus
Témavezető:	Tóvári Judit CSc, ny. főiskolai tanár

Budapest, 2022.

Tartalom

1. Bevezetés	4
1.1. Motiváció.....	4
1.2. Problémafelvetés, kutatási feladat	6
1.3. A disszertáció tézisei	7
1.4. Módszertan	10
1.5. Az értekezés felépítése	12
2. A jelölőnyelvek kora – az SGML	14
2.1. Az SGML jellemzői	15
2.2. Az SGML megjelenése és elterjedése a könyvtári világban	17
2.3. Gyakorlati alkalmazások, jelentős DTD-k könyvtáron kívül	20
2.3.1. TEI.....	21
2.3.2. EAD	23
2.4. Az SGML Magyarországon.....	24
3. Az XML megszületése	26
3.1. Kísérletek a MARC-XML konverzióra.....	29
3.2. A Library of Congress és a MARC/XML.....	31
3.3. Egy elfeledett modell – a XOBIS	33
3.4. Jelentősebb XML-implementációk	35
3.5. Magyarországi alkalmazások.....	38
4. Az entitás fogalmának megjelenése a könyvtártudományban	41
4.1. Az FRBR megszületése.....	41
4.2. Az FRBR-modell felépítése	44
4.3. Megszületik a Könyvtári Referenciamodell.....	51
4.4. Kapcsolat a múzeumi világgal – az FRBRoo.....	56
5. Az RDA forrásleírási szabályzat megjelenése és fejlődése	58
5.1. Az RDA és a MARC.....	61
5.2. Problémák a MARC-kal.....	63
5.3. A MARC-alapú katalógus újraszervezése az entitások mentén – a FRBR-izáció	66
5.4. Az RDA magyarországi szakmai műhelyei	68
5.5. Néhány gondolat az OSZK RDA-projektjéről	71
6. A linked data – adatok szemantikus összekapcsolása	74
6.1. A szemantikus web építőköve – az URI.....	75
6.2. A szemantikus web szabályai és kérdései, az ötcsillagos Open Data csomag.....	76
6.3. A szemantikus web ragasztója – az RDF.....	78
6.4. A tripletek kifejezésének formátumai	81
6.5. A linked data eszköztára.....	86
6.5.1. Szótárkódolási sémák (névterek)	87
6.5.2. Szótárak.....	89
6.5.3. Könyvtártudományi szótárak	92
6.6. Az adatgazdagítás.....	98
6.7. A linked data egyéb segédeszközei	100
7. A linked data és a könyvtár	102
7.1. A linked data-technológia bevezetésének lépései	102
7.1.1. Adatgazdagítás MARC-környezetben.....	103
7.1.2. Alkalmazási profil – retrospektív konverziós segédeszköz.....	105
7.1.3. Az adatok konverziója	106
7.1.4. Az adatok publikációja.....	109
7.2. Komplex publikációs szolgáltatások	111

7.3. A linked data lehetőségei az információkeresésben – adatgazdagítás, entifikáció és szerendipitás	114
7.3.1. Innovative Inspire.....	115
7.3.2. A Qulto szemantikus katalógus	116
7.3.3. Az adatgazdagítás előnyeinek kihasználása	119
7.4. A linked data szerepe a katalogizálás megújulásában	121
8. Keresőoptimalizálás a linked data segítségével	124
9. A Bibliographic Framework (BIBFRAME) – az új MARC?.....	127
9.1. A BIBFRAME első változata	129
9.1.1. Rob Sanderson kritikája.....	131
9.2. A BIBFRAME 2.0.....	134
9.3. A BIBFRAME kiegészítései	137
9.3.1. A Zepheira és a bibfra.me	137
9.3.2. Art and Rare Materials	138
9.3.3. bibliotek-o	142
9.3.4. Performed Music Ontology	144
9.3.5. A bflc, a Library of Congress speciális relációkészlete.....	145
9.4. A BIBFRAME-kompatibilis gyűjteménykezelő rendszer	147
9.5. Átjárhatóság más metaadat-rendszerek felé	148
9.5.1. Az Europeana és az EDM adatmodell.....	148
9.5.2. Interoperabilitás az RDA adatmodelljének irányába.....	151
9.5.3. Átjárás a schema.org irányába	154
9.6. Konverzió a MARC és a BIBFRAME között – segédletek és eszközök.....	154
9.6.1. Összehasonlító segédeszköz	154
9.6.2. Aktív konverzió	155
9.6.3. MARCEdit	156
9.6.4. Konverzió BIBFRAME-ről MARC-ra.....	158
9.7. BIBFRAME-re épülő katalogizáló felületek.....	159
9.7.1. A BIBFRAME Editor	159
9.7.2. A BIBFRAME Profile Editor	163
9.7.3. A Sinopia.....	166
9.7.4. A LIBRIS XL	168
9.7.5. A RERO.....	170
9.7.6. Összegzés	176
10. A BIBFRAME az Országos Széchényi Könyvtár linked data-miniprojektjében.....	178
11. A Petőfi Irodalmi Múzeum bibliográfiai rekordjainak konverziója	181
11.1. A Petőfi Irodalmi Múzeum könyvtára	181
11.2. Célkitűzések és a kísérleti adathalmaz	182
11.3. A konverzió elemzése.....	185
12. Az eredmények összegzése, értékelése	218
12.1. A konverziós esettanulmány eredményeinek táblázatos összefoglalása	218
12.2. Tézisvizsgálat	219
12.3. További kutatási irányok, lehetőségek.....	224
Mellékletek.....	227
Ábrajegyzék	250
Rövidítések jegyzéke	252
Felhasznált irodalom.....	255

1. Bevezetés

*„Mindig úgy képzeltem, hogy a Paradicsom
csak valamiféle könyvtár lehet.”*

Jorge Luis Borges

1.1. Motiváció

A könyvtár, az egyetemes emberi kultúra egyik fő letéteményese, sok évszázad óta szolgálja a művelődést, a tájékozódást, a világ megismerését. A kilométereken keresztül sorjázó könyvek és más információforrások, a raktári rend szigorú, következetes logikája, valamint a méltóságteljes csend tiszteletet, elfogódottságot ébreszt a belépő olvasóban, és az évszázadok hagyománya, mely szinte ott lebeg a polcok között, biztonságot jelent. Számomra legalábbis ezek az emlékek törnek fel abból az időszakból, amikor először találkoztam könyvtárral, a Fővárosi Szabó Ervin Könyvtár egy kerületi fiókjával. Olyan helynek tartottam, amely a tudást, a hitelességet, a bizonyosságot, a kiszámíthatóságot képviseli abban a világban, amelyre egyre kevésbé jellemzők ezek a számomra igen fontos értékek. A rendszerezéshez, illetve a szabályokhoz való erős kötődésem végül a felsőoktatásban is az informatikus könyvtáros szakhoz vezetett.

Hamar világossá vált számomra, hogy a könyvtárak működése évek hosszú munkájával kidolgozott, aprólékos, nemzetközi szinten egységes szabályokra épül, s hogy egy szabványosan épített gyűjteményi katalógus nem csupán helyi szinten kell szolgálja a tájékoztatást, tájékozódást (illetve a Charles Ammi Cutter, a Párizsi Alapelvek, illetve legújabban a Nemzetközi Katalógizálási Alapelvek által előírt/leírt funkciókat), hanem illeszkednie kell a bibliográfiai számbavétel, és a különféle aggregációs törekvések nemzeti és nemzetközi véráramába. Ennek pedig alapvető feltétele az elődök munkájára való tiszteletteljes, szükség esetén azonban kritikus figyelem, illetve a katalógizálási szabályzatok és metaadat-tárolási technológiák, valamint az egyes formátumok egymásba alakítási lehetőségeinek naprakész ismerete. A főiskolai diploma megszerzése utáni első munkahelyemen, a Széchenyi István Egyetem Egyetemi Könyvtárában – bár fő feladatomban a tájékoztatás volt – a sokféle tevékenységen keresztül egyre jobban megismertem a könyvtári munkát, és egyre nagyobb figyelmet fordítottam a katalógusra és a benne található adatok

tárolásának apró részleteire, problémáira, néha bizonyos következetlenségeire. Ebben minden bizonnyal informatikai érdeklődésem is nagy szerepet játszott.

Közel négy évnyi munkatapasztalattal, s a könyvtárinformatikai területek iránti változatlan lelkesedéssel vágtam bele a mesterképzésbe. Ekkoriban – 2014 utolsó hónapjaiban – történt, hogy egy kedves, akkoriban az OSZK-ban dolgozó pályatársammal történt beszélgetésből tudomást szereztem egy nemzetközi, digitális könyvtárakkal foglalkozó mesterképzésről, illetve az ottani, digitális tudásszervezéssel foglalkozó tanegységek részletes tartalmáról. A képzésen használt oktatási segédanyagok áttanulmányozása forradalmi hatással volt rám: úgy tűnt, a könyvtári-könyvtárinformatikai világ talán még soha nem látott változások előtt áll. Szakmánk az utóbbi fél évszázad legjelentősebb átalakítására szánta el magát, amelynek során évtizedek óta alkalmazott katalógusépítési szabályokat, illetve adatformátumokat vizsgál felül, s fő céljának tekinti, hogy az intézmények kapcsolódjanak a világháléhoz, s gyűjteményük annak integráns részévé váljon. E cél elérésével, valamint az információkereső felületek modernizálásával, és azok lehetőségeinek kibővítésével a könyvtár még több használót lesz képes a hiteles adatok és megbízható információforrások használata felé vezetni.

A könyvtár és informatika világa, úgy tűnik, minden eddiginél szorosabb szimbiózisban fog létezni: a közgyűjteményi adattárolási formátumokat általánosabb, webes formátumokra cserélve sokkal eredményesebb kommunikáció valósulhat majd meg mindazokkal, akiknek az összegyűlt adatvagyon újrafelhasználása, kreatív szolgáltatása a céljuk. Első pillanattól kezdve megrészegítettek ezek a kilátások. Magam előtt láttam a könyvtári szolgáltatás jövőjét, egy olyan világot, ahol a katalógusokat, vagy a discovery szolgáltatásokat használók olyan játszi könnyedséggel igazodnak el a közgyűjteményekben tárolt források között, mintha csak egy webes keresővel dolgoznának. Elámított a gondolat, hogy néhány nagyon egyszerű technológia használatával a világháló addig még soha nem tapasztalt mértékben telítődhet értelemmel, s képes lehet minden eddiginél pontosabb válaszokat adni a könyvtárak, múzeumok, levéltárak adatvagyonában kutató érdeklődőknek, s információkeresésüket új színekkel gazdagítani. Hamarosan azt is megtudhattam, milyen eszközök állnak a közgyűjteményi dolgozók rendelkezésére, ha az intézményekben összegyűjtött tudást a keresőszolgáltatásokba szeretnék juttatni.

2017-től kezdve a Petőfi Irodalmi Múzeum, majd az intézmény Digitális Bölcsészeti Központja munkatársaként a korábbiaknál jóval mélyebben van módom közgyűjteményi

adattárolással-adatkezeléssel foglalkozni. Fontosnak és szerencsésnek tartom, hogy egy olyan szakmai műhelyben dolgozhatok, ahol a kollégák érzékenyek a közgyűjteményi terület változásaira, új trendjeire, és nem csupán felismerik őket, hanem reagálnak rájuk, és a napi munkában, továbbá az intézmény online szolgáltatásaiban is igyekeznek megjeleníteni azokat. Az itt szerzett szakmai tapasztalatok elengedhetetlennek bizonyultak jelen értekezés írása közben.

1.2. Problémafelvetés, kutatási feladat

Amikor a közgyűjteményi adatvagyron szemantikus publikálásáról esik szó, még a szakemberek jelentős hányada is értetlenségét fejezi ki: a gyűjteményi adatbázisok minden használó számára korlátozás nélkül hozzáférhetők az online katalógusokon keresztül, mi szükség lehet bármiféle egyéb publikációra? A probléma feloldása abban a tényben rejlik, hogy a szemantikus weben történő adatközzétéssel, azaz a hagyományos adatsereformátumok gráfokká alakításával nem elsődlegesen a végfelhasználókat, hanem az informatikai szakembereket célozzuk meg, akik az adatok újrafelhasználásával különféle szolgáltatásokat – új felhasználói élményt kínáló keresőfelületeket, vizualizációs megoldásokat, aggregációs platformokat stb. – terveznek. Egy bináris formátumban lévő MARC-rekord megfelelő, emberi értelmezésre is alkalmas külalakban történő megjelenítéséhez szoftvereszközök kellene. Ugyancsak valamilyen szoftver közbeiktatása szükséges ahhoz, hogy a korszerűbb XML-formátumban közzétett dokumentumpéldányok a felhasználók számára is érthetően, látványosan jelenhessenek meg. S ezt a logikát folytatva: szoftvereszköz kell ahhoz is, hogy a gráfként hozzáférhető adatok valamilyen módon eljussanak a gyűjtemények használóihoz.

A fentiekből következik egy fontos kérdés: ha az informatikusok kiszolgálása a cél, mi szükség van egy újabb adatformátumra, amikor napjainkban is kiváló, hatékony, multifunkcionális felületek épülnek a Z39.50, SRU, OAI-PMH¹ protokollokon keresztül hozzáférhető, túlnyomórészt XML-ben megfogalmazott adatokra?

Disszertációmban azt kívánom bemutatni, hogyan alakultak át a közgyűjtemények (főként a könyvtárak) metaadat-tárolási megoldásai az informatika fejlődésének jelentősebb lépcsőfokain, illetve milyen változások mentek végbe a bibliográfiai források leírásának elméletében és gyakorlatában. Az informatikai technológiák fejlődési ívét az SGML

¹ A gyakrabban előforduló rövidítések feloldása az értekezés végén, a Rövidítések jegyzékében található.

jelölőnyelvtől az RDF forrásleíró keretrendszer megjelenéséig mutatom be, míg a könyvtártudományi változások bemutatását az FRBR fogalmi modell megszületésénél kezdem, végpontként pedig az entitásokra épülő katalogizálási szabályzat, az RDA megjelenését jelöltem ki. A vizsgálat célja e két terület elkerülhetetlen összeolvadásának ábrázolása, illetve az egymásra találás okainak bemutatása volt.

Kutatásokat folytattam a katalogizálási munkafolyamatot és az információkeresést érintő, jelenleg létező szoftvereszközök, a linked data-technológia különböző implementációinak számbavétele területén is. Kíváncsi voltam rá, hogy ezek a szoftvereszközök hogyan épülnek rá az új típusú adatvagyonra, milyen kiegészítő funkciókkal bírnak, és milyen hozzáadott értéket jelentenek a napjainkban alkalmazott, hagyományos megoldásokhoz képest.

Harmadik kutatási területem egy konverziós projekt tervezése, végrehajtása, illetve kritikai értékelése volt. Azt vizsgáltam, implementálható-e HUNMARC-környezetben egy alapvetően MARC 21-re optimalizált új adatsere-formátum, továbbá feltérképeztem, hogy milyen konverziós nehézségek, problémák merülhetnek fel. Az a kérdés is foglalkoztatott, hogy e problémák akadályai lehetnek-e egy, a HUNMARC-ra épülő, BIBFRAME-gráfot előállító konverziós segédeszköz előállításának. Szerettem volna képet kapni arról, vajon tényleg a BIBFRAME adatmodellje-e a jó választás az adatok szemantikus formátumba történő transzformációjakor, elegendőek-e a feladat megfelelő minőségű teljesítéséhez az új bibliográfiai keretrendszer szótárelemei, esetleg szükség van-e más elemkészletekből vett relációk bevezetésére.

1.3. A disszertáció tézisei

Disszertációm az alábbi kijelentések vizsgálatára épül:

1. tézis: A könyvtáriumban a bibliográfiai univerzum entitásalapú szemléletét legmegfelelőbben a linked data-technológia segítségével ültetheti át a katalogizálási gyakorlatba. E technológia számottevő előnyökkel bír a jelenlegi megoldásokhoz képest.

Napjaink könyvtári gyakorlatában egyre nagyobb problémát okoz, hogy a MARC-ra épülő informatikai eszközeink jelenleg képtelenek annyi rekordtípust és rekordkapcsolati mezőt kezelni, amennyi a bibliográfiai univerzum fogalmi modelljeiben és az entitásokra épülő RDA katalogizálási szabályzatban leírt kapcsolatrendszer megfelelő kezeléséhez

szükséges. E problémát a linked data-technológia használatba vétele oldja, oldhatja fel. Segítségével nem csupán az összekapcsolódó adatok, hanem azok *kapcsolatai* is expliciten, formalizáltan rögzíthetők, így azok az automatizált, számítógépes feldolgozás számára is értelmezhetővé válnak (innen a *szemantikus* jelző). Az adatok összekapcsolásának módszertana az általánosan alkalmazott webes megoldásra, az RDF-re épül, a közgyűjtemények által közzétett adatok felhasználhatósága ilyen módon univerzálissá válik. A linked data technológiája lehetővé teszi, hogy az egyes metaadatsémákat, elemkészleteket kombinálni lehessen egymással, így a jelenleginél sokkal jobban optimalizálhatóvá, sokszínűbbé válnak a forrásleíró űrlapok és a rajtuk szereplő metaadatelemek.

2. tézis: A könyvtárak számára előnyös választás, ha az RDF-re épülő adatszolgáltatást elsősorban a BIBFRAME bibliográfiai keretrendszerre alapozva valósítják meg.

A BIBFRAME egyike a világhálón hozzáférhető számos, könyvtártudományi tárgyú szemantikus elemkészletnek – ilyenek például a Dublin Core, a BIBO, vagy a FaBiO –, technológiai alapjai teljesen megegyeznek azokéval. Adatmodelljének elméleti kidolgozója, az amerikai Library of Congress már a fejlesztés kezdeti szakaszától fogva a MARC utódjaként emlegeti. Emiatt ismertsége és jelentősége meghaladja más szótárakét. A BIBFRAME adatmodellje nem fedi teljesen az RDA-ban, illetve a Könyvtári Referenciamodellben leírtakat, hanem egyszerűsített, három alapvető entitásra (Work, Instance, Item) épülő szerkezetet valósít meg, és jóval kevesebb kapcsolatjelölő relációt definiál, mint például az RDA szótára.

A közgyűjteményekben felgyűlt adatvagyon megfelelő minőségű, hatékony adatszéréjéhez feltétlenül szükséges egy széles körben elterjedt metaadatséma alkalmazása, amely a különféle aggregációs szolgáltatások, például a közös katalógusok üzemeltetésének is alapfeltétele. Ezért mindenképp érdemes megjelölni egy kitüntetett szótárat, amely az egyes intézmények közötti adatátadás gerincét jelenti. Letisztult adatmodellje, valamint a MARC-ból történő egyszerű átalakíthatósága végett a BIBFRAME ideális választás erre a szerepre.

3. tézis: A MARC 21-BIBFRAME irányú konverziós segédeszközök (stíluslapok) kisebb módosításokkal alkalmassá tehetők a HUNMARC-ban tárolt könyvtári adatvagyon megfelelő minőségű transzformációjára is.

Feltételezésem, hogy a MARC 21, illetve a HUNMARC közötti kisebb-nagyobb különbségek – például az autorizált névalak tárolásának módja, vagy az ISBD egyezményes jeleinek eltérő kezelése – nem jelentenek akadályt a HUNMARC-ban rögzített bibliográfiai rekordok konverziója során. E tézis igazolására megvizsgálom, hogy a magyarországi MARC bizonyos sajátosságai, a mezők, almezők, illetve indikátorok MARC 21-től helyenként eltérő értelmezései mennyiben befolyásolják a konverzió kimenetét, történik-e adatvesztés vagy minőségromlás a visszakeresés szempontjából, s amennyiben igen, az orvosolható-e egyszerű stíluslap-módosítások segítségével.

4. tézis: A linked data-technológia bevezetése nem kizárólag a könyvtáros szakemberek feladata. Igazán hatékonyan és eredményesen közgyűjteményi dolgozók és informatikusok együttműködésével valósulhat meg az átállási vagy adatszolgáltatási projekt.

A sikeres linked data-projekthez, annak erőteljes informatikai vonatkozása miatt nem elegendő kizárólag a könyvtárosi, vagy könyvtárinformatikai szaktudás. A konverziós stíluslapok készítése a múltban (is) szakosodott cégek feladata volt, amelyek valamilyen partneri együttműködésben végzik a dolgukat és készítik el a megfelelő informatikai segédeszközöket az intézmények részére. A legszerencsésebb, ha a közgyűjteményi szakemberek rendelkeznek bizonyos informatikai részismeretekkel, például az RDF-re épülő szemantikus elemkészletek és szótárkódolási sémák működési elvére, illetve a más adatformátumokkal történő megfeleltetések lehetőségeire vonatkozóan. A többi részfeladat, a megfelelő API-k vagy SPARQL-végpontok kialakítása, más adatszolgáltatásokhoz való csatlakozás, az adatok újrafelhasználását lehetővé tevő, szerendipitást és adatgazdagítást támogató felületek tényleges megvalósítása más szakmai területek ismeretében lehetséges.

5. tézis: Az RDF-logikát támogató formátumokra épülő információkereső felületek valóban újszerű felhasználói élményt jelentenek.

Az entitások megjelenítése nem az RDF-logika szerializációs szintaxisainak vívmánya: a könyvtártudomány számára az újdonság a szerendipitás, illetve adatgazdagítás hívószavak köré csoportosítható. Előbbi kifejezés – bár XVIII. századi eredetű – az 1850-es évekig nem tett szert komolyabb ismertségre, ekkortól azonban egyre többször használják a véletlenül tett, eredményes és örömteli felfedezések jelölésére (Carr, 2015). A 2016-ban publikált Nemzetközi Katalógizálási Alapelvekben, továbbá az FRBR-ben, valamint a Könyvtári Referenciamodellben is leírt felhasználói feladatok egyike a felderítés, amely célt

bármely közgyűjteményi keresőfelületnek szolgálnia kell. Ebben lényeges szerepe van a szerendipitásnak.

A szemantikus formátum alkalmazásának másik, igen nagy előnye, amelyet az információkereső felületeknek és a fejlettebb discovery eszközöknek² is ki kell használniuk, a különféle forrásokból származó adatok összekapcsolásának, azaz az adatállomány gazdagításának lehetősége. Egy adott entitás-előfordulás felkutatása más adatforrásokban, illetve az ezekre mutató ugrópontok létesítése a felhasználó továbbirányítása céljából kiemelten fontos feladat egy konverziós-publikációs projekt végrehajtása során. Ezzel szolgálható egy újabb, az ICP-ben leírt felhasználói feladat: a navigálásé és kutatásé. A Nemzetközi Katalógizálási Alapelvek szövege kitér arra, hogy ezt a lehetőséget a saját intézményi platformon kívül más katalógusokban, sőt a könyvtári világ határain túl is biztosítani kell (IFLA, 2016).

1.4. Módszertan

Az értekezés – a fent említettek szerint – két úton közelíti meg a lényegi tartalmat adó linked data-technológia ismertetését. Az informatikatörténet és a katalógizálástörténet egyaránt hálás kutatási téma, könnyedén hozzáférhető, jelentős mennyiségű szakirodalommal, amelyek elemzése, egymással való összevetése, összességében az átfogó fejlődési ív megrajzolása alkotja az értekezés körülbelül egyharmadát, amely a szükségszerű összeolvadás bemutatásának korábban vázolt célját szolgálja. Ennek tárgyalása természetesen nem nélkülözheti a gyakorlati vonatkozásokat – az ismert és kevésbé ismert implementációs projektek részletezését – ezért a legfontosabb elméleti ismeretek közlésén túlmenően néhány olyan kezdeményezést is tárgyalok, amelyek reprezentatíván tükrözik az adott technológiában (így például az SGML-ben, illetve az XML-ben) rejlő lehetőségeket. Szót ejtek olyan kísérletekről is, amelyek valamilyen különlegességgel szolgálnak, mint például a magyar fejlesztésű BDML bibliográfiai séma, vagy a kimondottan előremutató gondolati alapelvek mentén kidolgozott, a későbbi korok informatikai lehetőségeit felvillantó XOBIS.

² Discovery eszköznek, discovery szolgáltatásnak az olyan információkereső felületeket nevezik, amelyek technikai megoldásaikkal már túlmutatnak az OPAC-ok által biztosított lehetőségeken. Képesek például több forrásadatbázis (intézményi katalógus, intézményi repozitórium, előfizetett adatbázisok stb.) együttes keresésére, majd az érkező eredmények egységes szerkezetben történő megjelenítésére. A discovery szolgáltatás jellemzője még az egyszerűsített, egypontosá t tett keresőfelület, amely így logikai operátorok használatát sem igényli. A keresőfelület egyszerűsítéséről a 7.3. fejezetben szólok részletesebben.

Bár a linked data-technológiának is bőven létezik már nyomtatott szakirodalma – a teoretikus alapokat, főbb alkotóelemeket bemutató, terjedelmesebb, összefoglaló munkák, valamint esettanulmány-jellegű cikkek, kutatási jelentések –, sok, újabb keletű szolgáltatás, illetve kísérleti projekt még nincsen megfelelően, részleteiben feldolgozva, bizonyos esetekben pedig csak előzetes információk állnak rendelkezésre. Ezért az új technológiák bemutatásánál megváltozik a hagyományos és elektronikus források részaránya, több esetben konferencia-előadások vázlataiban, illetve videoközvetítésekben elhangzott ismeretekre támaszkodtam. A szemantikus webtechnológia keresőoptimalizálási (SEO) területen történő alkalmazásáról szóló részt a Petőfi Irodalmi Múzeumban végzett, tanulságos gyakorlati kísérletek bemutatása gazdagítja.

Az értekezés gyakorlati részének megvalósításában nagyon nagy szerepet játszott az a kísérleti projekt, amelyet az Országos Széchényi Könyvtár, a Petőfi Irodalmi Múzeum, illetve külsős szoftverfejlesztő munkatársakkal közösen hajtottam végre. Ekkor készítettem el – egyelőre deskriptív jelleggel – egy, a nemzeti könyvtár adatsere-formátumához igazított konverziós specifikációt, amely nagyon sok tanulsággal szolgált mind a kiinduló MARC-adatszerkezetet, mind pedig a BIBFRAME bibliográfiai keretrendszer alkalmazhatóságát és lehetőségeit, továbbá belső konzisztenciáit illetően. E dokumentum elkészítésében rendkívül sokat segítettek azok a tapasztalatok, amelyeket a Magyar Nemzeti Múzeum könyvtárának BIBFRAME-gráffá konvertált adathalmaza tanulmányozása közben szereztem.

Az értekezés keretein belül, a 11. fejezetben bemutatott önálló szakmai munka a Petőfi Irodalmi Múzeum által biztosított, megfelelően válogatott rekordhalmaz konverziójának előkészítése, végrehajtása és kritikai elemzése. A transzformációs folyamat megtervezésénél felhasználtam az OSZK miniprojektjében szerzett tapasztalatokat, ezért a munkát egy előzetes átalakítás lefuttatásával kezdtem, hogy kiderüljön, mely pontokon kell korrigálni a konverziós eszközt, a MARCEdit szoftverbe beépített stíluslapokat. Az eltérések a HUNMARC–MARC 21 adatformátumok különbségeiből, valamint a PIM tárgyleírási munkájának házi feldolgozási szabályzatban rögzített sajátosságaiból fakadtak. A nemzeti könyvtár részére végzett munkához képest – sőt az eddigi magyarországi BIBFRAME-konverziós kísérletekhez képest – többletet jelent, hogy minden felmerült problémára megoldási javaslatokat adok, amelyek a kiinduló adatszerkezet bizonyos módosításaira, illetve a konverziós eszközök kiegészítésére, bizonyos kódrészletek átírására vonatkoznak.

A jelölt módosítások végrehajtásával a PIM kiinduló adatállományából megfelelő minőségű, BIBFRAME-re alapuló RDF-gráf készíthető.

1.5. Az értekezés felépítése

Disszertációmhoz, amint korábban említettem, két elméleti-történeti témában végeztem kutatásokat. Az egyik történetiszál az informatika általános fejlődését, továbbá az új vívmányok könyvtártudományi alkalmazásait mutatja be az általános alkalmazású jelölőnyelvek korszakától kezdve: az SGML-t az értekezés *második*, míg az XML-t a *harmadik* fejezete tárgyalja. A folytatásban a könyvtártudomány elméletének fontos mérföldköveit veszem szemügyre: a könyvtártudomány legfontosabb fogalmi modelljeivel, az FRBR-rel és az LRM-mel a *negyedik*, míg az RDA katalogizálási szabályzattal az *ötödik* fejezet foglalkozik.

Az értekezés törzsanyagát része a szemantikus web, azaz az adat-összekapcsolás informatikai megvalósításának bemutatása adja. A fontos, általános építőkövek leírása (*6. fejezet*) után a linked data könyvtári területen való alkalmazására térek át, és igyekszem körképet adni arról, hogy milyen kísérletek, fejlesztési irányok, jó gyakorlatok jellemzők a közgyűjteményi világban, ki hogyan gondolkodott az új technológia alkalmazásáról. E téma tárgyalása a disszertáció *7. fejezetét* alkotja. A szemantikus webtechnológia alkalmazásának egy különleges területéről – a keresőoptimalizálásban történő felhasználhatóságáról – a *8. fejezet* szól.

A könyvtártudomány történetében talán az egyik legizgalmasabb fordulópont 2011-2012-ben következett be, amikor az Egyesült Államok nemzeti könyvtára, a Library of Congress bejelentette, hogy megtalálták, elkészítették a MARC utódját. Informatikai szakemberek segítségével egy olyan, korszerű, szemantikusweb-technológiára épülő bibliográfiai keretrendszert fejlesztettek ki, amely a remények szerint a világon mindenhol felváltja majd a jelenleg használt, egyre korszerűtlenebbé, rugalmatlanabbá váló MARC adatsere-formátumot. Disszertációm *9. fejezetében* ezért kiemelten foglalkozom ezzel a megoldással – a BIBFRAME-mel –, jóllehet, miként korábban már utaltam rá, ez a szótár csupán egy a jövő adathálózatának építőkövei közül. A tervbe vett, igen ambiciózus célkitűzés, nevezetesen az, hogy a világ könyvtárai megszabadulhassanak a cédulakatalógus hagyományait őrző, fél évszázados MARC-tól, igen izgalmas kutatási témává teszi a BIBFRAME keretrendszert.

Az értekezés gyakorlati részének elején röviden ismertetem az OSZK kísérleti publikációs projektjét, amelyben munkatársaimmal a BIBFRAME-szótár segítségével hajtottunk végre sikeres adatközzételt egy kisméretű, kísérleti adathalmazon. E kezdeményezésről az értekezés *10. fejezete* számol be. A *11. fejezetben* egy speciális, múzeumi gyűjteménykezelő rendszerből származó, módosított és kiegészített HUNMARC-ra épülő, célszerűen válogatott adathalmaz BIBFRAME-konverziójának tanulságait ismertetem, amely konverzió végrehajtásakor sok esetben az OSZK-projekt eredményeire is támaszkodhattam. E feladat végrehajtása egyszerre igényli a teszhalmaz MARC-struktúrájának, az egyes mezőkben és almezőkben található adatelemeknek a pontos ismeretét, továbbá a BIBFRAME keretrendszerben leírt relációk alapos feltérképezését. Ezen felül precizitást és kreativitást tesz szükségessé, mivel a Library of Congress által kidolgozott konverziós specifikációk a MARC 21 adatcsere-formátumra épülnek, amelytől a HUNMARC sok ponton eltér, ezért nem lehetséges – vagy nem nyújt kielégítő eredményt – az egyszerű, módosítások nélküli implementáció. Az eredmények és javaslatok sorát értékelés, illetve összegzés és kitekintés zárja (*12. fejezet*).

2. A jelölőnyelvek kora – az SGML

Az Internet megszületésével nagyjából egy időben az IBM-nél is korszakalkotó változás volt kibontakozóban. A már akkor is körülbelül fél évszázados múlttal rendelkező informatikai cég jelentős tudásvagyonot halmozott fel, amelynek gondozásában – és kereshetőségének biztosításában – jelentős szerepet játszott Steve Furth, az IBM akkori marketingmenedzsere. Tevékenysége annyira fontos és megkerülhetetlen volt, hogy a vállalaton belül sokan csak az információkeresés atyjának nevezték az idős szakembert. Steve Furth egy alkalommal beszélgetésbe merült egy kollégájával, Charles F. Goldfarbbal, amelynek során szó esett a vállalat automatizált dokumentum-feldolgozási és adatrepresentációs mechanizmusairól és azok visszasságairól. Goldfarb és kollégái a cégnél alkalmazott szövegfeldolgozó algoritmusok egységesítése során észrevették, hogy azok mindegyike teljesen más, szövegbe ágyazott parancsokat igényel a dokumentumok végleges változatainak elkészítéséhez. A kompatibilitási problémák miatt felmerült, hogy a szövegelemek feldolgozására vonatkozó utasítások rendszerét teljesen el kell vetni. Furth azonban úgy gondolta, az alapötlet jó, a jelölésrendszert azonban más funkciók betöltésére kellene használni. Goldfarb ekkor tette azt a javaslatot, ami végül a forradalmi újításhoz vezetett: az utasítások rögzítése helyett magukat az elemeket címkézzék fel megfelelő, egyértelmű módon (Goldfarb, 1996). Így a feldolgozó szoftverek között nem lép fel inkompatibilitás, mindegyik a maga módján dolgozhatja fel az előre definiált szövegegységeket, például a címeket vagy a képaláírásokat.

Charles Goldfarb az 1969-es év folyamán két kollégájával, Ed Mosherrel és Ray Lorie-val együttműködve megalkotott egy jelölőnyelvet, amelynek legmesszebbre mutató jelentősége az volt, hogy szakított a *procedurális*, utasításközpontú megközelítésmóddal – és a nyomdaipari hagyományokkal. Az IBM-nél kifejlesztett rendszer a szövegek strukturális egységeinek leírását állította a dokumentum-feldolgozás középpontjába, amely az ún. *deskriptív* megoldás néven ismeretes. Ebből több előny is származott. Először is az új típusú jelölőnyelvvvel preparált dokumentumokban történő keresés lehetőségei jelentősen kibővültek, hiszen lehetőség volt például annak beállítására, hogy egy bizonyos szövegrészlet csak fejezetcímbe előfordulva jelentsen találatot. Másodszor: a procedurális megközelítés nyomtató-centrikusságával ellentétben a deskriptív módszerrel készült szövegek képernyős megjelenítése is könnyebben szabályozhatóvá vált (Pitti, 1995a).

Az eredetileg Text Description Language-nek (szövegleíró nyelv) hívott jelölésrendszer alkalmazási perspektívája – nem kis részben a felvázolt lehetőségek miatt –

gyorsan nőtt a vállalaton belül, s ennek eredményeként a fejlesztéséről szóló jelentést már Generalized Markup Language (általános jelölőnyelv) megnevezéssel publikálták 1973-ban (Goldfarb, 1996), s érdekes egybeesés, hogy ennek rövidítése egyben a három kidolgozóra is utal (Goldfarb, Mosher, Lorie). A későbbi években a rendszer jelentősége, elsősorban általános alkalmazhatósága miatt, tovább növekedett, végül túllépett az IBM keretein, és az ANSI, valamint később az ISO közreműködésével amerikai, végül nemzetközi szabvány – vállalati tulajdon helyett köztulajdon – lett belőle. Így született meg 1986-ban az SGML (*Standard Generalized Markup Language, szabványos általános jelölőnyelv, ISO 8879:1986*) (Barry, é.n.), amelynek használatával sokkal könnyebben karbantartható és „hordozható” (ti. a különféle szoftverkörnyezetek között) dokumentumok állíthatók elő (Cole & Kazmer, 1995).

2.1. Az SGML jellemzői

A Goldfarb és két kollégája által kifejlesztett jelölőnyelv a dokumentumok logikai komponenseinek azonosítását szolgálja. A procedurális jelöléssel szemben a deskriptív megoldás „csupán” meghatározza a logikai szövegegységek típusát, feldolgozásuk, megjelenítésük szabályairól azonban nem mond semmit. Két fajtája ismeretes: a *strukturális* típus a dokumentumkomponenseket és azok logikai kapcsolatait jelöli, míg a *nominális* típus a szövegben előforduló személy-, testületi és földrajzi neveket, valamint tárgyszavakat azonosítja, aminek segítségével gyorsan és praktikusán készíthetők különféle mutatók (Pitti, 1995a).

Daniel V. Pitti szerint az SGML nem egy „polcra levehető” jelölőnyelv, amit bármikor, bármilyen dokumentumtípusra lehet alkalmazni, inkább a jelölőnyelvek metasabványa, lényegében keretrendszer azok megalkotásához. Grammatikát biztosít a dokumentumok logikai struktúrájának definiálásához, konvenciókat nyújt a szövegkomponensek elnevezéséhez, és nem utolsósorban: hivatkozások alkalmazásával képes hozzáférést biztosítani más, külső információforrásokhoz. Elterjedése esetén (az írás 1995-ben jelent meg) az SGML integrált és átjárható információs környezetet ígér, amelynek használatával lehetőség nyílik arra, hogy az Internet koherens, szabványalapú egészévé, információs univerzummá váljon.

Corthouts és Philips (1996) megfogalmazásában az SGML legfontosabb előnye, hogy segítségével a szövegstruktúra alkalmazásoktól és eszközöktől független, explicit és kétértelműségtől mentes módon jeleníthető meg. Cole és Kazmer (1995) az SGML-t széles

körben alkalmazható, komplex, de erőteljes protokollként definiálja, amely fontos szerepet játszik a hatékony információcserében.

Az SGML segítségével a nyers szövegbe, amely az elsődleges információt tartalmazza, s amelyet a szakterminológia *dokumentumpéldá*nyának nevez (Cole & Kazmer, 1995), jelölőket, ún. *tageket* kell elhelyezni, amelyek párosak: egyértelműen jelzik a jelölt szövegrész elejét és végét. A tagek jelentését és működési szabályait előre meg kell határozni, dönteni kell például attribútumaikról, egymásba ágyazhatóságukról és kardinalitásukról. Ezeket a deklarációkat, melyek az SGML követelményeivel összhangban készülnek, külön állományok tartalmazzák, amelyeket DTD-nek neveznek (Pitti, 1995a). Egy DTD tehát nem más, mint egy bizonyos dokumentumtípusra vonatkoztatott, specifikus jelölések listája, voltaképpen *ez maga a jelölőnyelv*. A DTD deklarációs funkciói mellett ellenőrző szerepet is betölt: segítségével megállapítható, hogy az előállított dokumentumpéldány érvényes-e, vagyis a benne szereplő jelölések, tagek megfelelnek-e a DTD-ben megfogalmazott szintaktikai szabályrendszernek (Gaynor, 1996). Az SGML-dokumentumpéldányok szerkesztésére több célszoftver létezik³, amelyek automatikusan összevetik a bevitt adatokat a DTD-vel. Így már gépelés közben képesek jelezni az esetleges hibákat, de utólagos validálásra alkalmas szolgáltatás is használható erre a célra.

Bíró Szabolcs *Szövegfeldolgozás XML alapokon* (2005) című könyvének történeti előzményeket tárgyaló fejezetében felsorakoztatja az SGML-lel szemben támasztott követelményeket, és kijelenti, hogy a rendszer ezek mindegyikének meg is felel:

- általánosan elfogadott, dokumentált, gyártófüggetlen szabvány;
- referenciái adott esetben átültethetők más projektekbe;
- az információ szerkezete tervezhető, ember és gép számára egyaránt leírható;
- a felhasználás számára lényeges elemek megjelölhetők;
- számítógépes feldolgozása könnyű, gyors és platformfüggetlen;
- újrafelhasználható, konvertálható, hordozható, továbbá gyorsan és pontosan megtalálható információt tartalmaz;
- biztos, használhatóságot alátámasztó múlttal és „kiszámítható” jövőképpel rendelkezik.

³ E szoftverek ma már természetesen az XML szabályaira épülnek.

2.2. Az SGML megjelenése és elterjedése a könyvtári világban

Az új jelölőnyelv-szabvány természetesen nem kerülte el az információk feldolgozásával, rendszerezésével és tárolásával hivatásszerűen foglalkozó könyvtárosok figyelmét sem. A World Wide Web megszületése utáni években nyilvánvalóvá vált, hogy a bibliográfiai leírást meghatározó, a '60-as évek végén, illetve a '70-es évek elején kidolgozott szabványok és szabályzatok nehezen bírnak lépést tartani az egyre sokrétűbbé, összetettebbé váló információs világ bibliográfiai számbavételi igényeivel (Davis, 1996). Az informatikai eszközök használatának következtében exponenciális mértékben megugrott a források mennyisége is. A munkaterhek sebes növekedése egyszerre vetette fel a bibliográfiai számbavétel jövőbeni biztosíthatóságának kérdését, ugyanakkor maga után vonta a feldolgozási és visszakeresési módszerek megújításának, fejlesztésének szükségességét is. Az SGML-re különös igény mutatkozott a könyvtári világban, mert

- alkalmas volt a MARC-ban felgyülemlett adatvagyron fogadására és további megőrzésére;
- rugalmasabb környezetet biztosított az új típusú dokumentumok feldolgozási gyakorlatának kikísérletezéséhez, miközben tökéletesen alkalmas volt a nyomtatott formátumok kezelésére is (McCallum, 1996);
- új keresési módszereket tett lehetővé;
- használatával a könyvtári adatok részeivé válhattak összetettebb metaadatrendszereknek.

Kézenfekvőnek látszott, hogy a továbblépéshez a MARC szabályait kell SGML-környezetben modellezni: ebből a célból többen is kísérletet tettek egy MARC DTD összeállítására. Legelőször Michel Vulpe munkáját érdemes megemlíteni, aki 1990-ben, az SGML-re specializálódott, Softquad nevű fejlesztőcég képviselőjeként készített egy kísérleti dokumentumtípus-definíciót (McCallum, 1996). Négy évvel később a Berkeley-n dolgozó Jerome McDonough az egyetemen működő *Cheshire II* elektronikus információkereső felület céljaira állított elő egy MARC DTD-t, amelyet egy négy évvel később megjelent cikkében (1998) részletesen bemutatott. Írásában megjegyezte, nem tartja valószínűnek, hogy az SGML egyszerűen és könnyedén a MARC helyébe lép, inkább a kettő valamilyen szimbiózisát véli elképzelhetőnek. Ebben az együttműködésben – írta – a MARC általánosan elfogadott adatkommunikációs formaként továbbra is jelen volna, de a például a bibliográfiai adatok intézményi kezelését kiválóan el lehet végezni SGML-alapú könyvtári rendszerrel is. Az ilyen szoftverek kedvező árai lehetővé teszik, hogy a könyvtárak komoly összegeket

spóroljanak, és az sem gond, hogy az intézmény nem hagyományos OPAC-ot használ. Mint azt a Cheshire II projekt példája is jelzi, könnyedén készíthető grafikus keresőfelület az új rendszerekhez is.

1995 őszén a Library of Congress egy tíz tagból álló munkacsoportot hívott össze egy „hivatalos” MARC DTD előkészítésére. A résztvevő szakemberek az alábbiakban összegezték a projekt elvárásait:

- a készülő DTD-nek általánosan használhatónak kell lennie;
- adatvesztés nélkül biztosítani kell a kétirányú konverziót a MARC és az SGML-formátum között;
- követnie kell a MARC mindenkori fejlesztéseit;
- az összesen öt USMARC-hívójelkészletet (Bibliographic, Holdings, Community Information, Authority, Classification) két DTD-be kell egyesíteni;
- a jelölőknek a MARC-hívójeleket kell tükrözniük (például mrcb245), így elkerülhetők a nyelvi különbségekből származó értelmezési problémák;
- a MARC-almezők mindegyikét külön jelölőnek kell megfeleltetni;
- a DTD-nek biztosítani kell az egyes elemek ismételhetségét és azok opcionális felhasználását (rugalmasság);
- kezelnie kell az ekkorra már elavult MARC-mezőket, de újonnan történő bevitelüket meg kell tiltania;
- fejlesztése során tekintettel kell lenni az ún. TEI-fejlécre és annak a MARC-rekordhoz hasonló funkcióira (lásd később), s javasolt a TEI-fejléc és az SGML-rekord közötti kapcsolat kialakítása (McCallum, 1996).

Az eredeti elgondolás szerint a Library of Congress által kidolgozandó DTD csupán kiegészítése lett volna a Berkeley-n alkalmazott változatnak. Hamar kiderült azonban, hogy túl nagyok az eltérések a Cheshire II által igényelt adatszerkezet és a Library of Congress elvárásai között. Végül egy külső cég, az ATLAS Systems készítette el a szükséges dokumentumtípus-definíciós állományokat 1996-ban (McCallum, 1996). Nem sokkal később a konverziós szoftverek is elkészültek – ezeket a könyvtár weblapján és FTP-szerveren tették közzé (*MARC SGML Archive*, 2003).

A Columbia University munkatársai digitális képkönyvtári szolgáltatásukat igyekeztek korszerűsíteni az SGML-technológia segítségével. Az általuk SCR (SGML Catalog Record) névre keresztelt tárolási formátum jóval bővebb, és több szintből álló adatszerkezeteket volt képes kódolni, mint a USMARC. Erre olyan esetekben lehetett szükség, amikor például több, különböző nézetből exponált kép készült ugyanarról a tárgyról, esetleg egyazon kép több különböző méretben vagy felbontásban szerepelt a gyűjteményben. Mivel azonban a közös egyetemi könyvtári, és magasabb, nemzeti szintű katalógusok kommunikációs formátuma a USMARC volt, nem lehetett tőle teljesen elszakadni. Az ellentmondást a projekt munkatársai úgy oldották meg, hogy elkészítették a rekordok „nyers” SGML-változatát (a feldolgozáshoz a Berkeley-n fejlesztett USMARC-DTD némiképp átdolgozott variációját használták), és az ebből konvertált, összefoglaló szintű USMARC-rekordok kerültek a katalógusokba. Ezekbe minden esetben hivatkozást helyeztek el, amely a bővebb SCR-változatra – helyesebben a belőle generált, képernyős megjelenítésre szánt HTML-dokumentumra – mutatott, amelyet az intézmény saját maga tárolt el, a leírt képdokumentumokkal együtt (Davis, 1996).

Az 1990-es évek első felében körülbelül negyven könyvtár együttműködésében üzemelő belga közös katalógus, a CCB (Catalogue Collectif de Belgique) rekordjait évente két CD-ROM-on is megjelentették. A projektben részt vevő tagkönyvtárak SGML-formátumban küldték el gyarapodásukat a kiadvány összeállítóinak – egy előre definiált CCB DTD szabályait követve –, ami elsimította a helyi feldolgozási szabályzatok okozta eltéréseket és minőségbeli különbségeket, és lehetővé tette nem MARC-alapú adatok közlését is, valamint biztosította az egységes megjelenítést. Hasonló elven működött az ország könyvtáraiban fellelhető időszakos kiadványok közös katalógusa, az Antilope is (Corthouts & Philips, 1996).

Az ETDML – elektronikus szakdolgozatok és disszertációk kezelésére szolgáló jelölőnyelv – fejlesztését a *Virginia Polytechnic Institute and State University* végezte az 1990-es évek második felében. A nyelv alapjául szolgáló DTD és a hozzá fejlesztett konverter (ETD2HTML) segítségével a beérkezett dokumentumokat egységes kinézettel lehetett szolgáltatni a felhasználók számára. Az ETDML első változata az SGML-t, míg a második már az XML-t használja (Kipp, Rahman & Bjorklund, 1998).

Az ELSA (Electronic Libraries SGML Applications) projektet egy háromtagú konzorcium koordinálta: az angliai De Montfort Egyetem, a francia Jouve System

d'Information fejlesztőcéggel – és az Elsevier. Célja a nemzetközi kiadóvállalat által fejlesztett SGML DTD-vel kódolt elektronikus dokumentumok (javarészt folyóiratcikkek, összesen kb. 5000 tétel) eljuttatása volt az olvasókhöz. Az informatikai háttérrel a Jouve gondoskodott, míg a hozzáférést biztosító kliensalkalmazást az angolok fejlesztették (Adams, 1995).

A HyperLib nevű kezdeményezés angol-belga együttműködésen alapult, feladata a könyvtári információs anyagok elektronikus, hipertextes formában történő terjesztése volt. A megfelelően tagolt szövegeket publikálás előtt a HyperLib DTD alapján jelölték (Corthouts & Philips, 1996).

2.3. Gyakorlati alkalmazások, jelentős DTD-k könyvtáron kívül

Az új jelölőnyelvi keretrendszer másik jelentős alkalmazási területe a szövegelőkészítés, a nyomdászat és a könyvkiadás volt. Nem sokkal az után, hogy az ISO megjelentette a 8879-es szabványt, amely az SGML-t definiálta, az amerikai könyvkiadók egyesülete (AAP) az Electronic Manuscript Preparation and Markup (*elektronikus kézirat-előkészítés és jelölés*) projekten belül rögtön készített is három DTD-t, amelyek az elektronikus publikálás céljára meghatározták az egyes dokumentumtípusok – könyvek, időszaki kiadványok, valamint cikkek – feldolgozásának mikéntjét (Cole & Kazmer, 1995). A kialakítás az OCLC közreműködésével történt, amely szervezet egyébként is az SGML korai támogatója volt: több folyóiratot publikált a rendszer segítségével, és kísérleti programokban is részt vett (Pitti, 1995a).

Az AAP által megfogalmazott dokumentumtípus-definíciók aztán maguk is nemzeti szabványok lettek ANSI/NISO Z39.59-1988 számon. A nemzetközi változat (ISO 12083) csupán hat évvel később, 1994-ben született meg, és nem váltotta be a hozzá fűzött reményeket, hiába promotálta számos kiadói szervezet, nem terjedt el olyan széles körben, mint várták (Corthouts & Philips, 1996). Említésre érdemes azonban, hogy 1992-ben John Duke és George Alexander az AAP kiadói DTD-jét felhasználva öntötte elektronikus formába az AACR2 angolszász katalógizálási szabályzatot, amellyel egyrészt megkönnyítették a jövőbeli nyomtatott változatok elkészítését, másrészt megteremtették a külső információforrásokra való hivatkozások elhelyezésének lehetőségét (Rubinsky, 1992).

A részleges sikertelenség oka az volt, hogy a gyakorlatban ahány kiadó, annyiféle különböző dokumentumtípus-definíció készült, ezek némelyike csupán a fejlécre, mások az

egész szövegre vonatkozóan is tartalmaztak előírásokat. Ilyen volt az európai fejlesztésű MAJOUR (Modular Application for Journals) DTD, amely kiadók és szerzők együttműködésével készült, és a cikkek fejlécében található bibliográfiai információk jelölésében nyújtott segítséget. De említhetjük az SSSH-t (Simplified SGML for Serial Headers), az Elsevier által fejlesztett Science Article DTD-t (Wusteman, 1997), a CALS-t, amely kezelési útmutatók és egyéb technikai dokumentációk kezelésére szolgált (McCallum, 1996), és számos más kezdeményezést. Vagy éppen a HTML-t, amely hipertext- és multimédiás dokumentumok leírását, jelölését teszi lehetővé napjainkban is.

2.3.1. TEI

A tudományos-filológiai célú szövegjelölés talán legismertebb eszköze napjainkban a TEI, amelynek fő funkciója általános tartalmú szövegek, szépirodalmi művek, kritikai kiadások, történeti források és előszöveg-átiratok elektronikus feldolgozása. Fejlesztését 1987-ben kezdte három számítógépes nyelvészettel, valamint irodalommal foglalkozó tudományos társaság, az Association for Computational Linguistics (ACL), az Association for Literary and Linguistic Computing (ALLC), valamint az Association for Computers and the Humanities (ACH). A TEI kialakításához kapóra jött az SGML-szabvány, hiszen biztosította azt a keretet, amelynek segítségével a szövegtulajdonságokat minél kevesebb veszteséggel, platformfüggetlenül és átjárhatóan lehet kódolni (Bíró, 2005). Jelen disszertáció írásának idején a TEI ötös számú verziója van érvényben, amelyet legutóbb 2021 augusztusában módosítottak (TEI, 2021). A teljes dokumentáció valamivel több, mint 2000 oldalas, ezért természetesen nincs mód jelen értekezés keretein belül részletesen bemutatni az összes használható jelölőt. A TEI-adatszerkezetek egy részlete, az ún. TEI-fejléc azonban mindenképpen említést érdemel. Ezt a szerkezeti egységet a TEI Szövegdokumentációs Bizottsága készítette, európai és észak-amerikai könyvtárosok és levéltárosok részvételével (Giordano, é.n.). A tervező munka során áttekintették az ISBD(G)-ben, valamint az AACR2-ben megfogalmazott előírásokat, de figyelembe vették a hivatkozások készítésének amerikai szabványát, az ANSI Z39.29-et is (TEI, 2004). A TEI-fejléc négy fő részből áll, amelyek az alábbiak:

- az elektronikus szöveg (és digitalizált dokumentum esetén a forrásdokumentum) bibliográfiai adatai, amely a visszakereséshez nélkülözhetetlen információkat tartalmaz, alkalmazása éppen ezért kötelező;

- a kódolás adatai – adminisztratív metaadatok, amelyek többek között a digitalizálási folyamat lépéseit, az alkalmazott eszközöket, a közben meghozott döntéseket és az esetleges hibákat tartalmazzák;
- a szöveg egyéb, nem bibliográfiai jellemzői, például megjegyzések a nyelvhasználatról, a szöveg keletkezésének körülményei stb. Javarást hangzó szövegek átírásakor és jelölésekor van jelentős szerepe;
- az ellenőrzések, módosítások listája – tartalma leggyakrabban az elvégzett változtatás leírása, a revízió dátuma és a módosító neve (Giordano, é.n.)

A TEI-fejlécek könyvtári szempontból fontos szerepe (és a rendszer fejlesztői szerint az egyik alapfeladata), hogy használatukkal – megfelelő konverziós eljárás után – előzetes, további adatokkal kiegészítendő MARC-rekordok készíthetők (Davis, 1996). Az adatok biztonságos konverzióját a Virginiai Egyetem könyvtára által 1996-ban közreadott szoftverrel (TEI2MARC) lehetett elvégezni, amelyet Jeff Herrin és Jackie Shieh fejlesztett, PERL nyelven. A szoftver a TEI Lite néven ismert DTD alapján dolgozik (*SGML: TEI to MARC*, 1996) – ez a kezdő alkalmazók számára nagy segítséget jelentő, leegyszerűsített, de a teljes változattal kompatibilis kódkészletet tartalmaz (Bíró, 2005).

A TEI-jelölőelemek és a MARC-mezők pontos megfeleltetése a rendszerdokumentáció P3-as változatában jelent meg elsőként (a 24. fejezetben), amely a nyers szövegekről leválasztott TEI fejlécek elkészítését tárgyalja (*The Independent Header*, é.n.); ezekből az alapszintű bibliográfiai adatokat tartalmazó adatszerkezetekből könnyedén építhetők különféle elektronikus gyűjtemények.⁴ A MARC és a TEI közötti összefüggések mélyebb megértésének céljából 1994 májusában háromnapos workshopot rendeztek a New Jersey-beli Somerssetben, amelyen az általános bevezető előadások után a résztvevők csoportmunkában foglalkoztak a MARC-rekord és a TEI fejléc alapvető (avagy: alapvetően eltérő) funkcióival, kapcsolódási pontjaikkal, valamint a jövőbeli használatukra, esetleges átalakításukra vonatkozó javaslatok megfogalmazásával (Horowitz, é.n.).

A TEI korábban említett, egyszerűsített változatát használta néhány évvel később az Empolis Magyarország Informatikai Kft., amikor 2003-2004-ben elkészítette a Magyar Elektronikus Könyvtár XML DTD-it a különféle dokumentumtípusok, próza, vers, dráma stb. szövegeinek jelölésére. Ezek a dokumentumtípus-definíciók alkotják ma is a MEK

⁴ A legfrissebb, P5-ös verzió ezeket az információkat már nem tartalmazza.

szolgáltatásának alapját (*DTD-k és metaadatkezelés a MEK-ben*, é.n.). Az MTA Irodalomtudományi Intézete DigiPhil-projektjének, valamint az OSZK Digitális Bölcsészeti Központjának munkatársai pedig ugyancsak a TEI segítségével végeznek igényes, filológiai gondosságú szövegjelölést.

2.3.2. EAD

Az EAD néven ismert, főként levéltárakban alkalmazott DTD fejlesztése a Kaliforniai Egyetem Berkeley campusán kezdődött 1993-ban, Berkeley Finding Aid Project néven, kettős céllal: egy különleges dokumentumtípus, az ún. keresési segédlet (*finding aid*) feldolgozási szabályzatának kidolgozása, valamint az elkészült rekordok adatbázisba szervezése érdekében (Pitti, 1995b). A keresési segédlet olyan tájékoztató segédeszköz, amely a levéltári anyagokban (kéziratok, levelezések, jogi dokumentumok, nyilvántartások, fényképek, hang- és képfelvételek stb.) történő eligazodást segíti. A magyar levéltári terminológia az iratanyag jellege és a feltárás mélysége szerint több ilyen segédletet is megkülönböztet, ilyenek például a *repertórium*, a *jegyzék*, a *lajstrom*, a *mutató*, a *regeszta* vagy az *ismertető leltár*. A keresési segédletek egyes típusainak sajátossága, hogy összefoglaló szintűek: nem a dokumentumok, hanem azok nagyobb csoportjainak adatait tartalmazzák, és segítenek annak eldöntésében, hogy érdemes-e áttanulmányozni az adott egységbe foglalt forrásokat.

Az ilyen levéltári segédletek számítógépes feldolgozásának története nem volt problémamentes. Az 1973-ban kiadott, levéltári célú Manuscripts (azaz kéziratok) című MARC-szabvány alkalmazása – amellet, hogy fejlesztése a Library of Congress kézirat-részlegének bevonása nélkül zajlott – nyilvánvaló tévutat jelentett, hiszen a többi dokumentumtípusnál bevált gyakorlatot követve maguknak a dokumentumoknak, iratoknak a feldolgozását helyezte a középpontba, és nem tartotta szem előtt a levéltárakban általánosan alkalmazott csoportos feldolgozás szükségességét. Ennek a problémának az áthidalására egy speciális MARC-hívójelkészlet, az AMC (Archival and Manuscripts Control) szolgált, amely egyes mezők tartalmának átértelmezésével (az összefoglaló szint kontextusába helyezésével), új mezők hozzáadásával (különösen a megjegyzések között) a diszkrét dokumentumok feldolgozása mellett a csoportos katalogizálásra is lehetőséget nyújtott, továbbá az 555-ös mező segítségével elvezetett a keresési segédletben rögzített, bővebb információkhoz (Pitti, 1997).

Az EAD kidolgozási munkálatainak elején a MARC is felmerült az adatok potenciális hordozójaként. Daniel Pitti beszámol róla, hogy a megfelelő adatmodell kiválasztása előtt két alapvető szempontot kellett figyelembe venni:

- az alkalmazandó rendszer szabványos, nyilvános hozzáférésű legyen, ne függjön semmilyen hardver- vagy szoftvergyártótól;
- legyen képes a levéltári segédletek széles spektrumában megfogalmazott szellemi tartalom, valamint a komplex dokumentumszerkezet hűséges visszaadására, elősegítve a kereshetőséget és megjelenítést.

A szerző hozzáteszi, hogy a MARC deskriptív mivolta, nyilvánossága, könnyű konvertálhatósága, mindenekelőtt pedig tiszteletet érdemlő kora és a változó környezet ellenére sikeres túlélése miatt került a képbe. Alkalmazásával kapcsolatban azonban súlyos problémák merültek fel:

- a MARC-rekordok maximális mérete 100.000 karakter, a segédletek jó része pedig hosszabb ennél;
- a MARC nehezen birkózik meg a hierarchikusan szerveződő információkkal, a keresési segédlet pedig összefoglaló szintű adatrögzítést igényel, amelyben nem mindig elegendő egyetlen szint használata. Ezért a feldolgozás többszörös koncentrációt igényel a hagyományos dokumentumokhoz képest, ezen kívül számos hibalehetőséget és rendszerproblémát rejt magában;
- a MARC használata a potenciális célszoftverek szűk piaca miatt jelentős költségekkel jár.

Fentiek megfontolása után a könyvtári adatsere-formátum használatát végül elvetették. Az Amerikai Levéltárosok Társaságának közreműködésével készült EAD nagyjából háromévi munka után végül SGML-alapú dokumentumtípus-definícióként látta meg a napvilágot 1996-ban. Napjainkban alkalmazott harmadik verziója már XML DTD-ként, valamint XML-sémaként érhető el.

2.4. Az SGML Magyarországon

Hazánkban az első SGML-kísérletek az 1997-ben alapított Neumann János Digitális Könyvtárhoz kapcsolódnak. Digitális szövegek közzétételével ekkoriban már a Magyar Elektronikus Könyvtár is foglalkozott, ők azonban TXT- és Microsoft Word formátumokban

szolgáltatták a dokumentumokat. A Neumann-ház Tószegi Zsuzsanna szerint a kezdetektől fogva egy komplexebb szolgáltatást lehetővé tévő, időtálló formátumot kerestek, amelyből könnyen állíthatók elő más formátumú kimenetek (Tószegi & Kora, 2003). A Digitális Irodalmi Akadémia projektben részt vevő szerzők életművét a TEI DTD segítségével készítették elő megjelenítésre, de a Neumann-háznak 1998-tól saját fejlesztésű, irodalmi művek feldolgozására szolgáló DTD-je, valamint szöveg megjelenítő stíluslapja is volt, amelyet a Bibliotheca Hungarica Internetiana gyűjtemény építéséhez vettek igénybe. A szakemberek szerint az SGML formátum használatával a szövegek megjelenítése rugalmasabb lehet, hiszen a címkék semmilyen formai előírást nem tartalmaznak. Így azokat a célközönség aktuális igényeitől függően lehet előkészíteni a képernyőn történő megjelenítésre (Kora, 2003).

A Magyar Tudományos Akadémia Nyelvtudományi Intézete házi készítésű SGML DTD-t használt a magyar irodalmi és köznyelv nagyszótára előkészítéséhez; a szerkesztési munkálatokban a francia fejlesztésű WriterStation szoftver volt a munkatársak segítségére (Pajzs, é.n.). Ugyancsak az MTA berkein belül fejlesztették a Magyar Történelmi Szövegtárat is, amelyben XIX-XX. századi szépirodalmi szövegek összetett, nyelvészeti keresését lehet elvégezni. A szolgáltatás technológiai alapját az SGML biztosította (Moldován, 1999). A szabványosított jelölőnyelv alkalmazásán nyugodott a MorphoLogic szoftverfejlesztő cég összes szótárának számítógépes leírása is (*MetaMorpho*, 2001).

3. Az XML megszületése

Az SGML komoly hatással volt a könyvtári és az általános informatikára, és DTD-k, illetve szolgáltatások sokaságát fejlesztették a segítségével. Igazán széles körben azonban mégsem terjedt el, amiért többnyire összetett, bonyolult, nehezen tanulható struktúráját okolták, amely bizonyára sok kezdeményezésnek állta útját. Nagyon kevés volt az igazán jó SGML-szakember: az előző fejezetben bemutatott példák igazolják, hogy a dokumentumtípus-definíciók elkészítésével speciálisan erre a területre szakosodott cégek (például az ATLIS vagy a SoftQuad) foglalkoztak, amelyek munkáját alaposan meg kellett fizetni. Ezen felül szűk volt a feldolgozáshoz használható szoftverek piaca, amely – áttételesen – ugyancsak a költségek növekedéséhez vezetett. Az SGML-dokumentumok megfelelő használatához ráadásul minden esetben szükség volt külső DTD-re, egy külön állományra, amely záloga volt az egységes formában történő feldolgozásnak – a több résztvevős projektek által használt, vagy a szabvánnyá vált dokumentumtípus-definíciók terjesztése pedig elég körülményes volt.

1996-ban – éppen abban az időben, amikor a World Wide Web-konzorcium a HTML-nyelv fejlesztésén is erőteljesen gondolkodott – a fent ismertetett visszasságok miatt már igen sokan szorgalmazták az SGML egyszerűsített változatának elkészítését, amely megfelel nem csupán a szakemberek, de a nagyközönség igényeinek is. Az új jelölőnyelv megalkotásakor a fejlesztők az alábbi célokat tartották szem előtt (Bíró, 2005):

- legyen nyíltan, könnyen használható, biztosítsa a dokumentumok könnyű előállítását;
- legyen széles körben alkalmazható, támogassa a legkülönbözőbb információk leírását;
- legyen kompatibilis elődjével, az SGML-lel;
- legyen könnyen támogatható, könnyen lehessen feldolgozó szoftvereket írni hozzá;
- egy műveletet lehetőség szerint csak egyféleképpen lehessen elvégezni (ez segítséget nyújt a feldolgozó szoftverek fejlesztésében is);
- ne tartalmazza az SGML-ben meglévő minimalizálási lehetőségeket, azaz egyes jelölőelemek opcionális elhagyhatóságát;
- legyen érthető, olvasható az emberi szem számára is;
- legyen gyors a fejlesztése;
- legyen formális és tömör.

A felsorolt alapelvek mentén kifejlesztett ún. kiterjeszhető jelölőnyelv (eXtensible Markup Language) első változatának specifikációja 1997 utolsó napjaiban (mások, például Salgáné (2004) szerint valamivel később, 1998-ban) jelent meg, és hatalmas siker lett az informatika szinte minden területén. (Egy jellemző példa: napjainkban már a napi rendszerességgel használt Office-alkalmazások jó része is XML-struktúrában tárolja a dokumentumokat, amire a fájlnév kiterjesztéséhez hozzáadott „x” utal: doc → docx; xls →.xlsx; ppt → pptx).

A szakemberek egymással versengve sorolták az új jelölőnyelv előnyeit:

- átjárhatóságot, könnyű adatmigrációt és platformfüggetlenséget biztosít az SGML bonyolultsága nélkül (Zuber, 2005);
- informatikai „eszperantót” jelent, amelyen mindenki olvashat, írhat és számíthat a megértésre (Salgáné, 2004);
- az adat- és dokumentumcsere univerzális formátuma, sőt de facto szabványa, kivételes rugalmasságával az információs kor lingua franca-jaként funkcionál (Miller, 2000);
- olyan hordozó, amely a források megtalálhatóságát, az információk keresését és a webes információcserét egyaránt elősegíti (Jacsó, 2002);
- nyílt és transzparens formátum, amely a HTML és az SGML házasságából született (Tennant, 2001), de nélkülözi az SGML bonyolultságát és a HTML egycélúsága helyett univerzális használhatóságot biztosít.

Az XML alapfunkciója mindazonáltal változatlan maradt: metanyelv, amely számtalan specifikus jelölőnyelvet, nyelvtant definiálhat. Ennek eszköze, akárcsak az SGML esetében, a dokumentumtípus-definíciós állomány, amely elhelyezkedhet magában a dokumentumpéldányban (ekkor belső DTD-nek nevezik), vagy külön file-ban (külső DTD) (Bíró, 2005). A 2000-es évektől kezdve azonban már inkább az ún. XML-sémákat alkalmazzák formázásra és a dokumentumok érvényességének biztosítására, mivel azok, a DTD-vel ellentétben, alkalmasak a nagyméretű, összetett dokumentumtípusok leírására is, lehetőséget biztosítanak az elemtartalmak és attribútumaik tipizálására, és nem utolsósorban az XML saját szintaktikáját használják, ezért a hatékony munkavégzéshez nem kell egy másik, speciális nyelvtant is elsajátítani (Tóvári & Szabó, 2011).

Az XML megjelenésére a könyvtári világ rendkívül gyorsan reagált: már 1998-ban több írás jelent meg az új jelölőnyelv metaadatszabványokra gyakorolt hatásáról, továbbá a MARC-kal való (lehetséges) kapcsolatáról (De Carvalho et al., 2004). A szakmai közösség még lelkesebben vette számba az XML könyvtári szolgáltatásokban való alkalmazhatóságának lehetőségeit, mint azt az SGML esetében tette a '90-es évek első felében, de sok esetben csupán megismételte a korábban már elhangzott érveket:

- az XML kiválóan használható strukturált dokumentumok elektronikus formában történő továbbítására, cseréjére (Jaksič, 2004), miközben jóval gazdagabb információ-reprezentációt tesz lehetővé, ami növeli a visszakereshetőséget (De Carvalho & Cordeiro, 2002);
- szoftveres támogatottsága nagyságrendekkel jobb, mint a MARC-é, így könnyebbé és olcsóbbá válik az új alkalmazások készítése (Keith, 2004);
- alkalmazása nyíltabb könyvtári rendszerek létrehozását teszi lehetővé. A mai világban „követelmény, hogy a könyvtári OPAC-okhoz hasonló közszolgáltatások nyíltak, átjárhatók, más keresők által elérhetők, mélyen linkelhetők legyenek” (Salgáné, 2004, p. 61.);
- segítségével kiterjeszhető a bibliográfiai rekordok tartalma, hiszen – a MARC-kal ellentétben – támogatja a hierarchikus dokumentumszerkezetet (Zuber, 2005);
- az XML alkalmazása révén a könyvtári adatok kikerülnek a mély, rejtett web tartományából, és elérhetővé válnak a keresőrendszerek számára (De Carvalho & Cordeiro, 2002). Dick Miller véleménye szerint a könyvtári információ nagy vesztese a fejlődő World Wide Webnek. A weben elérhető információk, szkennelési projektek, digitális repozitóriumok és weboldalak hada ellenére a könyvtári adatok még mindig egyedi formátumban (a MARC-ban) tárolódnak, és nehézkesen, szinte csak rendszerkönyvtárosok vagy specifikus szolgáltatások számára használhatók igazán jól (Johnson (2001) ezt a jelenséget nevezi *bibliográfiai apartheidnek*). Egyre fontosabb, hogy a bibliográfiai források integrálódjanak a weben elérhető források közé, különben a könyvtár jelentős versenyhátrányba kerül (Miller, 2000).

Az SGML esetében, bár a könyvtáros közösség lelkesen üdvözölte az új technológia előnyeit a MARC-kal szemben, többen – köztük Jerome McDonough (1998) is – úgy vélekedtek, hogy valamiféle szimbiotikus együttműködésre, komplementaritásra, a feladatok gyakorlati szempontú, logikus elosztására kell számítani a két formátum között.

Az XML tekintetében már komplexebb volt a helyzet. Az ebben az időszakban keletkező szakirodalom⁵ az új technika melletti radikális elköteleződésre biztatta a könyvtárosokat. E nézet képviselői úgy gondolták, a MARC-ot amint lehet, le kell váltani egy hasonlóan gazdag, de kötöttségektől mentes formátummal, ami hátráltatás helyett inkább elősegíti a kísérletezést (Dorman, 1995). A másik oldalon azok álltak, akik – az SGML-korszakhoz hasonlóan – ismét elismerték az aktuális újítás jelentette előnyöket, s megint csak nem a kizárólagosságban, hanem párhuzamosságban gondolkodtak (Johnson, 2001). Úgy vélték, egy teljesen XML-alapú könyvtári rendszer kifejlesztése nagyon sok bizonytalanságot hordoz magában és egyáltalán nem biztos, hogy megtérülnek a ráfordított költségek (Zuber, 2005). Bruce Johnson még 2001-ben is úgy vélte, hogy nem állnak rendelkezésre a szükséges adatlétrehozó, -kezelő és -megjelenítő szoftverek (például böngészők), az új technológiát pedig nagyon kevesen ismerik és alkalmazzák igazán értő módon. Jacsó Péter, a hawaii egyetem magyar származású professzorának szavaival így foglalhatnánk össze a 2000-es évek elejének könyvtári status quóját: *„a cél a Kánaán, de ha engem kérdeznek, hogy ott vagyunk-e már, azt kell mondjam, nem. Még nem, de már úton vagyunk.”* (Jacsó, 2002, p. 49.)

3.1. Kísérletek a MARC-XML konverzióra

A könyvtárosok munkamódszere az XML esetében is ugyanaz volt, mint az SGML-nél: a szolgáltatásokhoz, adatkonverziókhöz számos, a konkrét célokat megfelelően kiszolgáló DTD-t (később sémát) fejlesztettek, amelyek közül egyesek világszerte ismertté váltak, és használatuk széles körben elterjedt, mások azonban elszigeteltek maradtak és alkalmazásukról nagyon kevés információ lelhető fel.

David J. Fiander (2001) tanulmányában azt vizsgálja, vajon milyen úton kell elindulni egy könyvtári alkalmazású XML DTD kifejlesztésekor. Három utat ajánl:

- az egyik megoldás szerint pontosan meg kell őrizni a MARC szerkezetét, mezőit és almezőit, és ezeket kell belekódolni a dokumentumtípus-definícióba. Ez egyrészt könnyebbséget jelent a feldolgozó könyvtárosoknak, hiszen nem kell új jelölésrendszert megtanulniuk, másrészt lehetőséget ad az adatvesztés nélküli,

⁵ „Stratégiák a MARC-ból való kilépésre” (Tennant, 2010) – „A MARC-nak pusztulnia kell” (Tennant, 2002) – „Szükségünk van-e még a MARC-ra?” (Hopkinson, 1994) – „Mehalt a MARC?” – „A MARC vége?” (Johnson, 2001)

ellentétes irányú konverzióra – azaz XML-ből MARC-ba –, amire a vélhetően hosszú átállási periódus miatt van szükség. Ez alatt a hagyományos katalógusok építését sem szabad elhanyagolni.

- a másik megoldás abból a problémából indul ki, hogy a MARC-rekord alapszerkezete eltér a nyomtatott szabályzatokban meghatározott feldolgozási sorrendtől (Fiander szövegében az AACR2 szerepel, de a megállapítás az ISBD-re is igaz). A MARC-ot eredetileg a katalóguscédulák gyártásának megkönnyítésére fejlesztették ki, ezért mezőstruktúrája a tétel kötött szerkezetét követi: leíró mezők után authority mezők következnek, majd utána ismét leíró mezők stb. Fiander javaslata, hogy az XML-adatszerkezet ne a MARC-ot, hanem a szabályzatokban rögzített eredeti feldolgozási sorrendet kövesse, ami a munkafolyamat egésze szempontjából, de különösen a hozzáférési pontok megállapításánál hordoz magában előnyöket.
- a harmadik javaslat: egy új katalogizálási workflow DTD-be kódolása. Ez jelentheti az FRBR (lásd a következő fejezetet) elveinek erősebb megvalósítását – azaz a mű, illetve a példány adatainak külön-külön rekordban történő elhelyezését, majd e rekordok összekapcsolását; de akár csak annyit, hogy ezentúl nem szerepeltetnek a rekordokban 100-as mezőt annak lényegi funkcióvesztése miatt, tehát nem vesznek fel személynév típusú kiemelt hozzáférési pontot, hanem minden kapcsolódó személynevet a 700-as mezőbe rögzítenek, további hozzáférési pontként. A modern katalógusrendszerek többsége ugyanis úgy van konfigurálva, hogy egyszerre keresi e mezők tartalmát, így lényegi eltérés nem mutatkozik majd az eredményekben. Fiander gondolata a jelenlegi trendek ismeretében profetikusan bizonyult: *„A jelenleg egyetlen MARC-rekordban tárolt adatok felbonthatók lennének a mű adataira, az alkotó [creator] adataira, a manifesztáció adataira, majd ezeket újra össze lehetne kapcsolni. Megfelelő, fejlett felhasználói interfész segítségével a könyvtárhasználók bejárhatják a rekordok hálóját, hogy megtalálják a kapcsolatokat a különféle típusú adatok között”* (Fiander, i.m. p. 26.).

A Logos Research Systems nevű amerikai cég 1998-ban készítette el egyszerű konverter szoftverét Bob Pritchett vezetésével, amely képes volt MARC-ból XML-be és XML-ből MARC-ba adatokat átalakítani. A szolgáltatás weblapján online szolgáltatásként kipróbálható volt az egyébként letöltés után működő szoftver (de csak Windows 95 és Windows NT operációs rendszereken): az alkalmazás számára egyszerűen a vágólapról

voltak beilleszthetők a MARC-rekordok vagy az XML-struktúrák. A visszaalakítás előtt az XML-file tetszés szerint módosítható volt. Ugyanerről a weblapról egy USMARC-elemző alkalmazást is tesztelni lehetett, amely HTML- vagy nyers szöveges formátumban rövid tájékoztatást nyújtott az elemzésre átadott rekordban használt mezőkről, almezőkről, indikátorokról stb., illetve bizonyos mezőtulajdonságokról, köztük az ismételhetőségről vagy az elavultságról (Logos Research Systems, 2001).

A Stanford University Lane Medical Library-ben 1999 végén-2000 elején végrehajtott informatikai fejlesztések részeként született meg egy olyan, Java-alapú kliens-szerver alkalmazás (XMLMARC néven), amely MARC-rekordokat konvertált XML-lé. A szoftver alapját képező két DTD-t úgy készítették el, hogy az megszüntesse a MARC-ban szükségtelennek ítélt redundanciákat (így például a dokumentum nyelvéről szóló információk esetében, amelyek nem kevesebb, mint ötször tárolódnak el). A projektben dolgozó szakemberek a korábban említett radikális oldalt képviselték, hiszen már az ezredfordulón úgy hitték, hogy a MARC már nem releváns a könyvtári munka szempontjából, és a jövőben sem lesz az, így programjuk nem tartalmazott fordított, XML-MARC-irányú konverziós lehetőséget (Johnson, 2001).

A szoftver nem kereskedelmi használatra ingyenesen letölthető volt a projekt weblapjáról, és körülbelül negyven országban használatba is vették a rekordok tömbösített konverziójának megkönnyítése érdekében. A szoftverhez egy különleges DTD (később séma) is tartozott, a MARCUTL, amely nem csupán átalakításra, hanem a rekordok tartalmának frissítésére is alkalmas volt, például bizonyos feltételek teljesülése esetén el tudott helyezni a MARC-rekordokban egy új mezőt (Miller & Clarke, 2003).

3.2. A Library of Congress és a MARC/XML

Az ATLAS által a Library of Congress részére készített két SGML DTD négy éven keresztül, 1996-tól 2000-ig állt a felhasználók rendelkezésére. Ezután XML-be konvertálták azokat, és bár 2002-től sémát is lehet alkalmazni konverziós célokra, a DTD-eket még 2006-ig rendszeresen aktualizálták, karbantartották. A dokumentumtípus-definíciós állományok – megőrizve az SGML korában kialakított struktúrát – kettő (*mrca* elnevezéssel: Authority, Classification), illetve három (*mrcb* néven: Bibliographic, Holdings, Community Information) MARC 21-es formátumot fedtek le, amelyek részletes kódolása igen nagy

méretű, 500 és 240 kilobájtos file-okat eredményezett (Keith, 2004). A következő kódrészlet a MARC 21 100-as adatmezőjének paramétereit határozza meg az XML-ben:

```
<!ELEMENT mrcb100 (mrcb100-a | mrcb100-b | mrcb100-c | mrcb100-d | mrcb100-e | mrcb100-f | mrcb100-g | mrcb100-j |
mrcb100-k | mrcb100-l | mrcb100-n | mrcb100-p | mrcb100-q | mrcb100-t | mrcb100-u | mrcb100-4 | mrcb100-6 | mrcb100-
8)*>
<!ATTLIST mrcb100
  name CDATA #FIXED "MAIN ENTRY--PERSONAL NAME"
  obsolete CDATA #FIXED "no"
  repeatable CDATA #FIXED "no"
  i1 (i1-0 | i1-1 | i1-2 | i1-3) #REQUIRED
  i2 (i2-blank | i2-0 | i2-1) #REQUIRED
>
<!ELEMENT mrcb100-a (#PCDATA)>
<!ATTLIST mrcb100-a
  name CDATA #FIXED "Personal name"
  obsolete CDATA #FIXED "no"
  repeatable CDATA #FIXED "no"
```

A DTD-k megközelítésével szemben a Corey Keith által kidolgozott MARC/XML-séma általánosabb kategóriákat (mező, almező stb.) jelölő tagekkel dolgozik, s ezek attribútumai hordozzák a lényegi információt. A mezők esetében például három kötelező attribútumot definiál: a mező hívójelét és a két indikátorértéket. Talán ennek az univerzális ábrázolásmódnak köszönhető, hogy a MARC/XML „hivatalos”, kongresszusi könyvtári verziója nem csupán Amerikában, hanem az egész világon, így Magyarországon is elterjedt. Több közkedvelt integrált könyvtári rendszer (például az Aleph vagy a Huntéka) tartalmaz MARC/XML-konverziós modult, amely a MARC-ban tárolt rekordokat egy kattintással exportálhatóvá teszi további felhasználás céljára.

A formátum azonban nem mindenhol váltott ki osztatlan lelkesedést, ami részben annak volt köszönhető, hogy a MARC/XML nagyon erősen a MARC 21 sajátosságaira támaszkodott. Emiatt az attól kisebb-nagyobb mértékben eltérő (például kettőnél több indikátorértéket vagy egynél több karakterből álló almező-azonosítót használó) nemzeti MARC-hívójelkészletek, köztük például a danMARC2, háttérbe szorultak (Freire, Borbinha & Manguinhas, é.n.).

Ennek a problémának az áthidalására dolgozták ki a MarcXchange elnevezésű XML-sémát, azt a célt szem előtt tartva, hogy a készülő produktum általánosabb legyen a MARC/XML-nél, abból a célból, hogy minden helyi MARC-dialektust képes legyen támogatni (Gjerde, 2008), ugyanakkor a lehető legkisebb mértékben térjen el tőle a kompatibilitás érdekében. Karen Coyle amerikai könyvtári szakértő 2007. augusztusában kelt blogbejegyzésében bemutatja, hogy a javasolt séma mennyiben rugalmasabb a

MARC/XML-nél: a hívójelek nem csupán három számjegyből állhatnak, hanem betűket is magukba foglalhatnak, így lefedve a 00A-ZZZ tartományt is, továbbá az adatmezőkhöz nem csupán kettő, hanem akár (és legfeljebb) kilenc indikátorérték is definiálható (Coyle, 2007). A séma kifejlesztésének munkálatai 2003-ban kezdődtek el, a végeredmény – az ISO 25577-es nemzetközi szabvány – 2008-ban látta meg a napvilágot. Átdolgozott változata ISO 25577:2013 számon ma is érvényes, a gyakorlatban azonban kevésbé terjedt el.

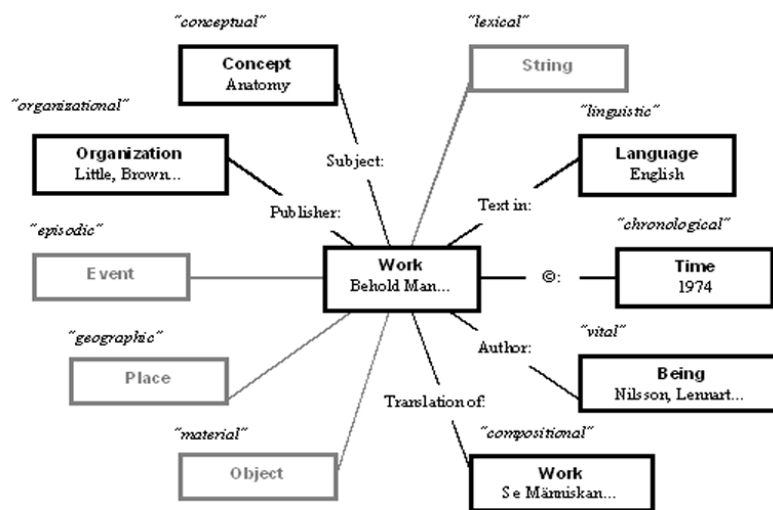
3.3. Egy elfeledett modell – a XOBIS

Az XML talán egyik legérdekesebb, rendkívüli mértékben előremutató könyvtári tárgyú alkalmazása a **XOBIS** néven fejlesztett séma, az XMLMARC továbbgondolása, amelyet a RELAX NG sémanyelv segítségével készített el a Stanford University projektjének tíz tagból álló munkacsoportja 2001-2002-ben (*About XOBIS*, é.n.). Speciális megközelítése, valamint a MARC-tól való erőteljes elrugaszkodása miatt érdemes bővebben tárgyalni, annál is inkább, mert a XOBIS számos alapelve a kortárs könyvtári informatikában is visszaköszön olyan technológiákba csomagolva, amelyek a 2000-es évek elején még csak gondolatban sem voltak meg. Nem véletlenül állítja Mirjana Jakšič, hogy a XOBIS jóval több annál, mint hogy a MARC-ban tárolt adatokat XML-ben újracsomagoljuk (Jakšič, 2004).

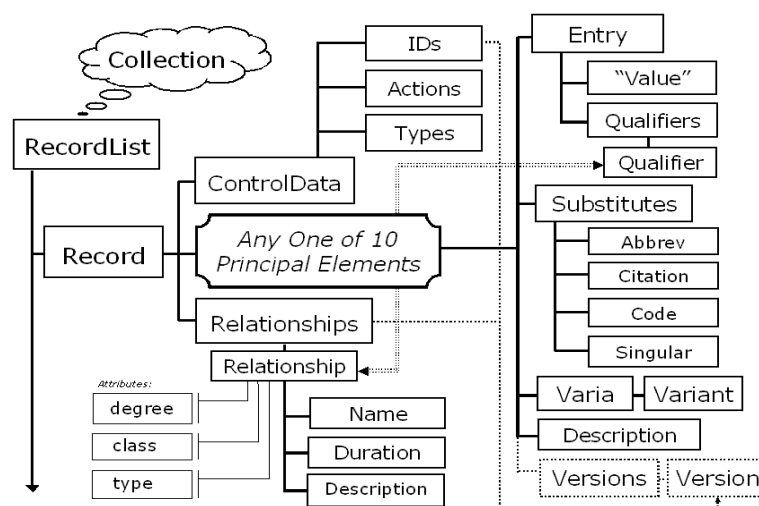
A séma egyik megalkotója, Dick Miller (2005) szerint: „*a XOBIS [...] a bibliográfiai és authority adatokat egy egyedülálló, egységes struktúrába szervezi újjá. Egyensúlyozni igyekszik az értékes hagyományok és az új technológiák között, hogy a jövőben alapja lehessen a [...] digitális környezetben megvalósuló információ-hozzáférésnek*” (p. 285.). A szerző írása egy másik pontján hozzáteszi, hogy a séma valójában a középutat jelenti a MARC összetettsége és a Dublin Core túlegyszerűsítése között.

A XOBIS működése a különféle bibliográfiai entitások és a köztük lévő kapcsolatok leírásán alapszik; ez utóbbira jelentőségéhez képest eddig kevés figyelmet fordítottak, pedig ez biztosítja a tágabb kontextust, rálátást az adott információkvantum szűkebb-tágabb környezetére. A séma kidolgozói szerint bármi, ami a világban absztraktnan vagy kézzelfoghatóan létezik, tíz különféle kategóriába sorolható be, amelyeket a séma *principal element*nek (főelemnek) nevez, s feltűnő jellegzetességük, hogy megalkotásuknál figyelembe vették az FRBR-ben (lásd később) meghatározott és három csoportba sorolt alapentitásokat (1. ábra). Ezek a főelemek a következők: mű, nyelv, időpont, lény (*being* –

rendszerint személy, de akár tulajdonnévvel rendelkező állat, például Bodri is), fogalom, esemény, földrajzi hely, tárgy, szervezet és karakterlánc (például szavak megfelelői különféle nyelveken). A főelemek között kapcsolatok állhatnak fent, s mivel bármelyik csoport elemei kapcsolódhatnak bármelyik másikkhoz, összesen száz különféle kapcsolattípus létezik a XOBIS-on belül. Ezek neve a főelemekhez rendelt alapértelmezett típusok összepárosításából származik: például *conceptual-chronological*; *compositional-geographic*; *compositional-vital*, s így tovább (Miller & Clarke, 2003).



1. ábra – A XOBIS főelemei és azok alapvető kapcsolatai (Miller, 2005)



2. ábra – A XOBIS séma strukturális felépítése (Miller, 2005)

A XOBIS alapszerkezetét a 2. ábra mutatja be. A séma gyökéreleme a RecordList, amely valamely gyűjtemény leképezése: ez egyedi rekordokból áll, amelyek mindegyikét három összetevő alkotja. A Control Data adminisztratív metaadatokat tartalmaz a rekordról

(létrehozás, módosítás, rekordtípus stb.); ezt követi a tíz Principal Element valamelyike, amelyek konkrét előfordulásait ún. Entry-k különböztetik meg egymástól (leggyakrabban az előfordulások konkrét megnevezése: *angol nyelv, Bulgária, George Orwell stb.*). Ha a kétértelműség elkerülése végett erre szükség van, az Entry-ben minősítőket kell alkalmazni (például London (Kanada)). A többszemponú keresés elősegítésére alkalmazhatók különféle alternatív elnevezés-alakok is (például rövidítések, kódok, vagy egyszerűen a megnevezés egyes számú alakja), vagy ún. variációk (például címváltozatok, szerzői álnévek stb. esetében). A harmadik összetevő maguknak a kapcsolatoknak (Relationships) a leírása, amely egyebek mellett a kapcsolat típusát, fajtáját (alá-, fölérendelő, előzmény/folytatás stb.), a kapcsolat erősségét (például egy mű fő- és melléktémái, elsődleges szerzője stb.) tartalmazhatja.

A XOBIS erősségét megalkotói univerzalitásában látták: nem csupán könyvtári, de múzeumi és levéltári anyagok leírására is alkalmas, ezáltal egymással kompatibilis rekordok jöhetnek létre az egyes memória-intézményekben, amelyek egyszerre szolgálhatják a helyi igényeket, ugyanakkor megfelelő, egységesítő XSLT-transzformáció után részét képezhetik egy közös felületnek, amely így – Dick Miller gondolataival élve – több lehet, mint az egyes részek összege (Miller & Clarke, 2003).

3.4. Jelentősebb XML-implementációk

Az UNIMARC-ot használó portugál nemzeti könyvtár és a BookMARC nevű könyvtárinformatikai cég együttműködéséből az XML alkalmazásának egy egészen különleges modellje született meg. Joaquim de Carvalho és Maria Ines Cordeiro a 2002-ben rendezett, 68. IFLA-kongresszuson mutatták be az ún. **TVS-modellt**, amelynek elnevezése az XML három legfontosabb implementációs területéből, a rekordok és az információ cseréjéből, szállításából (*Transportation*), a szerkezeti ellenőrzési funkcióból, a komplex érvényességi szabályok reprezentációjából (*Validation*) és a kialakítható megosztott szolgáltatásokból (*Shared Services*) áll össze (De Carvalho & Cordeiro, 2002).

A **MARC4J** nevű Java-szoftverkönyvtár (korábbi nevén James: Java MARC Events) fejlesztését 2002 óta végző Bas Peters (és egy ideig párhuzamosan Kevin S. Clarke) célja egy olyan open source környezet, API kialakítása volt, amelynek segítségével platformfüggetlen, több MARC-hívójelkészletet és MARC/XML-t egyaránt kezelni képes Java-alkalmazások készíthetők. A MARC4J biztosítja a két adatformátum közötti konverziót

is, sőt XSLT-utófeldolgozás eredményeként MODS-kimenet is előállítható a segítségével. Használatba vételéhez mindössze telepített Java Development Kit szükséges, a MARC4J-fájlok a GitHub weboldalról letölthetők (Peters, é.n.).

A *BiblioML* egyike azon kevés alkalmazási projekteknek, melyek kormányzati kezdeményezésre indultak el. A francia Kulturális és Kommunikációs Minisztérium kérésére 1999 végén kezdtek fejleszteni egy-egy, bibliográfiai, illetve authority információk tárolására, továbbá UNIMARC-rekordok konverziójára alkalmas DTD-t, amelyeket BiblioML-nek, illetve AuthorityML-nek neveztek el. Előbbi körülbelül 230, utóbbi nagyjából 80 címkét tartalmaz, s a dokumentumtípus-definíciók alapján a definiált tag-ek beszédes megnevezésüket, adattartalmukat, illetve hierarchiájukat (a gyökérelem a BiblioRecord, illetve az AuthorityRecord nevet viseli) tekintve az alábbiakban bővebben is ismertetendő MADS-, illetve MODS-formátumra emlékeztetnek. Az egyes címkék leírását azok UNIMARC-megfelelőire vonatkozó információk egészítik ki. Rendelkezésre állt egy Java-alapokon nyugvó konverziós segédeszköz is, mely a Saxon XML-feldolgozó alkalmazás segítségével végezte az UNIMARC-ról XML-re történő átalakítást (BiblioML, 1999).

A *METS* egy, a '90-es években indult, egyetemi könyvtárak által koordinált, Making of America II nevű projektből nőtt ki magát. Az eredeti célkitűzés egy többféle metaadattípust (deskriptív, adminisztratív, strukturális stb.) támogató DTD kifejlesztése volt, amelynek segítségével bizonyos típusú digitális könyvtári dokumentumokhoz – szövegekhez, álló- és mozgóképekhez, hangfelvételekhez – bőséges kiegészítő információkat lehet társítani. Az idő múlásával az eredeti dokumentumtípus-definíciót XML-sémává dolgozták át annak a Jerome McDonough-nak a részvételével, aki a Cheshire-projekt és a Library of Congress SGML DTD-jének elkészítésében is fontos szerepet vállalt.

A METS-dokumentum egyik opcionális szerkezeti egysége a leíró metaadatok szekciója, amely – bár saját elemkészletet nem tartalmaz az ilyen típusú információk rögzítésére – lehetővé teszi, hogy például MARC-ban tárolt katalógusrekordokra mutató hivatkozások jelenhessenek meg az adatszerkezetben, de más metaadatsémák, például a MODS alkalmazásával maguk a leíró metaadatok is beágyazhatók (Cundiff, 2004) – e beágyazások miatt a METS-et ún. konténer-formátumnak is nevezik.

A **MODS** nevű XML-sémát a 2000-es évek eleje óta ismeri a könyvtári világ. Fejlesztője a Library of Congress hálózatokkal és MARC-szabványokkal foglalkozó osztálya volt, akik az idő múlásával egyre gyakrabban szembesültek a MARC 21 már-már kezelhetetlenül nagy méretére és komplexitására vonatkozó panaszokkal. Sokan a rendszer hibájául rótták fel azt is, hogy használata – miután nem természetes nyelvi kifejezéseket, hanem számkódokat, hívójeleket alkalmaz – komoly előzetes felkészülést igényel. A MODS alkotói, köztük Rebecca Guenther, ezekre a problémákra igyekeztek választ keresni, miközben arról sem feledkeztek meg, hogy a digitális dokumentumok korának hajnalán kialakított Dublin Core metaadatstruktúra egyre kevésbé felel meg a növekvő elvárásoknak; korlátozott számú elemével nehézkesen birkózik meg a komplexebb szerkezetű digitális objektumokkal (Guenther, 2003)

A MODS nem más, mint a MARC 21 egyszerűsített, XML struktúrában megfogalmazott, kevesebb elemet tartalmazó változata, amely a Dublin Core-nál lényegesen gazdagabb feltárást tesz lehetővé, s feloldja a komplexitással és érthetlenséggel kapcsolatos problémákat. A hívójelek helyett szöveges jelölőket alkalmaz, ezzel megkönnyítve, és nem kis mértékben meggyorsítva a feldolgozó munkát. Az egyszerűsítést nem egy esetben úgy éri el, hogy egyetlen MODS-jelölőt egyszerre több MARC-mező adatainak fogadására is alkalmassá tesz. Az átalakítás során például a 256-os (elektronikus dokumentum jellemzői), 300-as (terjedelem) és részben a 856-os (elektronikus hely és hozzáférés) mezőjének adattartalma a `physicalDescription` jelölő attribútumaiban olvad össze (McCallum 2004). A Library of Congress konverter szoftvereket és részletes megfeleltetési táblázatokat készített a MARC 21 és a MODS között mindkét irányban.

A MODS másik, XML-szerkezetből fakadó előnye, hogy elemei, attribútumaik segítségével alkalmassá tehetők hivatkozások elhelyezésére, így például az 520-as vagy 505-ös adatmezők olyan MODS-elemeknek feleltethetők meg, amelyek linkeket tartalmazhatnak a releváns, szóban forgó információkhoz: a leírandó dokumentum összefoglalójához, kivonatához vagy éppen a tartalomjegyzékéhez (McCallum, 2004).

A disszertáció írásának idején a 2.1 verzióban hozzáférhető **MADS** a MODS kiegészítője, „párja.” Míg a MODS a bibliográfiai adatok MARC 21-es szabványára épül, a MADS alapját az authority adatok elemkészlete képezi. A két XML-séma feladatai hasonlóan oszlanak meg, mint a tradicionális adattárolási formátumok esetében: a MADS

segítségével autorizált adatok társíthatók a MODS segítségével kódolt bibliográfiai adatokhoz. A séma alapstruktúrájának köszönhetően egy dokumentumon belül – ennek elejét és végét a `madsCollection` gyökérelem jelzi – több authority adat is leírható, egyenként a `mads` elemek közé zárva őket. A szabvány weblapja részletesen jellemzi a használható elemeket, valamint kiterjedt MARC-mappinget is közöl (Library of Congress, é.n.).

3.5. Magyarországi alkalmazások

A **BDML** Salgáné dr. Medveczki Marianna munkája. A szerző 2006-ban megvédett doktori disszertációjában vázolta fel egy olyan DTD részleteit, amely XML nyelven adja meg a szabványos bibliográfiai leírás szerkezeti sémáját. Írásában kifejti, hogy nézete szerint még nincsenek kifejezetten a könyvtári bibliográfiai leírási elvárásokhoz alkalmazkodó metaadatrendszerek, illetve azokkal, amelyek már léteznek, nem lehet olyan szintű és minőségű bibliográfiai leírást készíteni, ami megfelelné a szabályzatokban megfogalmazott elvárásoknak. Ennek pótlására dolgozta ki a BDML-t, amely amellet, hogy egyszerre támogatja a hagyományos és elektronikus dokumentumok leírását, lehetőséget ad arra, hogy a bibliográfiai leírás teljesítse azonosítási funkcióját. Az általa megalkotott XML DTD tehát *„az ISBD szabályzatokban elfogadott adatcsoportokat, adatelemeket megjelenítő, a bibliográfiai adatok előfordulás-variációinak megfelelő, kötött sorrendű elem-készletet deklarál”* (Salgáné, 2005, p. 283.). Lehetővé teszi monografikus, összefoglaló és analitikus szintű, valamint egy- és többlépcsős leírások készítését is.

A BDML jellemzői az alábbiak:

- rugalmasság, tetszőleges bővíthetőség, platformfüggetlenség, költséghatékonyság;
- a metaadatok különválasztása a megjelenítési formátumoktól;
- "önleíró" adatelemek;
- a feldolgozandó dokumentumhoz való rugalmas illeszkedés;
- különböző nyelveken írt szövegek kevert használata;
- meghatározza az elemek sorrendjét, lefekteti előfordulási szabályait (kötelezőség, ismételhetőség);
- külön elemeket definiál, ha a leírandó adatelem megengedett forrásból származik.

A rendszer három elemtípussal dolgozik. Az elemek jelentős része (`title`, `alternative title`, `edition_statement`, `dimensions`, `binding`, `price` stb.) a bibliográfiai leírás adatelemeit tükrözi, alkalmazásukra az ISBD/M, illetve a KSZ/2 szabályzatban foglaltak szerint kerülhet sor az XML-adatszerkezetben. A leírás adatszoportjait az `area1-area7` elemek veszik körbe, amelyek a leírás szerkezetét definiáló címkék közé tartoznak. A BDML talán legérdekesebb tulajdonsága, hogy a bibliográfiai leírásban alkalmazható egyezményes jeleknek megfelelő tageket is definiál, amelyet azzal indokol a szerző, hogy e jelölők külön alkalmazásával meg lehet könnyíteni a későbbi megjelenítési formátumok előállítását, és garantálni lehet a szabványos leírások készítését. Ezért minden egyes adatszoportnál megtalálhatók az abban használható egyezményes jeleket reprezentáló elemdeklarációk:

```
<!ELEMENT c_curvedbracket EMPTY>
<!ELEMENT c_char_position3 EMPTY>
<!ELEMENT colon EMPTY>
<!ELEMENT comma EMPTY>
<!ELEMENT c_squarebracket EMPTY>
<!ELEMENT diagonal_slash EMPTY>
<!ELEMENT newindent EMPTY>
<!ELEMENT o_curvedbracket EMPTY>
<!ELEMENT o_squarebracket EMPTY>
<!ELEMENT plus EMPTY>
<!ELEMENT point EMPTY>
<!ELEMENT pointspacedashspace EMPTY>
<!ELEMENT equals_sign EMPTY>
<!ELEMENT semicolon EMPTY>
```

A BDML egyike azon megoldásoknak, amelyek a három, Fiander által meghatározott típus (lásd a 3.1. fejezetet) közül a legkritikábbak közé tartoznak – a szabályzatok rendelkezéseit átültetők közé. Bár több, igen egyedi tulajdonságot is felvonultat, valamint dicséretes a következetessége és szabványhűsége, mégsem terjedt el szélesebb körben. A tervbe vett továbbfejlesztésekről – alkalmazásáról a Debreceni Egyetem Zeneműtárában, az Elektronikus Magyar Irodalom projektben, továbbá kiterjesztéséről a könyveken kívül más dokumentumtípusokra – nem érkeztek további hírek.

Egy másik projektben a Szegedi Egyetem könyvtárának munkaközössége (Bakonyi Géza vezetésével) a 2000-es évek elején egy NKA-pályázaton nyert összeg felhasználásával elkészítette a *Széphalom* című folyóirat⁶ 1927-1929 évi lapszámainak digitalizált változatát. A karakterfelismerés és a különféle szövegelemek formai egységesítése után a szövegeket pdf-formátumban tették közzé a projekt weblapján, ami azonban csak a munka első lépése volt (Bakonyi, 2000). A továbbiakban létrehozták a *Széphalom* teljes szövegű kereshető

⁶ Országos terjesztésű, jelentős szegedi lap a két világháború között.

adatbázisát is, amelynek infrastrukturális alapját – a Zebra nevű adatbázis-kezelő szoftvert – az IndexData biztosította. A kereshetőséget biztosító strukturált dokumentumpéldányokat a nyers szövegből XML-jelölők segítségével állították elő, az ehhez szükséges DTD-t maga Bakonyi Géza készítette el. A keresés grafikus felület segítségével történik, amely az adatbázissal – egy PERL szkript és a YAZ-kliens közbeiktatásával – Z39.50 szabvány szerinti kommunikációt folytat, az eredményhalmazt pedig, megfelelő formázás után, HTML-elemek felhasználásával, emberi olvasásra alkalmas módon jeleníti meg (Hegyi & Sándor, 2001).

4. Az entitás fogalmának megjelenése a könyvtártudományban

Mielőtt rátérnék a szemantikus webtechnológia könyvtári alkalmazásának történeti ismertetésére, mindenképp szót kell ejtenem az entítasalapú forrásleírási irányzatról, amelynek elméleti alapjai már a '90-es évek végén rendelkezésre álltak. Igazán széles körű, hatékony gyakorlati megvalósítására azonban úgy tűnik, még e disszertáció írásának idején is várni kell egy ideig.

A könyvtártudomány fejlődésének (közelebbről a bibliográfiai adatfeldolgozásról történő gondolkodás változásainak) alapos vizsgálata és megértése fontos következtetések levonását teszi lehetővé. Ilyen következtetés például, hogy a szakmánkban lezajlott szellemi, felfogásbeli paradigmaváltás lényegében elkerülhetetlenül vezette a könyvtárosokat a világháló korszerű adatformátumainak alkalmazásához. Ezek használata által nem csupán informatikai nyereségek mutatkoznak (például a közgyűteményi adatbázisokban tárolt információk megjelenése a keresőszolgáltatásokban, kibővített lehetőségeket nyújtó keresőfelületek, több információforrás együttes, a korábbiaktól eltérő használata a keresés során stb.), hanem lehetőség nyílik a gyakorlatba is átültetni azokat a teoretikus elképzeléseket, amelyek jelen értekezés írásának idején már csaknem 25 éve a szakmai diskurzusok részei. Megvalósításuknak azonban mindeddig súlyos technikai akadályok álltak útjába.

Az alábbiakban áttekintem a könyvtárügyet a webtechnológia alkalmazásának irányába elmozdító entítasalapú forrásleírási szemlélet fejlődéstörténetét. E betekintés célja, hogy felvázolja ennek a szemléletváltásnak a nagyságrendjét és jelentőségét, azaz világossá tegye, hogy valóban szükségzerű, és egyszersmind nagyon szerencsés egymásra találásnak voltunk, vagyunk szemtanúi, amikor a könyvtártudomány és a kortárs informatikai technológia egyre erősödő szimbiózisát vizsgáljuk.

4.1. Az FRBR megszületése

Az 1990-es évek elején megjelenő informatikai újdonságok – a World Wide Web megszületése, illetve ezzel párhuzamosan az elektronikus bibliográfiai források számának ugrásszerű megnövekedése – kihívások elé állították a könyvtárakat az egész világon. A megváltozott technikai környezet, illetve a közgyűtemények hosszú ideje tapasztalható általános forráshiánya az IFLA Katalógizálási Szekciójának figyelmét sem kerülte el. Szükségessé vált a feldolgozási munkafolyamat áttekintése, amelynek során választ kellett keresni az egyre szaporodó elektronikus bibliográfiai források kezelésének kérdésére, illetve

meg kellett találni a forrásleírás azon pontjait, ahol egyszerűsítéseket, észszerűsítéseket lehet végrehajtani a folyamat gazdaságosságának, illetve a párhuzamos munkavégzés kiszűrésének érdekében.

A testület ezért 1990 augusztusában szimpóziumot rendezett Stockholmban, amelynek fő célja az volt, hogy felülvizsgálják a bibliográfiai információkkal szemben támasztott felhasználói elvárásokat. A tanácskozáson részletesen megtárgyalták az alábbi felvetéseket (Madison, 2005):

- hol és hogyan lehet a katalógusrekord tartalmát egyszerűsíteni;
- hogyan növelhető a nemzetközi bibliográfiai adatfeldolgozás hatékonysága, illetve hogyan csökkenthetők a folyamat költségei a redundáns munkavégzés elkerülésével, a rekordok nemzeti, illetve nemzetközi szintű megosztásán keresztül;
- hogyan lehet a kialakult feldolgozási gyakorlatokat és szabályzatokat eredményesen alkalmazni elektronikus környezetekben, milyen lépéseket kell tenni az országos és nemzetközi bibliográfiai adatbázisok kialakítása érdekében;
- hogyan érdemes számot vetni az egyetemes bibliográfiai számbavétel szükségességének növekedésével, tekintettel a világon létező bibliográfiai források számának ugrásszerű növekedésére (formátumtól függetlenül);
- milyen igényeket támasztanak a bibliográfiai univerzummal szemben azok, akik azt használják.

A stockholmi szeminárium résztvevői arra a konszenzusra jutottak, hogy a bibliográfiai rekord tartalmának, adatelemeinek olyan felülvizsgálatát kell kezdeményezni, mely elsődlegesen a használók igényeit tartja szem előtt. A feladat megoldására az IFLÁ-n belül a bibliográfiai rekordok funkcionális követelményeivel foglalkozó munkacsoport (*Study Group on Functional Requirements for Bibliographic Records*) jött létre, amelynek célja egy átfogó jelentés készítése volt, a szemináriumon elhangzottak mentén kijelölt iránymutatások alapján. Ezt az iránymutatást (*Terms of Reference*) Tom Delsey, valamint Henriette Avram készítette (Coyle, 2015), majd az IFLA Katalógizálási Szekciójának Állandó Bizottsága hagyta jóvá 1992-ben (IFLA, 1998). A jelentés szövegének kidolgozása önkéntes szakemberek, hozzászólók, illetve szaktanácsadók bevonásával indult meg;

utóbbiak között található Elaine Svenonius és Barbara Tillett is, mindketten a könyvtártudomány nemzetközi szinten elismert szakemberei (Madison, 2005).

A munkacsoport – figyelembe véve az iránymutatásban megfogalmazottakat – elsődleges vizsgálati módszerként az ún. entitásanalízist választotta, elsődlegesen Peter Pin-Shan Chen, tajvani származású amerikai kutató, valamint Richard Perkinson, adatmodellezéssel és adatbázis-tervezéssel foglalkozó informatikai szakember munkásságára alapozva (Madison, 2005). Előbbi 1976-ban adta közre *The Entity–Relationship Model – Toward A Unified View of Data* című, nagy hatású cikkét, amely teljesen új szemléletet hozott az adatkezelés-adatmodellezés területére (Chen, 1976), utóbbinak pedig *Data Analysis: the Key to Data Base Design* című könyvét érdemes megemlíteni. E közlemények címeiből világosan látható, hogy az IFLA-tanulmányt készítő szakemberek alapvetően adatbázis-tervezési, tehát technikai-informatikai elvek mentén fogalmazták meg mondanivalójukat. Az FRBR-munkacsoport által végzett munkának azonban nem volt célja az, hogy kiindulópontja legyen a bibliográfiai adatbázisok tervezésének, sem annak, hogy bármiféle adatszerkezet szervezőelveként szolgáljon – erről napjainkban úgy tűnik, sokan megfeledkeznek. Csupán azért tették le a voksot az entitásanalízis mellett, mert az „*az adatkövetelmények elemzéséhez strukturált megközelítést kínál*” (IFLA, 1998, p. 17.), vagyis csupán a bibliográfiai leírásban szerepeltetendő adatelemek csoportosításának egy lehetséges keretét vázolja. Érdemes odafigyelni Barbara Tillett véleményére, aki szerint „*[a]z FRBR nem adatmodell. Az FRBR nem metaadatséma. Az FRBR nem rendszertervezési struktúra. Csupán a bibliográfiai univerzum fogalmi modellje*” (Tillett, 2005, p. 197.).

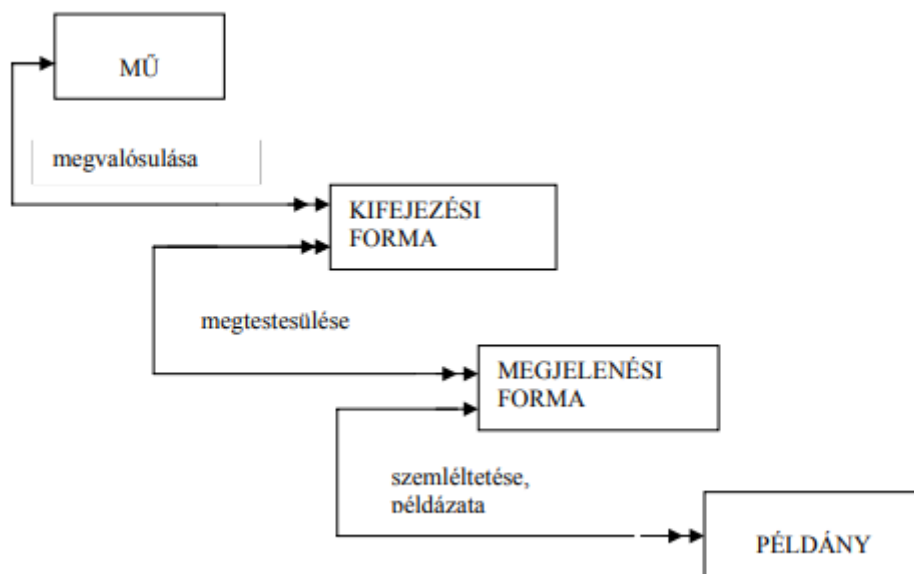
A hosszú előkészítési, valamint véleményezési szakasz után (Madison (2005) szerint negyven vélemény érkezett tizenhat országból) előállt végleges szöveget – bizottsági jóváhagyás után – az IFLA 1997 szeptemberében tette közzé. Az elkészült jelentés *Functional Requirements of Bibliographic Records* címmel látott napvilágot, s ezt a szakma a szavak első betűjéből képzett betűszó alapján FRBR-ként ismeri. (A magyar fordítás *A bibliográfiai tételek funkcionális követelményei* címen jelent meg, amely az eredeti címnél erőteljesebben emeli ki, hogy nem egy informatikai struktúra, hanem a bibliográfiai leírás szerkezetének elemzése következik.)

4.2. Az FRBR-modell felépítése

A jelentés szerzői a feladateleírásban rögzített megállapításokat követve olyan szerkezeti vázat dolgoztak ki, „*amely megállapítja, és pontosan meghatározza a bibliográfiai rekordok használói számára fontos entitásokat, minden egyes entitás ismérveit és az entitások között lévő kapcsolatok típusait.*” (IFLA, 1998, p. 11.) Ez a modell tehát annak meghatározására épül, hogy melyek azok a lényeges tematikai egységek, amelyekről az információkereső a bibliográfiai rekordból tájékozódni kíván, azaz nem a rekord egyes diszkrét adatait vizsgálja, hanem azokat a „dolgozat”, amelyeket az adatok leírnak. Ezeket a „dolgozat” nevezik *entitásnak*.

Az FRBR kidolgozói a bibliográfiai rekord adataleemeinek vizsgálatával összesen tíz entitást különítettek el, s ezeket három nagy csoportba sorolták. A modell rövid jellemzését adja e tíz entitásnak, felvázolja a közöttük lévő kapcsolatokat, illetve megadja az azokat leíró legfontosabb ismérveket, az entitások *attribútumait* is. A szakértők ezen kívül meghatározták a katalógus használata során jelentkező legfontosabb felhasználói igényeket – valamilyen entitás *megtalálását*, más, hasonló entításoktól történő megkülönböztetését (*azonosítását*), a használó kívánalmainak megfelelő entitás *kiválasztását*, illetve annak *megszerzését*, majd az entitás-attribútumok mindegyikét hozzárendelték egy vagy több igényhez (felhasználói feladathoz), annak alapján, hogy az adott adatalem mely igény(ek)e)t elégíti ki leginkább. E felhasználói feladatokat Hider (2017) elemzi részletesen tanulmányában.

A három entitáscsoport közül az első (az ún. WEMI) okozza mindmáig a legnagyobb fejtörést a könyvtárosok számára, mert megértése, logikájának belátása annak a szemléletnek a belsővé tételével kezdődik, hogy egy bibliográfiai forrásra ezentúl nem (csak) mint konkrét példányra kell tekinteni, amelynek bizonyos jellemzői vannak. A funkcionális analízis ugyanis megállapította, hogy egy forrás leírásakor az adatalemek nem egy (a *könyv* adatai, a *hanglemez* adatai stb.), hanem összesen *négy* entitásra vonatkoznak. Az ISBD szabályait követő, hét vagy nyolc adatscsoportból álló bibliográfiai leírásban megtalálható adatalemek tehát összesen négy tematikai egységbe sorolhatók az alapján, hogy *mi az, amit leírnak*.



3. ábra – Az FRBR első entitáscsoportja és elsődleges kapcsolataik (IFLA, 1998)

Elsőként el kell különíteni mindazokat az adatelemeket, amelyek egy bizonyos, diszkrét szellemi tartalomegységre, produktumra vonatkoznak – egy teljesen absztrakt, semmilyen módon nem materializálható entitásra, a *műre* (*work*). A mű nem más, mint valamilyen kitalált történet, művészi koncepció, vagy valamilyen logikai rend szerint összekapcsolt, vagy újonnan megállapított ismeretek összessége, egy sajátos, esetenként újra meg újra felbukkanó, öröklődő *mintázat*. Pettee (1936), Verona (1959), Smiraglia (2003), Holden (2020) és Wilson (1989) írásaiból tudható, hogy a művet (egyes szerzők szóhasználatában az *irodalmi egységet* [literary unit]) már Panizzi korában, a XIX. század első felében is megkülönböztették a diszkrét bibliográfiai forrástól, illetve felismerték a katalogizálásban betöltött jelentős szerepét – ez a distinkció tehát nem az FRBR leleménye. Martha M. Yee négyrészes cikksorozatát szentelt a mű mibenlétének (1994, 1995a, 1995b, 1995c), továbbá Seymour Lubetzky elméleti alapvetését vezérfonalul használva részletesen bemutatta (2000) a mű jelenlétét és legfontosabb paramétereit az egyes bibliográfiai forrástípusok esetében.

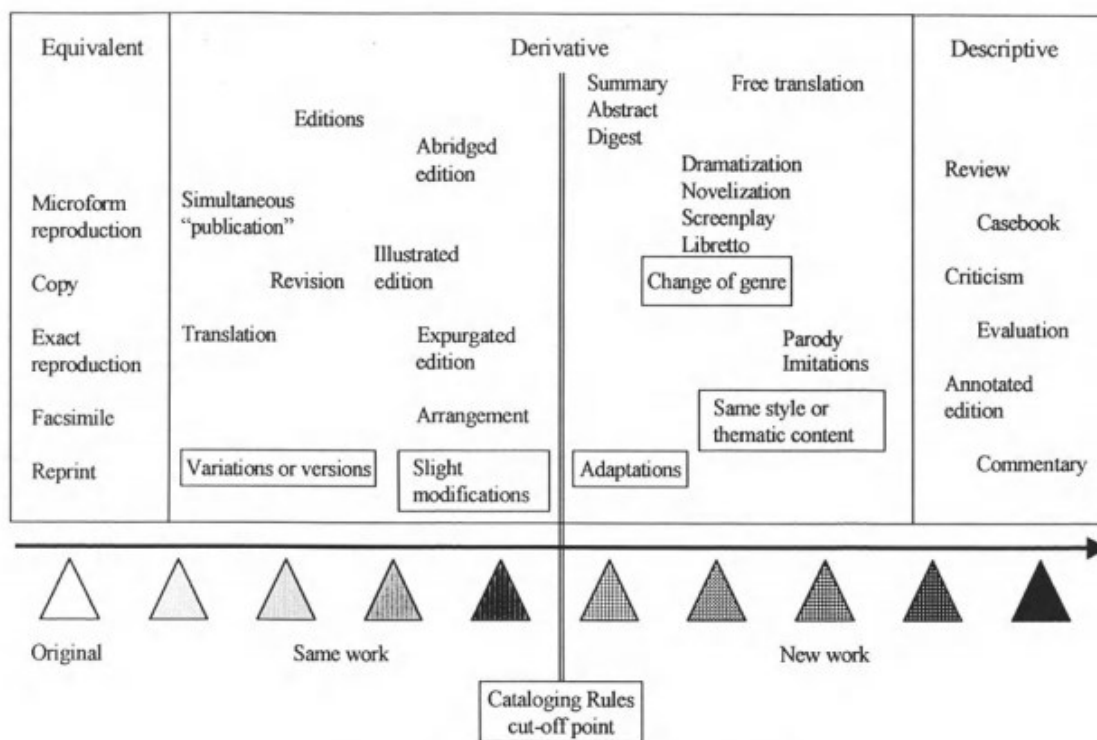
A művek absztrakcióját kívánja meg a Párizsi Alapelvekben is megfogalmazott követelmény a katalógustól: legyen alkalmas annak megállapítására, hogy egy mű mely kiadásai találhatóak meg a könyvtárban (Wilson, 1989; Lubetzky, 1957). Allyson Carlyle megállapította, hogy a referenzkérdések egy része magas absztrakciós szinten kezelendő: a használó ezekben az esetekben nem konkrét – egy adott évben, egy bizonyos kiadó által

megjelentetett – kiadásokat keres, hanem olyan kérdéseket tesz fel, mint hogy „megvan-e a könyvtárban az *Idő rövid története?*” (Carlyle, 2006). Épp ezért a mű adatainak jelenléte a bibliográfiai rekordban kiemelt fontossággal bír. Az ISBD által előírt *eredeti cím* adatelem azonban csak bizonyos esetekben, például egy mű különböző nyelvű fordításai, vagy azonos nyelvű, de különféle történeti korokból származó címváltozatok (például egy XVI. századi magyar nyelvű cím és XXI. századi átírása) esetében tudja biztosítani a bibliográfiai tételek egy helyen való megtalálhatóságát.

Az FRBR szerint a műveket – egyebek mellett – az alábbi attribútumokkal lehet jellemezni:

- a címe;
- a típusa (lényegében műfajmegjelölés);
- a keletkezésének ideje;
- a befejezés szándéka;
- a célközönsége;
- előadásának eszköze, a mű számozása, hangneme (zeneművek esetében);
- koordináták, napéjegyzetlőség (kartográfiai művek esetében).

Sorrendben a második entitás – még mindig igen magas absztrakciós szinten – a *kifejezési forma (expression)*. Ennek ismérvei (metaadatelemei) egy mű valamilyen megvalósulását írják le egy meghatározott jelölési rendszerben. Amikor tehát egy alkotó egy, a fejében lévő történetet, művészeti koncepciót, ismeretcsomagot formába önt, többféle jelrendszer, formanyelv között választhat: leírhatja magyarul, leírhatja – mondjuk – angolul, eltáncolhatja, szóban elmondhatja, kifejezheti filmes vagy színházi formanyelvi eszközökkel. Egyetlen mű különböző kifejezési formáiként kezelendők a fordítások (filmek esetében a szinkronizált vagy feliratozott változatok), illusztrált kiadások stb. Ugyanakkor új mű keletkezésének tartják a formanyelvben bekövetkező nagyobb változásokat: egy mű dramatizálását, megfilmesítését, vagy az eredeti mű paródiáját, és egyéb, különböző adaptációkat. Barbara Tillett, a bibliográfiai univerzumban leírható különféle kapcsolatok elemzése során szemléletes ábrán mutatja be a határvonalat egy mű két kifejezési formája, illetve két mű között.



4. ábra – A Barbara Tillett által meghatározott bibliográfiai kapcsolatok. Számos esetben a származék azonos műnek, más esetekben már új műnek minősül (Tillett, 2001)

Az FRBR megalkotói számoltak a művek határainak képlékenységével, amikor kijelentették: „a különböző kultúrákban vagy nemzeti csoportokban lefektetett bibliográfiai szabályok eltérők lehetnek annak meghatározásában, hogy hol a határ az egyik és a másik mű között.” (IFLA, 1998, p. 24.)

A kifejezési formát leíró főbb adatelemek az alábbiak:

- a kifejezési forma címe;
- műfaja;
- keletkezési ideje;
- nyelve;
- verziójelölése;
- méretaránya, vetülete;
- kottatípusa;
- előadásának eszköze (eltérhet a mű hasonló adatától).

A kifejezési formákkal kapcsolatban az Associazione Italiana Biblioteche katalogizálási munkacsoportja kritikát fogalmazott meg, mely szerint az entitást további alcsoportokra kellene bontani az alábbiak szerint (LeBoeuf, 2001):

1. a mű első, eredeti kifejezési formája;
2. az eredetihez típusban hasonló kifejezési formák;
3. azonos médiummal, de más kódrendszerrel elkészülő kifejezési formák;
4. kifejezési formák más médium használatával,

de mások is nemtetszésüket fejezték ki az entitással kapcsolatban. Edward O'Neill (2002) szerint „*a kifejezési formák azonosítása problematikus, és felveti a kérdést, valóban valid entitások-e*” (p. 158), más szerzők (Smiraglia, 2012) pedig úgy vélik, egész egyszerűen hibás lépés volt a mű és a kifejezési forma adatainak szétválasztása.⁷

A harmadik entitás a *megjelenési forma (manifestation)*, amely egy kifejezési forma valamilyen egyedi fizikai megtestesülése: egy szövegszerű kifejezési forma valamilyen meghatározott kiadása, egy zenei előadás felvételének megjelenései különféle hordozókon (hanglemezen, CD-n). Ez az entitás, avagy absztrakciós szint „*valamennyi fizikai tárgyat [tehát konkrét példányt – H.M.] képviseli, amelynek azonosak a jellemzői mind a szellemi tartalom, mind a fizikai forma tekintetében.*” (IFLA, 1998, p. 28.). A jelenlegi katalogizálási gyakorlatnak, a bibliográfiai rekord jelenlegi kialakításának, de magának az FRBR-nek is ez az entitás áll a középpontjában.

Az FRBR a megjelenési formák összesen 38 jellemző elemét állapítja meg, amelyek közül kiemelhető:

- a megjelenési forma címe;
- a felelősségi közlés⁸ (a tanulmányban e közléseket kizárólag a megjelenési forma adataiként sorolják fel, de természetesen megkülönböztethetők olyan felelősségi közlések, amelyek a műre (például a szerző), megint mások pedig a kifejezési formára (például a fordító) vonatkoznak. Taniguchi (2002) ezt a jelenséget nevezi

⁷ Az értekezés 9. fejezetében mutatom be, hogy a BIBFRAME bibliográfiai keretrendszer a művek és a kifejezési formák metaadatelemeit egybeolvasztja a Library of Congress által az FRBR-hez készített, de végül nem implementált javaslat alapján (Madison, 2005). A művek és kifejezési formák közötti határvonal az 5. fejezetben ismertetett RDA szabályzatban sem húzható meg egyértelműen, a szabályzat fejlődése során több, a műhöz, illetve a kifejezési formához tartozó metaadatelemet átcsoportosítottak.

⁸ A dolgozatban a *felelősségi* közlés fogalmat használom a *szerzősségi* közlés helyett: ennek magyarázatát az 5. fejezetben adom meg.

felfelé történő pszeudo-hozzárendelésnek (*upward pseudo-assignment*), ti. amikor egy entitás attribútumai más entításokra is vonatkoztathatók);

- a kiadásjelzés;
- a megjelenés helye és dátuma;
- a kiadó/terjesztő/előállító/gyártó;
- a sorozatra vonatkozó közlés;
- a hordozó típusa, mérete, terjedelme, anyaga;
- azonosító (például ISBN).

Shoichi Taniguchi idézett írásában olyan adatmodellt javasol – napjaink katalogizálási gyakorlatával, de az FRBR-rel is szembehelyezkedve –, amely a megjelenési forma helyett a kifejezési formának tulajdonít kiemelt fontosságot. Érvelésében kifejti, hogy számos attribútum, mely a megjelenési formához kapcsolódik az FRBR-ben, valójában a magasabb szintű entításokkal – művel és kifejezési formával – kapcsolatos használói igények kielégítését célozza – az FRBR például a megjelenési formáról leolvasott főcímet a *mű* címeként aposztrofálja (amennyiben a leírásban nem alkalmaznak egységesített címet). Taniguchi hozzáteszi, hogy a kifejezési forma számos attribútuma, bár logikai szinten elkülönül, valójában a gyakorlat szempontjából üres, hiszen nem feleltethető meg semmilyen, az ISBD-ben megállapított adatelemnek. Taniguchi ezért tanulmányában az attribútumok átstrukturálásával teszi kiemelkedővé a Kifejezési forma entitásszintjét, míg a megjelenési formát ún. „gyenge” entitásként írja le. Állítása szerint a címadatok, valamint a felelősségi közlések (továbbá a kiadási információk) kifejezésiforma-szintű attribútummá tétele semmilyen problémával nem jár.

Érdekes problémát vet fel írásában Caryle (2006), amikor arról elmélkedik, hogy a Braille-írással készült szövegváltozatok kifejezési formának vagy megjelenési formának minősülnek-e. Érvelése szerint a különféle fordításoknak is lehetnek egyenként Braille-változatai, és így megjelenési formának is értelmezhetők, mégis helyesebb, és a használók szempontjából logikusabb, ha egy sajátos formanyelvet igénybe vevő kifejezési formaként kerülnek leírásra. Ez utóbbi megfontolás alapján nyilváníthatók kifejezési formának azok az esetek is, amelyekben egy művet jeltolmácsolással is előadnak.⁹

⁹ A későbbiekben ismertetendő Könyvtári Referenciamodell (LRM) akkor javasolja új mű rögzítését, ha az eredeti mű átalakítása jelentős mértékű önálló szellemi vagy művészi munkát jelent. Az egyik formanyelvről, jelrendszerről a másikra ültetés – angolról magyarra, dánról norvégira, textuálisról Braille-formátumra vagy éppen jelnyelvre – úgy tűnik, nem tartalmazza ezt a hozzáadott értéket.

Az FRBR első csoportjának negyedik entitása, amely már kézzel is megfogható, a *példány (item)*. Sajátos adatai közé tartozik többnyire valamilyen egyedi azonosító (mint például a könyvtár által előállított vonalkód, amely a kölcsönzés munkafolyamatát segíti), a possessorbejegyzés, állapot, kölcsönözhetőségi információk stb. Ezeket az információkat napjaink integrált gyűjteménykezelő rendszereiben példányrekordok, vagy egymástól elkülönülő holding- és példányrekordok írják le, amelyek a bibliográfiai rekordhoz csatlakoznak.

Az FRBR által megkülönböztetett *második*, illetve *harmadik* entitáscsoportba a fenti négy entitás konkrét előfordulásaiért felelős személyek, illetve testületek kerültek (2. csoport), míg a harmadik csoport azokat az entitásokat tartalmazza, amelyek művek tárgyául szolgálhatnak. Szólhatnak művek valamilyen elvont fogalomról (emlékezet), valamilyen konkrét tárgyról (Márai Sándor írógépe), egy eseményről (a második világháború), vagy valamilyen földrajzi helyről (Líbia); művek témáját adhatja továbbá személy, illetve testület, illetve az is lehetséges, hogy egy mű egy másik műről, kifejezési vagy megjelenési formáról, esetleg példányról szól (IFLA, 1998).

Az FRBR összesen tíz entitásból álló szerkezete természetesen hosszú viták, egyeztetések során alakult ki. A tervezési dokumentumok arról tanúskodnak, hogy 1993 elején még létezett az ún. szupermű (superwork) entitás, amelyet egy későbbi ponton gyűjteménynek (aggregate) neveztek el, végül a későbbiek során ezt az entitást elvetették.¹⁰ Kiemelt jelentőséggel bír az a hozzászólás, amelyet a Library of Congress munkatársai tettek a készülő fogalmi modellhez: 1994-es keltezésű munkadokumentumukban ők javasolták a *reprezentáció* helyett a *kifejezési forma* megnevezést, a *termék* (product) helyett pedig a *megjelenési formát*. Az eredetileg külön entitásként létező *hordozó* (carrier) adatait beleépítették a *megjelenési forma* adatai közé. A washingtoni könyvtárosok saját entitás-kapcsolat modellt alakítottak ki, amelyben az FRBR első csoportjának négy eleme helyett mindössze három entitást – a művet, a megjelenési formát, továbbá a példányt – határoztak meg (Madison, 2005). Ez a három entitás adja a 2012-ben kidolgozott BIBFRAME alapszerkezetét.

¹⁰ A kortárs adatmodellezési kísérletekben - így a BIBFRAME-ben is - ismét megjelenik egy ilyen, a műnél is magasabb absztrakciós szint. A szuperműnek is nevezett entitásról Yee (2000), illetve Carlyle (1996) is említést tesz.

4.3. Megszületik a Könyvtári Referenciamodell

Az FRBR készítésével megbízott munkacsoport elsődlegesen a bibliográfiai rekord adatalemeit vizsgálta. A forrásleírás munkafolyamata azonban nem csupán a forrást azonosító ismérvek, hanem a leírás hozzáférési pontjainak rögzítését is tartalmazza, ezt a könyvtárosok egységes szerkezetű, ún. authority adatok rögzítésével és a bibliográfiai rekordhoz történő hozzárendelésével végzik el a Nemzetközi Katalogizálási Alapelvekben foglaltak figyelembevételével. A leírásban szereplő ágensek, földrajzi nevek, valamint címek egy-egy keresési szempontot jelentenek, amelyek szabványosan előállított, egységesített névalakját korábban a cédulák tételfejei, ma az integrált rendszerekben tárolt authority rekordok kitüntetett hozzáférési pontjai tartalmazzák.

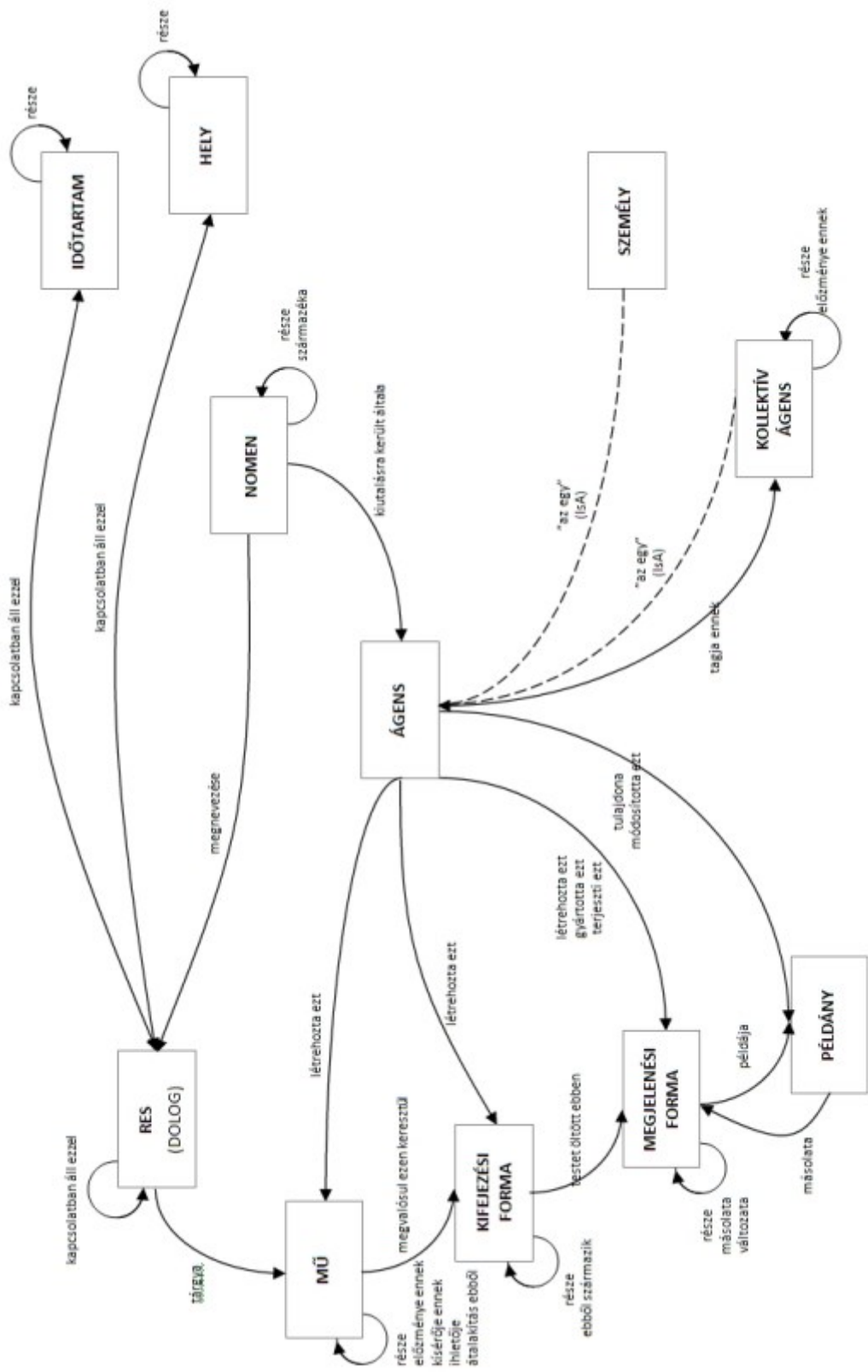
Már az FRBR megalkotása közben is felmerült, hogy a bibliográfiai univerzum átfogó vizsgálata érdekében az authority rekordokra is ki kellene terjeszteni az entitásanalízis látókörét, azonban a vizsgálatra fordítható idő korlátozott volta miatt ez elmaradt. Ezeket az elemzéseket csak évekkel később végezték el, az eredményeket pedig 2004-ben, illetve 2010-ben publikálták. Az entitás-analízisre épülő három fogalmi-elméleti modell (FRBR, FRAD, FRSAD) a maga területén jól megoldotta feladatát: feltérképezték a bibliográfiai, authority és tárgyi authority rekordok működését és a velük szemben támasztott használói elvárásokat. Azonban, mivel különböző időpontokban, az informatikai látókör szélesedésének más-más pontján, különböző munkacsoportok erőfeszítései nyomán készültek, nem egységes a fogalomhasználatuk, tartalmi-szerkezeti felépítésük egymáshoz viszonyítva nem következetes (Dudás, 2017). Szükség volt tehát egy olyan dokumentum elkészítésére, amely összehangolja a három, entításokon és kapcsolatokon alapuló fogalmi modellt és kiküszöböli a köztük lévő lényegi inkonzisztenciát.

Így született meg 2017-ben – nagyjából hét évig tartó munka után – a Könyvtári Referenciamodell, amely jelentős változásokat hozott a bibliográfiai univerzumba (Riva, LeBoeuf & Žumer, 2017). Per definitionem, a referenciamodell egy absztrakt keretrendszer, amely felvázolja egy adott környezet entitásait, illetve az azok között lévő kapcsolatokat, s fontos jellemzője, hogy nincsen semmiféle kötődése bármilyen szabványhoz, technológiához, sem azok valamilyen konkrét megvalósulásához (Riva, 2018). Erre a meghatározásra alapozva a Könyvtári Referenciamodellnek, akárcsak az FRBR-nek, a fő célja az entítások elkülönítése, kapcsolataik leírása, illetve a használók információkeresésének szempontjából fontos ismérvek meghatározása (Dancs, 2018a)

A Könyvtári Referenciamodellben leírt legtöbb entitás már megjelent a három korábbi fogalmi modellben – a definíciókban mutatkozó különbségek összevetését Seikel és Steele (2020) végezte el –, azonban az LRM-ben ezek az entítások hierarchikus szervezőelv mentén csoportosulnak. Pat Riva (2016) szerint ez azért szükséges, hogy csökkenjen a redundancia a rendszerben: a magasabb szintű osztályokra definiált attribútumok és kapcsolatok a hierarchikus megközelítésnek köszönhetően automatikusan érvényesek az alárendelt osztályokra is.

A tizenegy entitást tartalmazó hierarchia csúcsán az ún. dolog (*res*) entitás áll – ez gyakorlatilag korlátozás nélkül mindent magában foglal, ami releváns a bibliográfiai univerzumban. Alárendeltjei közé tartoznak az FRBR első entitáscsoportjának elemei – mű, kifejezési forma, megjelenési forma, példány –, az ún. ágensek, azaz olyan személyek, illetve kollektív ágensek, amelyek az előbb felsorolt entításokhoz valamilyen módon kapcsolódnak, azaz művet, kifejezési formát, vagy megjelenési formát hoznak létre, illetve birtokolnak egy egyedi példányt. A „dolgokhoz” külön entitásként hozzárendelhetők továbbá bizonyos időtartamok, időpontok (például ekkor készült egy kifejezési forma, ekkor élt egy személy, erről az időszakról szól egy mű stb.), illetve bizonyos földrajzi helyek (például egy megjelenési forma egy bizonyos helyen készült, egy testület egy bizonyos helyen működik, vagy egy műnek egy bizonyos hely adja a témáját).

Az FRAD, illetve az FRSAD elemeinek összefésülésével új, *nomen* nevű entitást definiál a Könyvtári Referenciamodell. A *nomen*nek egy entitás egyetlen, meghatározott előfordulására vonatkozó, azzal társított különféle megnevezései, névformák, amelyek típusukkal, értékükkel (valamilyen konkrét karakterlánc, string), nyelvükkel/írásrendszerükkel, valamint sémájukkal jellemezhetők (például ISBN-séma: 10 vagy 13 számjegy). A *nomen* entitás egyes konkrét előfordulásai valamely *res*-előforduláshoz kapcsolódnak a „megnevezése” kapcsolattal, így tehát a *nomen* egy *res* valamilyen kontextusban használt *megnevezése*. Minden *nomen* egy és csak egy *res*-re vonatkozhat, ugyanakkor egy *res*-nek több különböző *nomen*ja is lehet, amelyek eltérhetnek egymástól, például ugyanazon tudományterületet jelölő természetes nyelvű fogalom (például „könyvtártudomány”), illetve annak raktári szakjele („020”) (Riva, 2018).



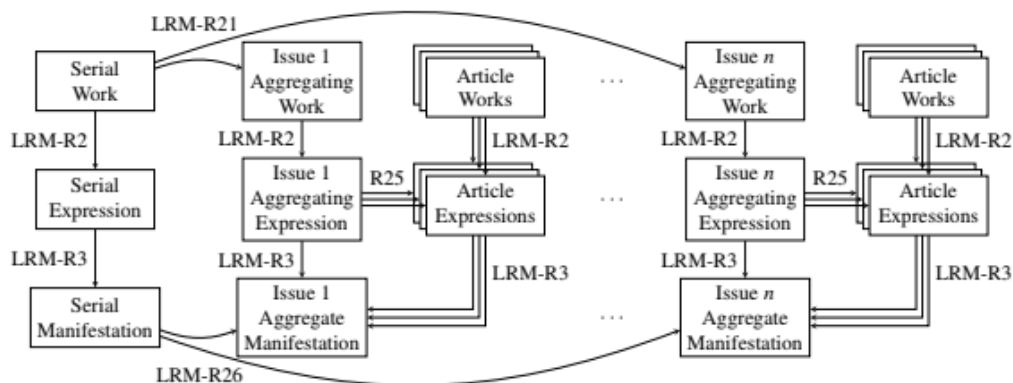
5. ábra – Az LRM-ben definiált 11 entitás és kapcsolataik (Riva, LeBoeuf & Žumer, 2017)

Az entitásszerkezet mellett természetesen megváltoztak az entitások tulajdonságai, attribútumai is. A mű jellemzői között megjelenik egy sajátos ismérv, a *reprezentatív kifejezési forma*: ez egy mű kitüntetett, „kanonizált” formája, amely a használók szemszögéből nézve legjobban jellemzi az adott művet. Az LRM-ben leírtak szerint „*a használók ösztönösen megértik, hogy William Shakespeare Hamletje az angol nyelvhez kötődik, és irodalmi műfaja szerint színdarab.*” (Riva, LeBoeuf & Žumer, 2017, p. 88.) Bármely más kifejezési forma – rövidített változat, fordítás stb. – eltávolodást jelent az eredeti kifejezési formától, amelyet ezért kiemelkedő jelentőséggel kell kezelni. A reprezentatív kifejezési forma megállapítása mindenkor sajátos kulturális ismereteken alapul. (Az LRM szövegében egyébiránt ezen a ponton megjelenik a korábban, Taniguchi nyomán felfelé történő pseudo-hozzárendelésnek nevezett szellemi aktus. Mivel a használók bizonyos jellemzőket a hétköznapi megfogalmazásban általában magához a műhöz társítanak (a Rómeó és Júlia Shakespeare *drámája*), ezért a kifejezési forma egyes attribútumai „*elméletben »átruházhatók« a műre, és a mű azonosítására alkalmazhatók, bár szigorúan véve ezek az ismérvek a kifejezési formát jellemzik és nem a művet.*” (Riva, LeBoeuf & Žumer, 2017, p. 88.).

A Könyvtári Referenciamodell röviden kitér a gyűjtemények és az időszaki kiadványok modellezésére, azaz az entitásszerkezetben történő leképezésükre is. A gyűjtemények három típusát *forrásgyűjteményben, bővülésből eredő forrásgyűjteményben, illetve párhuzamos kifejezési formák gyűjteményében* határozza meg az LRM, és úgy tekint rájuk, mint aggregátum megjelenési formákra, amelyek a gyűjteményt alkotó egyes művek kifejezési formáiból tevődnek össze. A forrásgyűjteménybe – azaz típusban vagy műfajban valamilyen hasonlóságot mutató kifejezési formák összességébe – értendők a válogatások, antológiák, monografikus sorozatok, önálló fejezetekből álló könyvek, időszaki kiadványok lapszámjai (cikkek gyűjteményei), illetve más hasonló bibliográfiai forrás-csoportok. A bővülésből eredő forrásgyűjteményt általában egyetlen önálló mű, és a hozzá társuló, alárendelt művek alkotják (előszó, illusztráció, jegyzetapparátus stb.), míg a párhuzamos kifejezési formák gyűjteményeire a legjobb példák a többnyelvű kiadványok (nyomtatottak, de akár több szinkronnal/felirattal forgalomba hozott DVD-filmek is). Ez utóbbi esetben az egyetlen megjelenési forma több, párhuzamos kifejezési formát tartalmaz.

Ehhez társul az egymás mellé helyezés szellemi koncepcióját jelentő ún. aggregáló mű, illetve annak kifejezési formája – ennek segítségével válnak leírhatóvá a gyűjtemény megalkotásában részt vállaló közreműködők és a rájuk vonatkozó felelősségi közlések. A

gyűjtemény (aggregátum megjelenési forma) modellezésekor tehát a befoglalt művek kifejezési formáin túl mindig eggyel több kifejezési formával kell számolni: ez utóbbiban valósul meg az aggregáló mű.



6. ábra – Időszaki kiadványok modellezése az LRM-ben (Riva, 2018)

Az időszaki kiadványok esetében szintén az aggregáció fogalmára kell tekintettel lenni. Az első szinten az időszaki kiadvány minden egyes lapszáma mögött aggregáló koncepció (tehát mű) rejlik, amely az adott részegységben megjelenő cikkeket kifejezési formákként gyűjti egybe, hozzáadva az aggregáló mű kifejezési formáját – ezzel születik meg az aggregátum megjelenési forma.

Magasabb absztrakciós szinten az LRM bevezeti az *időszaki kiadvány mű* (serial work) fogalmát – ez tehát az egész időszaki kiadvány átfogó koncepcióját, tartamát jelenti, amely ihletőjéül (LRM-R21) szolgál az egyes lapszámok koncepcióját jelentő aggregáló műveknek. Az időszaki kiadvány műnek egy és csak egy kifejezési formája, annak pedig egy és csak egy időszaki kiadvány megjelenési formája van (ezt a megkötést az 5. fejezetben ismertető RDA katalogizálási szabályzat WEM-zárnak nevezi (RDA Toolkit, é.n.), ennek részeit (LRM-R26) alkotják az egyes lapszámok önálló aggregátum megjelenési formákként.

Az LRM-ben megfogalmazottak szerint egy többféle kiadásban megjelenő időszaki kiadványt két időszaki kiadvány műként, és nem egy időszaki kiadvány mű két kifejezési formájaként kell modellezni. Ez a helyzet például akkor, ha egy időszaki kiadvány két regionális kiadásban jelenik meg: az időszaki kiadvány mű ebben az esetben is a szerkesztési elv, az átfogó koncepció, amely éles határvonalat von a két időszaki kiadvány mű közé. De akkor is két időszaki kiadvány műről, s nem két megjelenési formáról kell beszélni, ha az egyes részegységek nyomtatott, illetve online formában is megjelennek, s tartalmuk teljesen

azonos (Riva, LeBoeuf & Žumer, 2017). Ennek magyarázata, hogy nem tekinthető evidenciának, hogy a két változat közötti kapcsolat az idők folyamán mindenkor változatlan marad, ezért – érvelnek az LRM megalkotói – helytelen lenne csupán egy időszaki kiadvány mű két kifejezési formájaként értelmezni őket. Az időszaki kiadvány művek között fennálló számos kapcsolattípus (mű-mű kapcsolatok) modellezése az LRM-nek nem célja, ezt más modellek, például a PRESSoo hatáskörébe utalják a készítők.

4.4. Kapcsolat a múzeumi világgal – az FRBRoo

A XX. század utolsó éveitől – egészen pontosan 1996-tól (LeBoeuf, 2012a) – kezdve hasonló fogalmi modellezési munkát végeztek a másik nagy közgyűjteményi ágazatban, a múzeumoknál is. A múzeumi műtárgyak leírásának átfogó elméleti modelljét, amely egyszerűen a Fogalmi Referenciamodell (Conceptual Reference Model) nevet kapta, a Múzeumok Nemzetközi Tanácsának nemzetközi dokumentációs bizottsága (ICOM-CIDOC) alkotta meg. A CRM 2006-ban ISO-szabványként látott napvilágot (LeBoeuf, 2012b), előzetes munkaanyagként pedig már 1998-ban olvasható volt. A dokumentum formába öntését 2000-től kezdve a CIDOC által létrehozott CRM munkacsoport végezte (LeBoeuf, 2012a).

A CIDOC-CRM modell központjában az *esemény* nevű entitás áll, amelyhez ágensként elsődlegesen valamilyen *természetes* vagy *jogi személy* (a CRM definíciója szerint *aktor*) kapcsolódik, s ezek az ágensek többféle *néven* ismeretesek. Az eseménynek három fő típusa van: valaminek a *kezdeté*; valaminek a *vége*, illetve az ettől a kettőtől a tervezettség, szándékoltság tekintetében különböző *cselekvés*. Az esemény egy meghatározott *időbeli* és *térbeli* ponton megy végbe, s ezek a pontok *dátumok*, illetve *földrajzi helynevek* segítségével írhatók le. Az eseményhez kapcsolódnak a különféle *tárgyasult*, illetve *absztrakt alkotások*, amelyeknek szintén számos *megnevezése* létezhet.

A múzeumokban gyűjtött kulturális örökségi objektumok – a könyvtárakkal ellentétben – minden esetben egyedi darabok, s ezért nem kezelhetők úgy, mint például egy nyomtatott könyv kivitelileg megegyező példányai. LeBoeuf (2012b) szerint a múzeumi gyűjtés e sajátosságából következik, hogy ebben a közgyűjteményi ágazatban a könyvtáraknál jóval lassabban fogalmazódott meg a szabványosítási igény – tekintve, hogy minden egyes múzeum a saját gyakorlatát igyekezett fejleszteni – s így ezeket az intézményeket nem kötik olyan, mára már egyre inkább elavultnak tekinthető eszközök és technológiák, mint az ISBD vagy a MARC.

Az FRBR publikálása utáni években felmerült a gondolat, hogy a könyvtári, valamint a múzeumi fogalmi modellt harmonizálják. Az együttműködés fontosságáról ugyancsak LeBoeuf (2012a) fogalmazott meg fontos gondolatokat, aki szerint a két közgyűjteménytípusban őrzött kulturális javak leírását egy közös fogalmi modellnek kell szabályoznia, mivel a könyvtári és múzeumi gyűjtemények egyáltalán nem diszjunktak: a könyvtárak őrizhetnek – és sok esetben valóban őriznek is – egyedi kéziratokat, a múzeumok pedig több, kereskedelmi forgalomba szánt, több példányban létező objektumot, mint amilyenek a fényképnymatok. De gondolhatunk arra is, hogy az egyes közgyűjteményi objektumok tartalmukban utalnak egymásra: egy könyv leírhat például egy ókori egyiptomi műtárgyegyüttest, egy festmény pedig ábrázolhat egy irodalmi szerzőt, vagy annak egy nevezetesebb karakterét.

Az FRBR 1997-es, a CIDOC-CRM 1998-as (előzetes) megjelenését követően elég későn, csupán 2003-ban kezdődött meg a harmonizációs munka. A kiindulópontot a múzeumi közösség modellje jelentette: ennek terminológiai keretei közé kellett elhelyezni a könyvtári elméleti dokumentumban leírtakat. Az FRBR, mint korábban utaltam rá, entitások és kapcsolataik rendszereként definiálta a bibliográfiai univerzumot, a CIDOC-CRM ezzel ellentétben az objektumorientált megközelítést követte: osztályokat és tulajdonságokat határozott meg. Ezért az FRBR-ben megfogalmazott attribútumokat osztálytulajdonságokká kellett alakítani, így kiegészítve a CIDOC-CRM-ben kialakított fogalmi struktúrát. Az eredmény egy objektumorientált és jelentősen kibővített FRBR lett, amelyet az eredeti, entitás-kapcsolat modellre alapozó változattól (FRBR_{ER}) eltérően, FRBR_{OO}-nak nevezték el. Az előkészítés hat évig tartott, majd 2009-2012 között a második verzióon dolgoztak a szakemberek, amely már nem csupán az FRBR, de az FRAD és az FRSAD legfontosabb megállapításait is magába foglalta (IFLA, 2015). Az FRBR_{OO} 3.0 verziója 2017 októbere óta olvasható.

5. Az RDA forrásleírási szabályzat megjelenése és fejlődése

A lassan átalakuló katalogizálás-elméleti környezet gyakorlatba való átültetése még éveket váratott magára. Csupán a 2000-es évek második felében határozták el az angol-amerikai nyelvterületen érvényben lévő AACR2 katalogizálási szabályzat karbantartásával foglalkozó szakemberek, hogy az új elméleti irányzat fényében ismét áttekintik, és ahol szükséges, módosítják az 1970-es évek közepe óta érvényben lévő második kiadást. Az eredetileg AACR3 „munkacímen” emlegetett szabályzat jóval tágabb szemlélettel igyekszik lefedni a dokumentumkezelés egyes területeit, mint elődei, ezért az addig *Angol-amerikai katalogizálási szabályok* című összeállítást az általánosabb célokra tekintő, *Forrásleírás és hozzáférés* néven kezdték emlegetni (Dudás, 2012). Érdeemes megemlékezni róla, hogy éles kritikát is kapott: a könyvtárügy öt új törvényének megalkotójaként is ismert Michael Gorman a fejlesztés igen korai szakaszában úgy vélte, a munkát „tudatlanságból, újdonságvágyból és a gyakorlat kárára történő elméletmagasztalásból kevert bájital” hatása irányítja (Gorman, é.n.).

Az AACR átdolgozása 2005-ben kezdődött, a teljes tervezet – immár RDA-ként – a szakmai egyeztetések és az érkezett javaslatok beépítése után 2010-ben látott napvilágot. Az ezt követő időszakban több nemzeti implementációs projekt is kezdetét vette, továbbá megindult a szabályzat szöveges tartalmainak, illetve szótárainak és szótárkódolási sémáinak nemzeti nyelvekre történő fordítása.

Az RDA, bár a szakmai diskurzusban általában egyszerűen entitásközpontú katalogizálási szabályzatként hivatkoznak rá, valójában több ennél. Hivatalos definíciója alapján az RDA ugyanis a metaadatelemek, illetve az azok használatára vonatkozó előírások komplex készlete, amelyek szavatolják a könyvtári, illetve kulturális örökségi források metaadat-leírásainak megfelelő minőségét, illetve a linked data-alapú rendszerekben történő felhasználhatóságát (RDA FAQ, 2022). Ugyan e meghatározásban lényegi elemként szerepel a linked data-technológiára való építés (e technológia alapjairól lásd az értekezés 6. fejezetét), bevezetését más technikai környezetekben is meg lehet valósítani. Dancs Szabolcs (2020) alapján ezek a következők:

- relációs/objektumorientált adatok, ahol az entitásokat, kapcsolataikat, metaadatelemeiket, valamint azok értékeit megfelelően szerkesztett adattáblákban fejezik ki, s tekintetbe veszik az alkalmazandó VES-eket és SES-eket is;

- bibliográfiai/authority adatok, amelyek a mai gyakorlathoz a legközelebb állnak. A bibliográfiai információk egy része a több entitás metaadatelemeiből felépülő bibliográfiai rekordban kap helyet, más entítások jellemzése pedig az ezekhez kapcsolódó authority rekordokban történik;
- lapos fájl/strukturálatlan adatállomány.

Az RDA-ról szóló szakirodalom bőséges és szerteágazó. Kialakulásának, fejlesztésének és tesztelésének történetével mások mellett Morris és Wiggins (2016), Tillett (2016), Panchyshyn & Park (2015), valamint Wacker, Han & Dartt (2011) foglalkozott. Az elméleti modellekkel való kapcsolatát Riva & Oliver (2012) értékelte; a tájékoztató szolgáltatásra gyakorolt hatását Keenan (2014) és McCutcheon (2012), valamint Miller (2011) elemezte, az ISBD-vel való összefüggéseit Rodríguez (2016) tárgyalta. A magyarországi szerzők közül meg kell említeni Dancs Szabolcs írásait az FRBR-, illetve RDA-alapú katalogizálásról és a forrásfeldolgozás jövőjéről (2017a), a szabályzat európai implementációs megoldásairól (2017b), továbbá az RDA szerkezetének a 3R projekt során történő alapos átdolgozásáról (2018b). Dudás Anikó *Nemcsak weben lenni, hanem webből lenni* (2013) című írásában a forrásleírás átalakuló informatikai hátterének tárgyalása mellett az RDA-ra is kitér, *Forrásleírás és hozzáférés* (2012) című tanulmányát pedig egészében a szabályzat bemutatásának és kritikájának szenteli. Vass Johanna (2017a, 2017b) kétrészes cikksorozatban tárgyalja az RDA kapcsolatát a nemzetközi könyvtári modellekkel és elemkészletekkel, köztük az FRBR₀₀-val.

A 2010-es években egyre gyorsuló ütemben terjedt el az RDA használata, nem csupán az angolszász országokban, de Európa más területein is. Regionális koordinációs testületek alakultak: a kontinensen ezeket a feladatokat az EURIG (European RDA Interest Group) látja el, a NARDAC tevékenysége Észak-Amerikára, az ORDAC-é pedig Óceániára terjed ki. Az Országos Széchényi Könyvtár által koordinált, intézményközi RDA szakértői munkacsoportnak, a Könyvtári Szabványosítási Irodának, valamint a Petőfi Irodalmi Múzeum Digitális Bölcsészeti Központjának szakmai iránymutatásával Magyarország is megkezdte az áttérési munkálatokat: a könyvtárosok fokozatos megismertetését az entításalapú katalogizálás elméletével és gyakorlatával, illetve az RDA komponenseinek magyar nyelvre fordítását, és az új kifejezések meghonosítását a hazai szakmai diskurzusban.

A 2017. évben a Könyvtári Referenciamodell megjelenésének előkészítéséhez kapcsolódva az RDA eredeti, angol nyelvű változatának fejlesztését felfüggesztették (Dancs, 2018b). Ennek magyarázata, hogy a készítők az RDA alapos átstrukturálását vették tervbe az LRM miatt, hiszen az új fogalmi modellben jelentős mértékben megváltozott a bibliográfiai univerzum entitásstruktúrája. Új fogalmak, új modellezési alapelvek jelentek meg, s több esetben, ha csak kis mértékben is, de megváltoztak az entítások definíciói is. Ezt az átalakítási projektet 3R (RDA Restructure and Redesign) néven ismeri a szakmai közösség. A munkálatok 2020 végén fejeződtek be, a jelenleg érvényben lévő rendelkezések előfizetés birtokában a <https://access.rdatoolkit.org/> címen megtekinthetők.

A szabályzat aktualizált szövege minden, az LRM-ben definiált entitással foglalkozik – azaz közli az egyes entításokhoz tartozó metaadatelemek leírását, valamint javaslatokat fogalmaz meg azok rögzítési módjára vonatkozóan. Egyes metaadatelemek értékei csak kötött listákból – ún. szótárkódolási sémákból – származhatnak, ezeket az RDA Registry közli (bővebben a 6.5.3. fejezetben szólok róla), és a nemzeti fordítási tevékenységek kiemelt részét képezik. Más értékek bevitelét pedig az ún. karakterlánc-kódolási sémák szabályozzák, amelyek bizonyos formai követelményeket határoznak meg egy metaadatérték szerkezetével kapcsolatban (csak négy számjegyből állhat, csak szöveges értéke lehet stb.) A szabályzatokban meghatározott entításkapcsolatokhoz, illetve metaadatelemekhez tartozó azonosítókat ugyancsak az RDA Registry tartalmazza.

A szabályzat által meghatározott, leírandó metaadatelemek egy részének megadása kötelező, más részük azonban opcionálisan rögzíthető – intézményi vagy magasabb szintű megállapodás kérdése, hogy a forrásokról leírható, az RDA által meghatározott metaadatok közül melyek kerülnek bele ténylegesen a leírásokba. Ezeket az egyes forrástípusokra vonatkozó megállapodásokat az adott nemzet vagy intézmény *RDA-alkalmazási profilja (AP)* rögzíti¹¹, amely tartalmazza a leírásokban alkalmazható metaadatelemek leírását, azok hierarchia-viszonyait, esetenként azt, hogy az egyes elemekhez tartozó értékek generálhatók-e más metaadatelemek értékeit felhasználva. Az alkalmazásprofil tartalmazza továbbá az elemek legfontosabb tulajdonságait, kötelezőségét, ismételhetőségét, illetve rögzíti, hogy az elemek értékének megadását milyen formában várjuk el, szöveges karakterláncként, vagy valamilyen azonosító alakjában, esetleg megköveteljük valamilyen

¹¹ Az alkalmazási profilokról a későbbiekben – más kontextusban – még ejtek szót, az értekezés 7. fejezetében.

meghatározott szótárkódolási séma értékeinek használatát. Megjegyzéseket tartalmazhat továbbá a karakterláncként megadott értékek formátumára vonatkozóan.

A fejezet rész végén érdemes kitérni egy terminológiai jelenségre, amely az ISBD-szabályzatok magyar (félre)fordításával függ össze: a *szerezési közlés* fogalmának kérdésére. Az ISBD konszolidált változatának (IFLA, 2011) első fejezete foglalkozik a bibliográfiai leírás első adatcsoportjának adatelemeivel – az adatcsoportot *Title and Responsibility Statement Area*-nak nevezve. A *responsibility statement* kifejezés általánosabb, bármely bibliográfiai forrástípus kontextusában használható megfelelője a *felelősségi közlés* volna, amely kifejezés pontos használata fontosabb, mint eddig bármikor. Az RDA ugyanis egyértelműen a szerző fogalmának szűkebb értelmezése mellett teszi le a voksot, amikor két külön metaadatelemet definiál a *szöveges* művek alkotóinak rögzítéséhez (ezt, és csak ezt nevezi az RDA *szerezőnek*), illetve *alkotónak* hívja mindazokat az ágenseket, akik felelősek egy – tetszőleges típusú – mű megalkotásáért. A *szerezési közlés* forma használatával elsiklunk e finom, ámde lényeges különbség fölött.

5.1. Az RDA és a MARC

Az RDA egyesült államokbeli tesztelését koordináló munkacsoport 2011 júniusában terjedelmes jelentést tett közzé az elvégzett munkákról és azok eredményeiről. A szakemberek megállapították, hogy az általuk készített felmérésre választ adók jó része úgy gondolta, hogy az RDA biztosította előnyök igen nehézkesen biztosíthatók MARC-környezetben, következésképp ahhoz, hogy az RDA-t hatékonyan lehessen használni a bibliográfiai források leírására, valamilyen más hordozóformátumra van szükség (*Report and Recommendations of the U.S. RDA Test Coordinating Committee*, 2011).

Amikor tehát az RDA forrásleírási szabályzat – illetve ezen keresztül az entitásszemlélet – könyvtári alkalmazhatósága kerül terítékre, feltétlenül figyelembe kell venni, hogy a jelenleg alkalmazott hordozóformátum, a MARC használatával megfogalmazott metaadatközlésekkel csupán részleges sikerek érhetők el. A MARC-ot azonban, világméretű beágyazottsága miatt rendkívül nehéz más rendszerekkel felváltani: rekordok milliárdjai léteznek különféle adatbázis-szervereken, illetve aggregációs szolgáltatásokban, és a különféle integrált gyűjteménykezelő rendszerek, illetve könyvtári szolgáltatási platformok funkcióinak jelentős része is elválaszthatatlanul kapcsolódik ehhez a formátumhoz. A világ számos intézményében – köztük az Országos Széchényi

Könyvtárban is – kísérleteket folytatnak annak érdekében, hogy az RDA rendelkezései minél nagyobb mértékben felhasználhatók legyenek MARC-környezetben.

A MARC 21-et gondozó testületek számos új mezőt, illetve almezőt definiáltak a szabványban, amelyek közül érdemes kiemelni az alábbiakat:

- 336 – tartalomtípus (content type). Értéke egy 23 elemet tartalmazó szótárkódolási sémából választható. E séma olyan értékeket listáz, mint az *állókép, beszéd, előadott zene, szöveg* stb. (*RDA Content Type*, é.n.);
- 337 – médiumtípus (media type). Szótárkódolási sémája nyolc elemet tartalmaz, olyan általános típusokat, mint a *mikroformátumú, kivetíthető, audio, számítógépes médiumtípus* (*RDA Media Type*, é.n.);
- 338 – hordozótípus (carrier type). Ötvenhat – köztük több elavultnak jelölt – elemet tartalmazó szótárkódolási séma, mely az egyes adathordozó-típusok megnevezéseit tartalmazza: *videokazetta, online forrás, objektum, hanglemezt, diafilm* stb. (*RDA Carrier Type*, é.n.). A 336-337-338-as mezők \$a almezője az RDA által előírt szöveges megnevezést, míg a \$b almező annak három karakterrel jelölt kódját tartalmazza. A forrástípus általános megnevezését hordozó 245\$h (MARC 21: *Medium*, HUNMARC: *az információhordozó általános megnevezése*) almező tartalma e mezők tartalmával sokkal részletesebben és árnyaltabban fejezhető ki;
- 344, 345, 346, 347 – egyes hordozótípusok sajátos jellemzőit hordozó adatmezők, amelyek a hangzóanyagok paramétereit, a mozgóképek vetítési, illetve technikai információit, illetve a digitális állományok jellemzőit tartalmazzák;
- több új adatmezővel egészült ki a MARC 21 authority formátuma is. Az egyes személyekhez, testületekhez kapcsolódó dátumokat már tipizálva is meg lehet adni a 046-os mező almezői segítségével – \$f a születési dátum, \$g a halálozási dátum stb. A személyek, testületek címadatai ugyancsak granulárisabban közölhetők – 371\$b (város), 371\$d (ország), 371\$e (irányítószám) stb. A személyek esetében megadhatók bizonyos testületi tagsági információk (373), foglalkozásra, valamint tevékenységi területre (372, 374), továbbá nemre (374) vonatkozó közlések, mindezekhez a megfelelő almezőkben (\$s, \$t) időintervallumok is kapcsolhatók;
- \$0 - \$1 – Egységes forrásazonosítók (URI-k)¹² tárolására szolgáló almezők. A \$0 almezőt akkor kell azonosítóval feltölteni, ha ez az URI valamilyen, az adott entitás-

¹² Az URI-k mibenlétével, funkciójával és jelentőségével a disszertáció 6.1. fejezete foglalkozik.

előfordulást leíró, tetszőleges adatgazdagságú, autorizált névalakot azonban mindenképp tartalmazó *authority rekordra* mutat, míg a \$1 almezőbe magát az entitás-előfordulást reprezentáló URI kerül (ez utóbbit *real world object-URI-nak* – RWO-nak – is szokták nevezni, mivel a valódi világ valamilyen egyedi objektumára utal). Természetesen az is elképzelhető, hogy egy adott entitás-előfordulás vonatkozásában mindkét almező tartalmaz valamilyen értéket (PCC, 2017):

```
100 1#
$a Obama, Michelle,
$d 1964-
$e author
$0http://id.loc.gov/authorities/names/n2008054754
$1http://viaf.org/viaf/81404344
```

E példa esetében tehát a \$0 almező tartalma a Library of Congress authority állományában található személynév *authority rekordra* mutat, míg a \$1 almezőben található URI magának Michelle Obamának mint személynek a jelölésére használható.

5.2. Problémák a MARC-kal

A MARC a könyvtárak (sőt több múzeum¹³) által világszerte használt adatsereformátum. A szakmai diskurzusban nagyon gyakran így, *formátumként* hivatkoznak rá, de a helyes megfogalmazás mégsem ez. A MARC ugyanis csupán az ISO 2709-es szabványban definiált információcsere-formátum, az ún. INEX-struktúra egy konkrét alkalmazása (McCallum, 1996): hívójelek, illetve almezők speciálisan definiált készlete, amelynek segítségével bibliográfiai források metaadatait lehet rögzíteni. Egy másik hívójel- illetve almezőkészlet segítségével lehetőség van az autorizált névalakok szabványban előírt elemeit, továbbá a névalak által megnevezett entitás-előfordulás más adatait tartalmazó adatszerkezet (*authority rekord*) létrehozására. További elemkészletekkel lehetővé válik példányinformációk, osztályozási információk, sőt akár nem bibliográfiai források leírása is, ha azok kielégítik egy bizonyos közösség információs igényét (*MARC Standards*, é.n.). Voltaképpen a nemzeti MARC-ok sem mások, mint speciális elemkészlet-változatok, melyek mind ugyanarra az ISO-szabványban meghatározott struktúrára épülnek. Ezek történetét Campos és szerzőtársai dolgozták fel, s cikkükben következetesen a többes számú

¹³ A Monguz Kft. múzeumi használatra átdolgozott Huntéka szoftvere a múzeumi műtárgyak leírását is MARC-rekordok segítségével oldja meg – az intézményspecifikus metaadatelemeknek szabadon definiálható (900-as tartományban lévő), vagy a HUNMARC-szabványban nem definiált, más tartományokból vett (például 652) hívójeleket adva.

MARC-formátumok alakot használják, ezek közé értve az olyan elemkészleteket is, mint például a PICA vagy a MAB (Campos, Lopes & Galvao, 1995).

A bibliográfiai, azaz a forrásleírásra alkalmazható MARC-elemkészletet kidolgozója, Henriette D. Avram úgy alkotta meg, hogy annak struktúrája pontosan követi a katalógustétel szerkezetét. Az ISBD szabályzatok rendelkeznek a bibliográfiai leírás adatszoportjainak, illetve azok adatelemeinek sorrendjéről (IFLA, 2004), amiből a katalógustételen látható információk sorrendje is következik. Ez a sorrend határozza meg azt, hogy egy bizonyos adatelem vagy adatszoport milyen hívójellel rendelkezik a MARC-ban. Alacsony – 100/110/111-es – számot kap a kiemelt hozzáférési pont adatmezője, annak típusától függően; a cím- és felelősségi közlés adatszoport elemei, melyekkel a leírás elkezdődik, a 245-ös adatmezőbe kerülnek, a kiadási adatok a 250-es, a megjelenési adatok a 260-as hívójelet kapták, és így tovább. A bibliográfiai tételen a leírás alatt elhelyezkedő tárgyi, illetve további hozzáférési pontok adatmezői már egészen magas, 600-700-as hívójeleket kaptak.

A MARC életének csaknem fél évszázada hosszú és termékeny időszak volt. Fő célját, amit alapvető szerkezete is tükröz – a katalóguscédulák elkészítésének és sokszorosításának megkönnyítését – évtizedeken keresztül megfelelően teljesítette. Az informatikai környezet gyökeres megváltozása, a világháló megjelenése azonban jelentősen visszavetette a MARC használhatóságát. A géppel olvasható katalogizálási formátum alkotóelemeire, illetve logikai operátorokra épülő információkereső rendszerek kezelése az internet keresőszolgáltatásait használók számára igen nehézkesnek bizonyul (erről bővebben szólok az értekezés 7.3 fejezetében). A MARC a XXI. század első éveiben már igencsak elavultnak mondható, használata során egyre több nehézség merül fel.

Az egyik legjelentősebb probléma, hogy a jelenlegi adatsere-formátum sajátos szerkezete miatt nem támogatja az entitásalapú katalogizálást, nem teszi elkülöníthetővé, és ezzel külön kereshetővé az egyes entitásszinteket. Egy MARC-alapú integrált könyvtári – vagy tágabb értelemben véve: gyűjteménykezelő – rendszer alapjában véve három vagy négy rekordtípus kezelését támogatja: a bibliográfiai rekordét, az authority rekordét, a példányrekordét, egyes rendszerek ettől elkülönülve az ún. holdingrekordok készítését is lehetővé teszik. Ezek a rekordok csak limitált mértékben kapcsolhatók össze egymással: authority rekord összeköthető bibliográfiai rekorddal az ún. authority mezőkön keresztül, bibliográfiai rekordok egymással pedig kizárólag a MARC-ban definiált rekordkapcsolati

mezőkön keresztül kapcsolódhatnak. Ahhoz, hogy a MARC-ot megfelelően lehessen használni entitásalapú katalogizálásra, jóval több, az egyes entításoknak megfeleltetett rekordtípusra és ezek között jóval nagyobb számban definiált kapcsolódási pontokra lenne szükség.¹⁴ Ez az LRM és az RDA esetében is legalább 10 rekordtípust jelent, még akkor is, ha a rendszerszempontról értelmezhetetlenül absztrakt *res* (és RDA entitás) kimarad. Ez azonban már szétfeszítené a használhatóság kereteit, és pontosan azt a célt tenné elérhetetlenné, amely az FRBR kidolgozói előtt lebegett: a forrásleírási munkafolyamat gazdaságosabbá, egyszerűbbé tételét.

A MARC keretein belül tehát nincs lehetőség a forrásleírás FRBR-ben megfogalmazott szemléletéhez szükséges entitásstruktúra leképezésére. Ebből adódik, hogy az RDA szabályai, amelyeket éppen az entítások biztosította vezérfonal mentén szerkesztettek, csak részben érvényesíthetők a hagyományos, rekordokra épülő szerkezetben. A bibliográfiai és az authority formátumot ugyan kiegészítették számos új elemmel, az ún. RDA-mezőkkel (Library of Congress, 2014) – ezek egy részét az 5.1. fejezetben ismertettem –, azonban ez nem oldja meg gyökeresen a helyzetet. Az entitásalapú forrásleíráshoz a MARC szerkezetének teljes átformálása lenne szükséges.

A MARC a könyvtári közösség szabványaként jött létre – és az is maradt, nem terjedt el más, kapcsolódó szakmák, például a könyvkiadás és -kereskedelem területén. Az általánosan alkalmazott adattárolási és adatcsere-formátumokhoz képest jelentős lemaradásban van: napjainkban nehéz olyan egyéb tudományterületet mondani, amely a bibliográfiai adatokat MARC-ban tárolná (Library of Congress, 2008). Ennek igen lényeges és sajnálatos eredményei vannak. Az egyik, hogy bármilyen, könyvtári területen kívül eső fogadó részére igen nehezen, aprólékos kidolgozást igénylő, bonyolult konverziós eljárások után lehet csak adatokat szolgáltatni, továbbá az adatok automatizált átvitele csak speciális, szűk kör által alkalmazott hálózati protokoll (Z39.50) segítségével valósulhat meg. A könyvtári világon kívül fejlesztett alkalmazások nem képesek feldolgozni a MARC-ban kódolt bibliográfiai adatokat. A könyvtárosok választási lehetőségei pedig beszűkülnek,

¹⁴ A Petőfi Irodalmi Múzeum egyedi fejlesztésű Huntéka-rendszerében már történtek olyan lépések, amelyek az entitásalapú katalogizálás felé mutatnak: a személynevek authority rekordjában például nyilvántartják az illető személy születési és halálozási helyét, amelyet a földrajzi tárgyszavakat leíró (ugyancsak authority!) rekordokkal létrehozott kapcsolat segítségével oldanak meg. Ugyanakkor mindenképp szükséges további rekordkapcsolati pontok kialakítása, hogy például az egyes személyek lakóhelyeit is rekordbeemeléssel, ne begépeléssel lehessen rögzíteni a személynévrekordban.

amikor az információtechnológiai piacnak csak egy kis szeletéből választhatnak szoftvereket: azokat, amelyek képesek kezelni a felgyülemlett adatvagyonot.

A MARC/XML kidolgozása sokat javított a könyvtárak elszigeteltségi problémáján: az XML általános szerkezete miatt az ilyen formátumba csomagolt bibliográfiai leírásokat más intézmények is kezelni tudják, s kibővül az adattranzfer-protokollok köre: az XML-lé konvertált MARC-rekordok akár OAI-PMH, akár SRU-protokoll használatával is továbbíthatók a külső partnerek felé. Miután azonban a MARC/XML címkéi továbbra is a „field”, „subfield” stb. neveket viselik, attribútumaik között pedig a MARC-ban alkalmazott hívójelek és almezőjelek találhatók, az adatelemek *értelmezése* továbbra sem megoldott azok számára, akik nincsenek tisztában a MARC-szabvánnyal. Automatizált, humán ágenseket nélkülöző feldolgozási folyamatban ugyancsak megállapíthatatlan az adatelemek jelentése.

A MARC, bár a könyvtárak közötti, esetenként nemzetközi szintű adatcserét hivatott támogatni, az egyes intézményi sajátosságok adaptálása, valamint a nemzeti MARC-változatok közötti kisebb-nagyobb eltérések meggátolják e feladat hatékony végrehajtásában. E két tényező rendkívül megnehezíti például a közös katalógusok építését a különböző helyekről eltérő szerkezetben érkező adatsomagok miatt.

Szintén lényeges probléma, hogy a MARC-rekordokban tárolt adatok, bár az OPAC használatával online is elérhetővé váltak, valójában a mély web részét képezik, ezért a keresőszolgáltatások számára lényegében láthatatlanok. A könyvtári adatbázisok hatalmas adatsilókká váltak, amelyek, ha csak a használó nem ismeri a katalógus pontos URL-címét, kiaknázatlanul maradnak. Óriási mennyiségű adat tárolódik számos különálló könyvtári katalógusban elrejtve, amelyek az alkalmazott, elavult metaadatformátum miatt nem kereshetők más, egyszerűbben használható segédeszközök segítségével, mint például a Google (Dean, 2013). A MARC melletti kitartás abba a veszélybe sodorja a könyvtárakat, hogy nem lesznek képesek megfelelően kielégíteni a modern használók igényeit az általuk gyakran használt információs környezetekben, például a keresőszolgáltatások találati listáiban (Gonzales, 2014). „*Nem beszélünk olyan nyelvet, melyet a Web megért*” – panaszkodik Miller (2014).

5.3. A MARC-alapú katalógus újraszervezése az entitások mentén – a FRBR-izáció

Az előzőekben leírtak alapján belátható, hogy a MARC sajátos, rekordtípusokra épülő szerkezete miatt nem alkalmas arra, hogy az entitások mentén szerveződő bibliográfiai

leírásokat tároljon. Az OPAC- vagy discovery felületeket ugyanakkor már hosszú ideje igyekeznek alkalmassá tenni arra, hogy azok a keresésekre eredményül kapott MARC-rekordokat az FRBR alapentitásai mentén *utólag* csoportosítva jelenítsék meg, így vezetve a használót a magas absztrakciós szintektől egészen a konkrét könyvtári példányokig. Ezt a módszert nevezik a katalógus FRBR-esítésének (avagy főbörizálásának), amelyet már a kétezres évek első felének végén számos projektben és intézményben alkalmaztak. Ezeket az implementációkat Eden (2006) szedte csokorba, s több kísérlet részletes leírása a szaksajtóban, illetve tanulmánykötetekben is megjelent. A VTLS integráltrendszer-fejlesztő cég Virtua nevű termékében megvalósított technikai megoldásokat Espley és Pillow (2012) mutatta be, Callewaert (2013) a flamand közkönyvtári portálon elkészített implementációt elemzi, Hickey és O’Neill (2005) az OCLC WorldCat szolgáltatásának átalakításáról szól, Notess, Dunn és Hardesty (2011) pedig egy FRBR-re épülő zenemű-kereső felület, a Scherzo kialakításáról számolnak be. Martha M. Yee (2005) elméleti irányultságú cikke a MARC-mezők entitásonként történő csoportosítását elemzi, néhány problémás esetet szemügyre véve.

Az FRBR-esítési művelet sikeres és hatékony végrehajtásának egyetlen követelménye a rekordok bizonyos adatmezőinek megléte, azaz a kielégítő adatgazdagság. A hagyományos, MARC-ban tárolt bibliográfiai rekordok FRBR-izált megjelenítésének technikáit egy 2015-ben született összefoglaló tanulmányban (Decourselle, Duchateau & Lumineau, 2015) három alapvető szempont mentén csoportosították a szerzők: az entitások kialakításának logikája, az alkalmazott adatmodell kifejezőképessége, illetve a megvalósított specifikus kiegészítések alapján.

Az entitásokon alapuló megjelenítés előkészítése végbemehet az eredeti MARC-rekordok csoportosításának, illetve az adatmezők gondosan kidolgozott szétválogatásának útján. Az első esetben a rekordokban előforduló egyes MARC-mezők, leggyakrabban természetesen a szerzők/közreműködők, illetve a címek egyezése alapján igyekeznek műveket és kifejezési formákat reprezentáló csoportokat alkotni. Ezt a módszert alkalmazza például az Ex Libris által fejlesztett Primo discovery felület. A MARC-rekordból a szoftver első lépésben egy normalizált XML-t, ún. PNX-formátumot állít elő. A PNX-címkekészlet több eleme (az ún. FRBR-kulcs) a rekordokban eredetileg tárolt adattartalom normalizálási lépéseken átesett alakját jelöli, az FRBR-izálás pedig ezek egyezésére alapozva megy végbe (Haász, 2017). Például a szerzői/közreműködői névalakok (MARC: 100/110/111 illetve 700/710/711) átalakított változatait a Primo a PNX előállításakor a „k1” címkével látja el –

címek esetében a 130-as, illetve a 240/242/245/246/247/740 mezők adattartalmával dolgozik (ExLibris, é.n.). Hasonló, összevetésen alapuló algoritmust dolgoztak ki az OCLC szakemberei, amely Work-Set Algorithm néven vált ismertté, és művek elkülönítését teszi lehetővé. E munkafolyamat alapdokumentuma 2005-ben, második változata 2009-ben készült el (Hickey & Toves, 2009).

```
<frbr>
  <t>1</t>
  <k1>$$Kjokai mor 1825 1904$$AA</k1>
  <k3>$$Kerdely aranykora$$AT</k3>
</frbr>
```

7. ábra – A Primo normalizált XML FRBR-kulcsokat leíró részlete (képernyőkép, saját szerkesztés)

FRBR-izálni azonban lehetséges más módon is – ez az ún. szabályalapú megközelítés –, ezt a módszert alkalmazta például a Library of Congress a 2000-es évek második felében, amellyel egy FRBR Display Tool nevű megjelenítő eszközt működtettek. A MARC-rekordokból ebben az esetben először MARC/XML-t készítettek, amelyet különböző átalakítási szabályokat tartalmazó stíluslapok segítségével feleltettek meg a MODS (Metadata Object Description Scheme) metaadatformátum elemeinek, valamint ilyen stíluslapok segítségével végezték az eredeti rekordok információtartalmának entitások mentén történő csoportosítását is (Schneider, 2008).

5.4. Az RDA magyarországi szakmai műhelyei

Az Országos Széchényi Könyvtár Országos Könyvtári Szolgáltatások Igazgatósága (OKSZI), illetve az Országos Könyvtári Szabványosítási Bizottság (OKSZB) munkájának támogatására intézményközi RDA-munkacsoport jött létre 2018 márciusában, Dancs Szabolcs elnökletével. A tagság jelentős része az OSZK dolgozóiból állt, de képviseltették magukat az MTA Könyvtár és Információs Központ, az Országos Pedagógiai Könyvtár és Múzeum, az ELTE Egyetemi Könyvtár, illetve a Petőfi Irodalmi Múzeum munkatársai is. A szakembergárda 2019-ben a DEENK, az SZTE Klebelsberg Könyvtár, az MMA, illetve az OGYK képviselőivel is kibővült – a tagok aktuális névsora az OSZK honlapján hozzáférhető¹⁵. A tagok elsősorban terminológiai-fordítási kérdésekkel (különösképpen a Könyvtári Referenciamodell, illetve a Nemzetközi Katalogizálási Alapelvek magyarításából

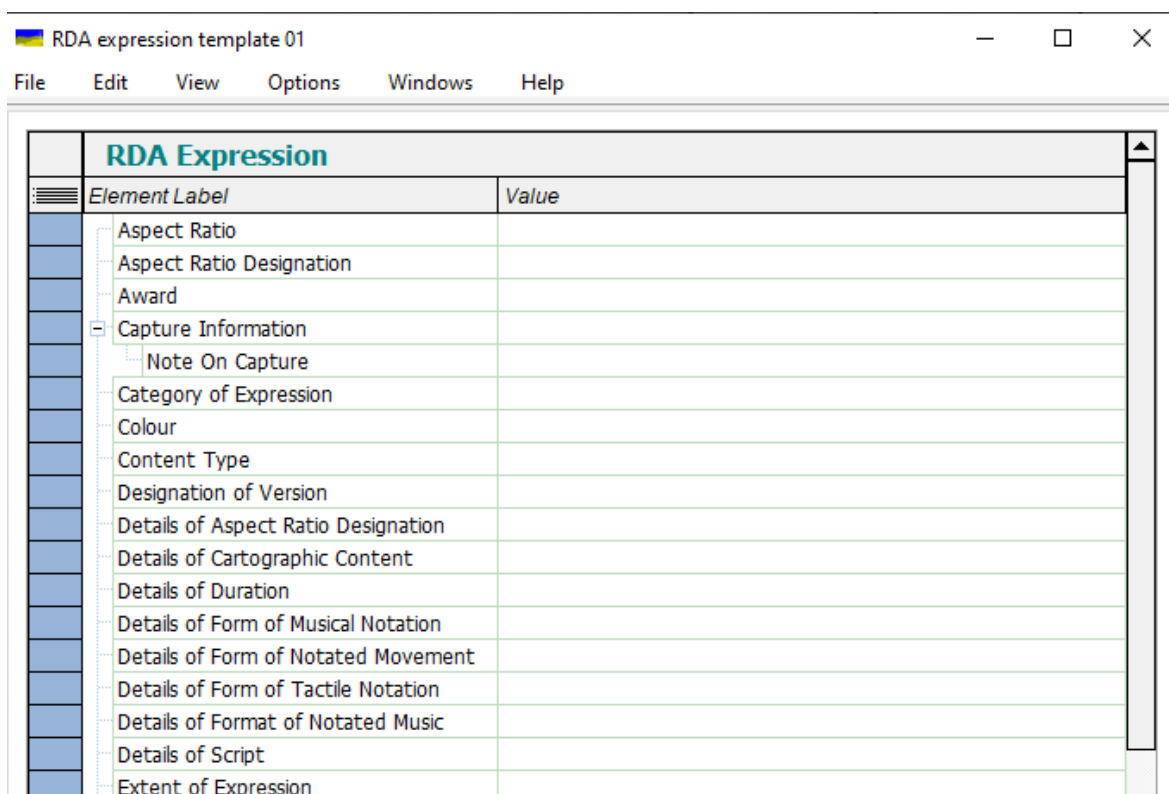
¹⁵ A <https://www.oszk.hu/rda-hu-munkacsoport> címen.

eredő felvetésekkel), a digitális másolatok bibliográfiai kezelésével, az RDA bevezetésének ütemezésével, a szabályzat MARC-keretek között is betartható rendelkezéseinek implementálásával, illetve az RDA-nak a MARC 21-re gyakorolt hatásaival foglalkoztak.

A 2020. első negyedévében beköszöntő pandémia miatt az RDA-HU munkacsoport hosszú ideig szüneteltette tevékenységét (március 16-i találkozására már nem kerülhetett sor a hatályos védelmi intézkedések miatt), formálisan azonban nem oszlott fel. A szabályzattal kapcsolatos, szerteágazó tevékenységeket 2021 novemberéig az Országos Széchényi Könyvtárban működő Könyvtári Szabványosítási Iroda (KöSZI), valamint a Petőfi Irodalmi Múzeum Digitális Bölcsészeti Központja (PIM-DBK) végezte. A végbement szervezeti változások miatt napjainkban ismét az OSZK-KöSZI koordinálja az RDA-val kapcsolatos tevékenységeket, továbbá újra működik az RDA-HU munkacsoport is, tagjai egy aktívan tevékenykedő szűkebb, továbbá egy – más elnevezés hiányában – *Interest Group*-nak nevezett tágabb csoportba tömörülnek. Az RDA-val összefüggő fontosabb feladatok a következők:

- *képzések, előadások szervezése.* A KöSZI, illetve a DBK munkatársai több alkalommal tartottak prezentációkat, workshopokat az RDA-val, illetve az entitásalapú katalogizálással kapcsolatos ismeretekről. E disszertáció írásának idején is zajlik egy széles közönséget megszólító, akkreditált, 30 órás tanfolyam anyagának, menetrendjének kidolgozása, amelyet a Könyvtári Intézet munkatársaival közösen valósítunk meg;
- *az RDA elemkészleteinek, szótárkódolási sémáinak fordítása,* illetve a fordítások lektorálása. Az RDA Registry-n található hatalmas korpusz (a fontosabb entitások több száz metaadatelemmel rendelkeznek!) fordításának oroszánrésze már elkészült, az elemek címkéinek, illetve definícióinak magyar változatát már publikálták. A tényleges szabályzatszöveg, azaz az RDA Toolkit tartalmának átültetése (körülbelül 200.000 karakter összterjedelemben) még várat magára, s várhatóan egy évnyi munkát jelent majd, amelyet hat hónapos lektorálási fázis követ;
- folyamatban van egy *magyarázatos példatár,* illetve *feladatgyűjtemény elkészítése;*
- a katalogizálási munkát jelenleg meghatározó MSZ-szabványok, továbbá az RDA rendelkezéseinek módszeres egybevetése, *szabványanalízis;*
- nemzeti *RDA-alkalmazásprofil kidolgozása;*
- az oktatást támogató, az entitásalapú katalogizálás megfelelő szemléltetését biztosító *szoftvereszközök, keresőfelületek felkutatása,* beemelése az oktatási folyamatba.

Olyan integrált könyvtári rendszer vagy szolgáltatási platform, amely natívan támogatná az entitásalapú katalogizálást, még nem áll rendelkezésre – ez a szoftveres környezetek megkerülhetetlen MARC-ra épüléséből következik. Az OSZK, illetve a PIM-DBK szakemberei olyan megoldások után kutatnak, amelyek lehetővé teszik az entítások egymástól elkülönülő metaadatolására, és egymáshoz kapcsolására alapuló feldolgozás élményének bemutatását. Ennek egyik ígéretes segédeszköze a *RIMMF* elnevezésű, ingyenesen letölthető és kipróbálható szoftver, amelynek négyes verziószámú változata már az LRM miatt átalakított forrásleírási szabályzattal való ismerkedést teszi lehetővé. A RIMMF mellett munkatársaimmal a holland nemzeti könyvtár által készített Entity Finder nevű szolgáltatást, továbbá a Sinopia, a BIBFRAME Scribe és a VitroLib, valamint az értekezés 9. fejezetében részletesen ismertetett BIBFRAME Editor nevű segédeszközöket vontuk közelebbi vizsgálat alá.



8. ábra – A RIMMF szoftver kifejezésforma-leíró űrlapja (képernyőkép, saját szerkesztés)

5.5. Néhány gondolat az OSZK RDA-projektjéről

Az Országos Széchényi Könyvtárban is végeztek forrásleírási kísérleteket a RIMMF szoftver használatával, ezek a próbálkozások azonban hosszú ideig meglehetősen elszigeteltek és esetlegesek maradtak. A nemzeti könyvtár munkatársai ugyanis abból az alaptételből kiindulva, hogy a feldolgozási munkakörnyezet még hosszú ideig MARC-alapú lesz, úgy vélekedtek, hogy az RDA-val való ismerkedést (legalábbis egyelőre) a MARC szabta keretek között kell megvalósítani, legyenek azok bármennyire szűkek. Mivel a RIMMF nem MARC-alapú, ezért helyette a KOHA nevű, open source, böngészőben használható rendszert választották, amiben fontos szempont volt ennek a rendszernek az átlagosnál jobban testre szabható katalogizálási felülete (Dancs, Mohay & Hubay, 2019). A bibliográfiai rekordok az RDA mezőkkel kiegészített, MARC 21-re épülő űrlapokon készültek el – az űrlapok és adatmezőik konfigurációját a Petőfi Irodalmi Múzeum munkatársa, Mohay Anikó végezte –, illetve a szabványnak megfelelő formátumú, authority rekordok csatlakoztak hozzájuk, melyek személy- illetve testületi nevek autorizált alakját, és több kiegészítő információt tartalmaztak. Ez a rekordhalmaz adta az alapját annak a kísérleti miniprojektnek, amely a szemantikus web adatformátumaira történő konverzióra, illetve az adathalmaz nyilvánosan hozzáférhetővé tételére irányult. A rekordok linked data-gráfokká történő átalakítását az e sorok írója által összeállított specifikáció, illetve megfeleltetési dokumentum alapján végezték el. E kísérlet kivitelezéséről részletesen szöveg az értekezés 10. fejezetében.

Az entitásközpontú forrásleírás során a megjelenési formák leírásától elkülönítve kell tárolni a művek, valamint a kifejezési formák metaadatait is. Ezt a KOHA-rendszerben új authority rekordtípusok definiálásával érték el olyan módon, hogy az egyes entitások metaadatait reprezentáló MARC 21-hívójeleket különválogatták. Így külön rekordtípus állt elő a művek, és egy másik a kifejezési formák leírására. A kifejezési forma rekordja a 240-es – esetenként a 700-as – adatmezőben csatlakozott a megjelenési formát leíró, hagyományos bibliográfiai rekordhoz, míg a mű-típusú authority rekord a KOHA által definiált 100\$9 almezőben csatlakozott a kifejezésiforma-típusú authority rekordhoz.

Az alábbi példában a megjelenési formát leíró MARC-rekordon túl egy mű és egy kifejezési forma rekordja is látható. Félkövérrrel kiemelve szerepelnek azok az adatmezők, illetve adatelem-részletek, amelyek az RDA rendelkezései nyomán kerültek az adatszerkezetbe:

```

=LDR 00436nz a2200145n 4500
=001 60186
=003 HU-IL-001206-1
=005 20181120105535.0
=008 181114||\aza||aabn\|\|\|\|\|\|\|\|\|\|\|a|a\|\|\|\|\|d
=040 \\$aHU-IL-001206-1$bhun$eRDA
=100 1\ $aMolnár Ferenc,$d1878-1952.$tAz ibolya
=380 \\$059018$2oszkdoktip$aszínmű
=670 \\$aSzínház : három egyfelvonásos / írta Molnár Ferenc (1921)
=942 \\$aMű

=LDR 00472nz a2200157n 4500
=001 60199
=003 HU-IL-001206-1
=005 20181120110437.0
=008 181114||\aza||aabn\|\|\|\|\|\|\|\|\|\|\|a|a\|\|\|\|\|d
=040 \\$aHU-IL-001206-1$bhun$eRDA
=100 1\ $960186$aMolnár Ferenc,$d1878-1952.$tAz ibolya.$lmagyar
=336 \\$054029$aszöveg$btxt
=377 \\$ahun
=670 \\$aSzínház : három egyfelvonásos / írta Molnár Ferenc (1921)
=942 \\$aKIFEJEZÉSI

=LDR 01677nam a2200361 i 4500
=999 \\$c233152$d233152
=001 233152
=003 HU-IL-001206-1
=005 20181210160109.0
=008 181113s1921\|\|\|hu\|\|\|\|\|\|\|\|\|\|\|00|\|d\hun\|d
=040 \\$aHU-IL-001206-1$bhun$cHU-IL-001206-1$eRDA
=041 0\ $ahun
=080 1\ $a894.511-25
=100 1\ $aMolnár Ferenc,$d1878-1952$eszerző$4aut$04016$94016
=240 1\ $060207$aSzínház.$lmagyar$960207
=245 10$aSzínház :$bhárom egyfelvonásos /$círta Molnár Ferenc.
=264 \1$aBudapest :$bFranklin-Társulat, $c[1921]
=264 \4$c@1921
=300 \\$a215, [1] oldal ;$c14 cm
=336
\\$2rdacontent$aszöveg$btxt$0http://rdaregistry.info/termList/RDAContentType/1020$0http://id.loc.gov/vocabulary/contentTypes/txt
=337 \\$2rdamedia$asegédészköz
nélküli$bn$0http://rdaregistry.info/termList/RDAMediaType/1007$0http://id.loc.gov/vocabulary/mediaTypes/n
=338
\\$2rdacarrier$akötet$bnc$0http://rdaregistry.info/termList/RDACarrierType/1049$0http://id.loc.gov/vocabulary/carriers/nc
=340 \\$2rdagen
=347 \\$2rdaft
=505 00$gI.$tElőjáték Lear királyhoz : szatíra egy felvonásban$gII.$tMarsall : színmű egy felvonásban$gIII.$tAz ibolya : vígjáték egy felvonásban
=700 12$aMolnár Ferenc,$d1878-1952.$tElőjáték Lear királyhoz.$lmagyar$960202
=700 12$aMolnár Ferenc,$d1878-1952.$tMarsall.$lmagyar$960204
=700 12$aMolnár Ferenc,$d1878-1952.$tAz ibolya.$lmagyar$960199
=710 2\ $aFranklin Társulat Magyar Irodalmi Intézet és Könyvnyomda$sekiadó$4pbl$960166
=850 \\$aHuBpOSK
=852 \\$aHuBpOSK
=942 \\$2ddc

```

Az OSZK-DBK együttműködés idején – 2020-2021-ben – kiemelt feladat volt az RDA, illetve a katalogizálás entitásalapú szemléletének népszerűsítése, oktatása. A két intézmény munkatársai, köztük e sorok írója, több képzést, szakmai napot, illetve Networkshop-tutoriált szervezett, amelyeknek *A jövő katalógusa felé* címet adtuk. A képzések programjának gerincét elméleti előadások adták, amelyekben az entitásalapú katalogizálás

elméleti modelljeiről, katalogizálástörténetről és szemantikus webtechnológiáról esett szó. Ezt az ismeretanyagot gyakorlati feladatokkal egészítettük ki, ahol a résztvevők MARC-rekordokban igyekeztek entitásokat felismerni, más alkalommal pedig a RIMMF használatával próbálhatták ki a bibliográfiai források több entitásszintű leírását (PIM, 2021).

6. A linked data – adatok szemantikus összekapcsolása¹⁶

2001-ben Tim Berners-Lee egy, a jelenlegitől eltérő célú, alternatív világháló vízióját alkotta meg a Scientific American hasábjain. A koncepciót *szemantikus*, azaz értelemmel teli webnek nevezte el, és olyan tartalmak összességéként definiálta, amelyek a számítógépek számára is értelmezhetők (Berners-Lee, Hendler & Lassila, 2001). A hálózati informatika zseniális újítása, a weben elhelyezett dokumentumok közötti hivatkozások, hiperlinkek létrehozása napjainkban már nem kielégítő, mégpedig azért, mert nem ad lehetőséget annak jelölésére, hogy a két, egymással összekötött webdokumentum milyen kapcsolatban áll egymással, azaz nem lehetséges a relációk tipizálása. Ahogyan Tim Berners-Lee és szerzőtársai fogalmazzák: „*a szokásos hipertextes világhálón összekapcsolt két dokumentum közötti kapcsolat nincs kifejtve, mivel az alkalmazott adatformátum – a HTML – nem elég kifejező ahhoz, hogy az egyes dokumentumokban leírt entitások kapcsolatát más entitásokhoz különféle típusú linkekkel adjuk meg*” (Bizer, Heath & Berners-Lee, é.n.). A weben elhelyezett dokumentumokat ugyanis csak az őket tanulmányozó ember tudja interpretálni, azok tartalma a gépi folyamatok számára feldolgozhatatlan: a keresőszolgáltatások oroszlánrésze nem az értelmezésen, hanem az indexelt webdokumentum-állomány, valamint a begépett keresőkérdés mechanikus egybevetésén, azok karakterszintű egyezésén alapul.

Berners-Lee felfogásában a szemantikus web nem egy diszjunkt konstrukció, épp ellenkezőleg: a ma is létező, jól ismert világháló egy hozzáadott rétege, amely hosszú ideje bevált és alkalmazott technológiákon alapszik – csupán kevés új elemet vezet be. Ugyanakkor abban mégis különbözik a World Wide Webtől, hogy nem webdokumentumokat, hanem jóval kisebb egységeket, *adatokat* köt össze egymással olyan módon, hogy az adatok közötti kapcsolatok típusa is rögzíthetővé, számítógéppel interpretálhatóvá válik. Ez pedig a megfelelő technológiai megoldások segítségével lehetőséget nyújt gépi következtetések elvégzésére, azaz olyan, adatok közötti összefüggések megállapítására, amelyek explicit módon nem találhatóak meg azokban. A szemantikus web tehát úgy is értelmezhető, mint az adatok hálózata (*web of data*), amely adatokból és tipizált kapcsolataikból szövődik (Solodovnik, 2013). A hálózat építőköveinek

¹⁶ A linked data kifejezés – bár könnyűszerrel fordítható *(össze)kapcsolt/kapcsolódó adatokra*, tapasztalataim alapján a magyar nyelvű szakmai diskurzusban nem honosodott (még) meg ez a forma. Ezért itt és a továbbiakban az eredeti, angol nyelvű megnevezést használom.

összekapcsolódását, illetve – optimálisan – nyílt hozzáférését hangsúlyozza egy másik, a szakirodalomban elfogadott kifejezés, a *linked open data*.

6.1. A szemantikus web építőköve – az URI

A világháló megalkotója *Linked Data* című írásában (Berners-Lee, 2006) négy ajánlást tett közzé, amelyek követésével hatékonyan bővíthető az egymással összekapcsolódó adatok hálózata. Az első szerint a dolgok megnevezésére URI-kat kell használni, amelyeknek – a második ajánlás értelmében – http-alapúaknak kell lenniük, hogy ennek segítségével az URI által reprezentált entitás-előfordulásról a hálózaton információkhoz lehessen jutni. A harmadik ajánlás szerint az adatgazdának gondoskodnia kell arról, hogy az URI-k felkeresésekor a felhasználó a vonatkozó informatikai szabványoknak megfelelő információt kapjon vissza, míg a negyedik ajánlás tartalmazza a linked data-ökoszisztéma lényegét: arra bátorítja az adatgazdát, hogy helyezzenek el hivatkozásokat más URI-kra, hogy az azok által leírt entitás-előfordulásokról még több információt lehessen szerezni.

Az URI, az egységes forrás-azonosító egységes, mert struktúrája kötött, és egységes, mert a világon mindenhez társítható. A *forrás* kifejezés értelmezése már nehezebb: ugyanolyan tág értelmű fogalommal állunk szemben, mint az *entitás* esetében. Mindkettőre igaz az, hogy valami olyat jelölnek, ami egyértelműen körülhatárolható, elválasztható minden mástól, és róla kijelentések, állítások tehetők. Az *entitás* és a *forrás* tehát leírható annak saját adatai segítségével, legyen az akár konkrét, kézzelfogható dolog, vagy elvont fogalom, még arra sincs szükség, hogy a valóságban létezzen (például egy hétfejű sárkány). Bármely, a világon absztrakt vagy kézzelfogható módon létező dologhoz hozzárendelhető tehát egy egységes szerkezetű azonosító. Tim Berners-Lee ajánlása alapján ennek az egységes szerkezetnek hasonlónak kell lennie az URL-ekhez, amelyek képesek megmutatni bármely, a világhálón elhelyezett webdokumentum helyét. Jól ismert informatikai megoldás, hogy amikor egy böngészőprogramban megjelenik egy URL, akkor a szoftver egy külső szolgáltatás (az ún. Domain Name Server) segítségével lefordítja azt egy webszerver IP-címére, majd ennek a webszervernek egy ún. GET-kérést küld, amelynek informatikai részleteit a http-protokoll szabályozza. Ideális esetben a megszólított webszerver fogadja a böngészőtől érkező kérést, és a kért tartalmat adja válaszul.

Egy URI azonban nem webdokumentumot, hanem valamilyen konkrét entitás-előfordulást azonosít, amelynek már nem szükséges a világháló valamilyen alkotóelemének

lennie. Használatával lehetségessé vált a valóságban létező bármely objektumot az interneten, az automatizált adatfeldolgozás számára is értelmezhetően azonosítani, így a világháló információs hálózatból a dolgok hálózatává alakul át.

Az URI-k használata során a http-protokoll és az ún. tartalomjegyeztetési (*content negotiation*) technológia segítségével biztosítható a kommunikáció egy, a kérést indító kliens és egy webszerver között. Például az alábbi URI-hoz:

<http://bibframe.org/vocab/instanceOf>

nem tartozik webdokumentum, hiszen az URI csupán valamilyen forrás azonosítására szolgál. Amikor egy ilyen azonosító megjelenik egy http-kérésben, a kérés azt is tartalmazni fogja, hogy a kérelmező szolgáltatás milyen típusú tartalmat vár vissza: hagyományos webböngészők (Firefox, Chrome, Edge, Opera stb.) esetében HTML-formátumban hozzáférhető, a szoftver által megjeleníthető weblapot. A kérés elküldésekor a webszerver észleli a bejövő kérést, és a kért tartalom típusától függő irányba továbbítja azt – a továbbirányítás részleteit a webszerver kezelője által megfogalmazott újirrási szabályok (*rewriting rules*) tartalmazzák. Ha a kliensprogram kérésében szöveges/HTML-alapú tartalom szerepel, akkor az átirányítás egy emberi szemmel olvasható, böngésző által megjeleníthető weblap URL-jére történik; ez a weblap az URI által jelölt forrásról közöl információkat. Ha azonban a kérést nem webböngésző, hanem valamilyen automatizált, gépi szolgáltatás indította, megváltozik az igényelt tartalom típusa. Ennek teljesítéséhez egy *másik* irányba, egy másik webdokumentumhoz történik a továbbirányítás, és ebben az esetben az URI által jelölt forrásról rendelkezésre álló információkat valamilyen, a HTML-től eltérő fájlformátumban kapja meg a kérő (Horváth, 2015a).

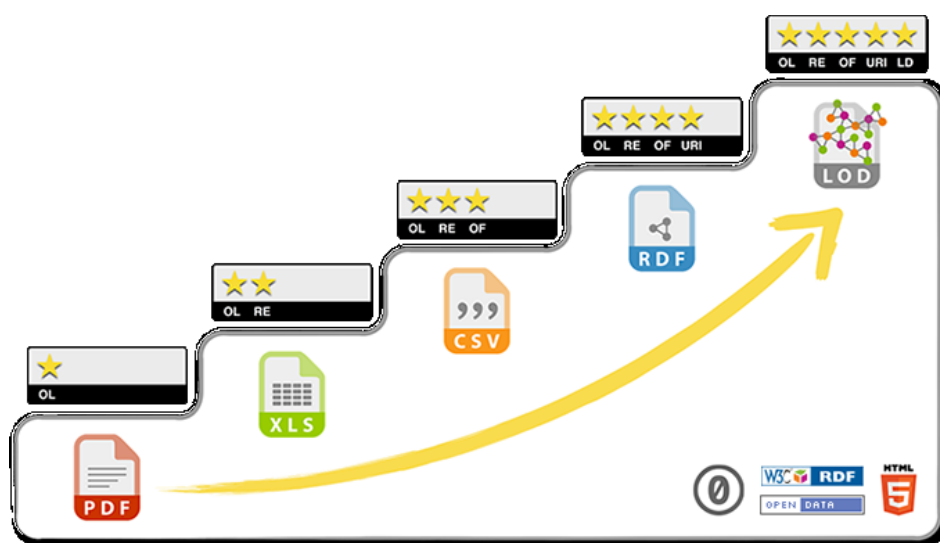
6.2. A szemantikus web szabályai és kérdései, az ötcsillagos Open Data csomag

Ugyancsak Tim Berners-Lee adja annak rövid leírását, hogy milyen paraméterekkel lehet kisebb-nagyobb adattömegeket a nyilvánosság számára hozzáférhetővé tenni a világhálón. A lehetséges fokozatokat a szerző – a szállodák által nyújtott szolgáltatásokhoz hasonlóan – csillagokkal minősíti; a legmagasabb szinten az adatok publikációja ötcsillagos (Ontotext, é.n.).

- Az első csillag akkor jár, ha egyáltalán közzé vannak téve az adatok (tetszőleges formátumban, akár szkennelt dokumentumokként), és biztosított azok megtekintése,

letöltése, felhasználása vagy megosztása bárki számára. Ez a szint egyszerűen a *nyílt* adatok szintje (open data).

- A második csillag az adatok valamilyen meghatározott szerkezetben, ún. *strukturált adatok*ként történő közzétételének szintje. A megfelelő szoftver birtokában ezek az adatok könnyedén, gyorsan áttekinthetők, könnyen végezhetők velük csoportos műveletek, valamint egyszerűen átalakíthatók más, strukturált adatformátumokba.
- A harmadik szint a *szoftverhez nem köthető* strukturált adatformátum használata, olyané, mint például a csv. A három csillagos adatok feldolgozásához nincs szükség egyedi szoftverre vagy szoftvercsomagra (mint például az Excelre az *xlsx* esetében), feldolgozásuk számos informatikai megoldás segítségével elvégezhető.
- A negyedik szinten a publikált adatok olyan módon vannak közzétéve, hogy abból kiolvashatók a közöttük fennálló kapcsolatok, illetve azok típusai is. A négy csillagos adatpublikáció tehát olyan fájlformátumok használatát igényli, amelyek támogatják az ún. RDF-logikát (erről a 6.3. fejezetben lesz szó), ezt már joggal nevezhetjük *linked* vagy *linked open datának*. E szinttől kezdve kijelenthető, hogy az adatok közzététele a szemantikus weben történt, azonban a negyedik szintű adathalmaz külső hivatkozásokat nem tartalmaz, így elszigetelt marad.
- Az ötödik szintet akkor éri el valamilyen publikált adathalmaz, ha a benne található adatok az RDF-logikát támogató adatformátumban találhatóak, és e logika segítségével mások által közzétett adatokhoz (is) kapcsolódnak, összekötve a globális adathálózat egyes láncszemeit.



9. ábra – Az adatközzététel öt csillagos lépcsője (Ontotext, é.n.)

A hálózatra az alábbi kijelentések, feltevések érvényesek (Baker, Coyle & Petiya, 2014):

- a nem egységes elnevezések feltevése (*non-unique naming assumption*): lényegében biztosra vehető, hogy egyetlen meghatározott entitás-előfordulásnak (például Jókai Mórnak) az adathálón igen sok, egymástól eltérő URI-ja létezik majd. Informatikai eszközök segítségével könnyen orvosolható ez a probléma, sőt a hálózat építésének szempontjából a jelenség kimondottan hasznos is: a 6.6. fejezetben tárgyalandó *adatgazdagítási* folyamatnak éppen az a lényege, hogy minél több, azonos entitás-előfordulást reprezentáló URI gyűljön össze, amelyek között azonossági kapcsolat létesíthető;
- bárki mondhat bármiről bármit (*anyone can say anything about anything*): a világ bármely részén, bármely entitás-előfordulás tetszőleges (akár valótlan) kijelentésekkel is jellemezhető, az informatikai eszközök nem alkalmasak az állítások valóságtartalmának ellenőrzésére. Pozitívabban szemlélve ez az állítás azt is jelenti, hogy lehetségessé válik az azonos entitás-előfordulásra vonatkozó, más és más adatgazdák által tett, eltérő tartalmú kijelentések összegyűjtése. Így a keresett témáról komplex, igen széleskörű tudás szerezhető;
- a nyílt világ feltevése (*open world assumption*): e feltevés értelmében soha egyetlen pillanatban sem jelenthető ki, hogy egy bizonyos entitás-előfordulásról már minden létező tudás rendelkezésre áll. A dolgokról való tudásunk folyamatosan fejlődik, változik, sőt egy későbbi pillanatban akár a korábbiaknak ellentmondó információk is felbukkanhatnak. Minden nap, minden percben szülehetnek új kijelentések egy entitás-előfordulásról, amelyek az URI-k segítségével könnyedén hozzáadhatók, felfűzhetők a tudáshálóra.

6.3. A szemantikus web ragasztója – az RDF

Az értelemmel teli világháló tehát nem más, mint egymással összekapcsolt, bizonyos konkrét forrásokat reprezentáló URI-k hálózata. Az összekapcsolás azonban nem minden: ahhoz, hogy a kialakuló háló valóban szemantikus legyen, szükség van a források között fennálló legkülönbözőbb kapcsolattípusok rögzítésére és azok számítógép általi értelmezésére. A World Wide Web-konzorcium szakemberei által megalkotott forrásleíró keretrendszer adatmodellje ezt teszi lehetővé. Az elképzelés alapötlete, hogy az URI-val azonosított forrásokat tulajdonságaik, ismérveik segítségével más forrásokkal köti össze (Szeredi, Lukácsy & Benkő, 2005).

Ez a keretrendszer egy egyszerű logika érvényesítésén alapszik, nevezetesen, hogy a világon minden leírható, azaz metaadatulható egyszerű, elemi kijelentések segítségével. Ezeket a tömör, elemi állításokat tripletnek nevezik. A tripletek, miként a nevük is sugallja, az RDF szabályai szerint három elemet tartalmaznak: az alanyt (*subject*), az állítmányt (*predicate*) és a tárgyat (*object*). A triplet alanya az a konkrét forrás, illetve az azt reprezentáló URI, amelyről megfogalmazzuk az adott kijelentést. Az állítmány az alany egy tulajdonságának megnevezését tartalmazza – szintén URI formájában –, s így összekapcsolja, valamilyen összefüggésbe hozza a triplet első és harmadik elemét. Az utolsó elem, a tárgy, az állítmányban megfogalmazott tulajdonság értékét adja meg. Ez történhet egy harmadik URI használatával, de egyszerű, közvetlenül megadott érték formájában is, amelyet az RDF-környezetben literálnak neveznek. Ennek alapján a tripleteknek két fajtája létezik: az első a *literáltriplet*, ahol a tárgy helyén nem azonosító, hanem hivatkozás nélkül rögzített tartalom fordul elő. A második az RDF hivatkozás, amely két forrás kapcsolatát írja le (Bizer, Heath & Berners-Lee, é.n.). Haller és munkatársai e hivatkozások tipizálását és funkcióelemzését végezték el összefoglaló tanulmányukban (2020).

Az alábbi kijelentések literáltripletek:

Péter – születési helye – Budapest.
Péter – édesapja – István.
A mű – címe – Sötétség délben.
A kifejezési forma – közreműködője – Bart István.

A következő lépésben az állítások elemeit URI-kkal helyettesítem. (E disszertáció céljaira a példákban használt azonosítókat az ELTE ITDI domainjével alkotom. Amikor azonban egy adatgazda a saját maga által összegyűjtött tudás publikálásáról dönt a szemantikus weben, igen nagy szabadságot élvez az URI-k kialakításában. Az egyetlen követelmény a http-formátum, ezen felül pedig törekedni kell arra, hogy az URI „tetszetős” (*cool*) külalakkal bírjon (W3C, 2008), tehát legyen informatív a humán ágensek számára is. Az adatszolgáltatók általában maguk alakítják ki saját azonosító-alkotási szabályrendszerüket, amelyet *URI-névkonvenció*nak is hívnak).

Az RDF hivatkozás:

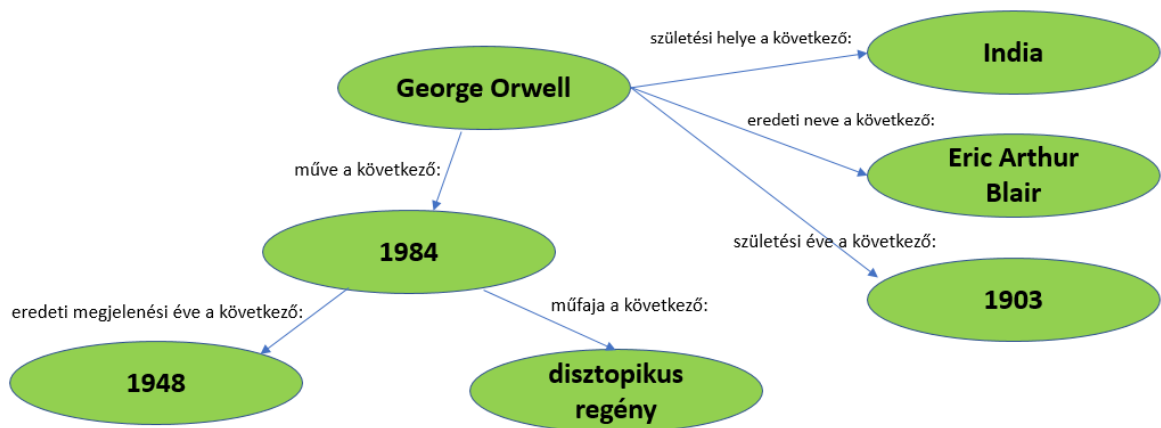
`http://itdi.elte.hu/szemely/Peter`
`http://itdi.elte.hu/szemelyiadat/szuletesihelye`
`http://itdi.elte.hu/foldrajzihely/Budapest`

`http://itdi.elte.hu/szemely/Peter`
`http://itdi.elte.hu/szulo/edesapja`
`http://itdi.elte.hu/szemely/Istvan`

<http://itdi.elte.hu/FRBRentitas/mu>
<http://itdi.elte.hu/FRBRentitas/mu/mucim>
szöveges érték: „Sötétség délben”
<http://itdi.elte.hu/FRBRentitas/kifejezesiforma>
<http://itdi.elte.hu/FRBRentitas/kifejezesiforma/kozremukodo>
http://itdi.elte.hu/szemely/Bart_Istvan

Minden, a fentiekhez hasonló háromelemű kijelentés értelmezhető egy gráfként is, ahol az alany és a tárgy egy-egy csomópontot alkot, a közöttük húzódó él pedig a tulajdonságot jeleníti meg. Mivel egyes állítások alanyai előfordulhatnak más kijelentések tárgyaiként, és fordítva: bizonyos állítások tárgyairól más kijelentések szólhatnak, a források tetszőleges méretű hálója jöhet létre, ahol az egyes csomópontok között a legkülönbélebb relációtípusok létezhetnek. Egy RDF-gráfot tehát forrásokról tett kijelentések sorozataként kell értelmezni, s amikor RDF-konverzióról esik szó, akkor az nem jelent mást, mint hogy egy bizonyos adatszerkezetet (például egy MARC-rekordot) állítások sorozatává kell átalakítani (Heath & Bizer, 2011).

A gráf adatszerkezet több szinten is értelmezhető. A legalsó, legelemibb szinten megállapítható, hogy egyetlen, három – vagy akár négy – elemből álló kijelentés (egyetlen metaadatközlés, technikailag pedig egyetlen triplet vagy quad) is értelmezhető olyan gráfként, melynek mindössze két csomópontja, és egyetlen, a két csomópontot összekötő éle van. De gráf keletkezik akkor is, amikor egyetlen, hagyományos formátumban tárolt MARC-rekord konverziója történik bizonyos, véges számú tripletté, amely az alkalmazott adatmodelltől függően egy vagy több entitás előfordulásainak leírását tartalmazza. Ez már az ún. metaadatleírás-halmaz (Dancs, 2020), amely számos csomópont és számos él segítségével reprezentálható. Harmadszor pedig gráfnak nevezik azt az adathalmazt is, amely egy intézmény *valamennyi* MARC-rekordjának entifikálásából és RDF-konverziójából keletkezik: ebben az esetben a diszkrét rekordok adattartalmából keletkezett állítások összessége *részgráfot* alkot.



10. ábra – Tripletek ábrázolása gráfként (saját szerkesztés)

Az RDF használatának számos előnye mutatkozik (Heath & Bizer, 2011):

- világméretben használható, lehetőséget biztosít arra, hogy bárki hivatkozhatson bármire;
- minden egyes triplet része a globális adathálózatnak, és minden triplet kiindulópontként használható e hálózat felfedezéséhez;
- az eltérő forrásokból származó adathalmazok könnyedén egymásba olvashatók;
- többféle séma (szótár) egyidejű használatát biztosítja (a szótárakról a 6.5.2. fejezetben még lesz szó);
- az RDF sémanyelveinek használatával (RDFS, OWL stb.) az adatszerkezet összetettsége könnyen szabályozható.

6.4. A tripletek kifejezésének formátumai

Az RDF-logika, vagy RDF-adatmodell megjelenítésére számos fájlformátum alkalmas, amelyeket szerializációs szintaxisoknak neveznek. Saját ízlés, illetve a partnerek igényei szerint lehet válogatni belőlük, hiszen alapvető különbség nincsen közöttük, mind ugyanazt a célt szolgálják: elemeket, valamint viszonyokat tesznek az URI-k segítségével a számítógépes folyamatok számára is jelentésteli módon feldolgozhatóvá.

Az *N-Triples* formátum a lehető legegyszerűbb megoldás. Az állítást alkotó azonosítókat relációs jelek között, egymás után, teljes alakjukban ki kell írni, majd a triplet végére pontot kell tenni. Ez a formátum nyomtatásban nehezen jeleníthető meg; a gépi feldolgozásra szánt állományban minden triplet sortörés nélkül szerepel (Allemang & Hendler, 2011).

Az *RDF/XML* a már jól ismert XML-formátumot bővíti ki, s teszi alkalmassá az állítások formalizált ábrázolására. A szemantikus web egyik legkedveltebb adatformátuma mind a mai napig, többek szemében e szintaxis jelenti magát az RDF-et (Sporny, 2012). Egy RDF/XML-dokumentum, hasonlóan a hagyományos XML-dokumentumpéldányokhoz, verziódeklarációval kezdődik:

```
<?xml version="1.0" encoding="UTF-8"?>
```

ezt pedig a dokumentumpéldányban alkalmazott névterek prefixumainak meghatározása követi:

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:bf="http://id.loc.gov/ontologies/bibframe/"
  xmlns:bflc="http://id.loc.gov/ontologies/bflc/"
  xmlns:madsrdf="http://www.loc.gov/mads/rdf/v1#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
```

E deklarációs rész biztosítja, hogy két vagy több, hasonló elnevezésű reláció definíciója ne keveredjen össze, ne képződjenek érvénytelen állítások, s minden kijelentés a névterekben kódolt reláció-szabályok szerint épüljön fel (Dudás, 2013). A névtérprefixumok alkalmazásának tárhely-takarékossági indokai is vannak. Az XML-dokumentumban alkalmazott minősített nevek állandó része (az ún. bázis-URI) minden alkalommal egy rövid karaktersorral helyettesíthető, így ahelyett, hogy minden alkalommal egy adott azonosító hosszú formáját kellene alkalmazni, például:

```
http://itdi.elte.hu/abszolutoriumIdeje
```

rövidebben is kifejezhető ugyanez az URI: `elte:abszolutoriumIdeje`, de csak abban az esetben, ha a deklarációs részben megjelenik a prefix értelmezése:

```
xmlns:elte="http://itdi.elte.hu/"
```

A deklarációkat követi a dokumentumpéldány leíró része:

```
<rdf:Description rdf:about="http://example.org/thing">
  <rdf:type rdf:resource="http://example.org/stuff/1.0/Document"/>
  <dc:title>A marvelous thing</dc:title>
</rdf:Description>
```

amelyet rendszerint az `rdf:Description` címke vezet be, majd az `rdf:about` címke jelöli meg azt a forrást, amelyről a további kijelentések szólnak. Minden további triplet alanya, legalábbis amíg meg nem történik a leírás lezárása az `</rdf:Description>` címkével, ez a forrás

lesz. A fenti példa a `http://example.org/thing` URI-val jelölt forrásról összesen két állításban közöl információt, megadja annak típusát (`rdf:type`), illetve címét (`dc:title`).

A 6.1. fejezetben említett tartalomjegyeztetési és átirányítási folyamat, amennyiben a kérés nem böngészőben megjeleníthető tartalomra, hanem gépi úton feldolgozható állományra irányul, RDF/XML állományt is visszaadhat a kérelmezőnek. Ennek a szintaxisnak az előnye, hogy – miként bármely más XML-állomány – könnyedén formázható, alakítható, továbbítható, illetve készíthető elő megjelenítésre. Előállítás is meglehetősen egyszerű abban az esetben, ha a kiinduló formátum is valamilyen XML-séma szerint készült dokumentumpéldány; ez teszi lehetővé például MARC/XML-állományok adattartalmának egyszerű átültetését RDF-logikára. Hátránya azonban, hogy az emberi szem számára kevésbé olvasható, nehezen áttekinthető. Ezért olyan munkafolyamatokban, ahol szükséges az adatszerkezet humán kontrollja, más szerializációs szintaxisok alkalmazása javasolt (Heath & Bizer, 2011).

A *Turtle* (Terse RDF Triple Language) az N-Triples átláthatóságát kombinálja a minősített nevek alkalmazásából eredeztethető tömörséggel (Allemang & Hendler, 2011). A leegyszerűsített deklarációs rész, illetve a leíró rész megvalósítása egy olyan formátumot eredményez, melyet különösen akkor érdemes használni, ha a gépi feldolgozhatóság mellett az adatok emberi értelmezésére is szükség van.

A Turtle névtér-deklarációja az alábbiak szerint történik:

```
@prefix bf: <http://id.loc.gov/ontologies/bibframe/> .
@prefix bflc: <http://id.loc.gov/ontologies/bflc/> .
@prefix madsrdf: <http://www.loc.gov/mads/rdf/v1#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix zs: <http://docs.oasis-open.org/ns/search-ws/sruResponse> .
```

a leíró rész szerkezete pedig az alább látható módon alakul:

```
<http://id.loc.gov/authorities/names/n82247773> a bf:Agent,
    bf:Person ;
    rdfs:label "Flanagan, Terry." ;
    bflc:name00MarcKey "1001 $aFlanagan, Terry." ;
    bflc:name00MatchKey "Flanagan, Terry." ;
    bflc:primaryContributorName00MatchKey "Flanagan, Terry." .
```

Minden kijelentés tehát, hasonlóan az RDF/XML-hez, a kiemelt pozícióban lévő URI-val jelölt forrásra vonatkozik. Az egyes kijelentések végét pontosvessző, a forrás leírásának végét az utolsó tripletet követő pont jelzi.

A *Trix* az RDF/XML-hez hasonló XML-séma, mely az egyes szerkezeti elemek megnevezésével dolgozik (Carroll & Stickler, 2004). Az értekezéshez felhasznált szakirodalom vizsgálata azt mutatta, hogy a Trix-szintaxis az elméleti cikkekben, valamint a gyakorlati implementációkban kevésbé használatos.

```
<TriX xmlns="http://www.w3.org/2004/03/trix/trix-1/">
<graph>
  <uri>http://example.org/graph4</uri>
<triple>
  <uri>http://example.org/aBook</uri>
  <uri>http://www.w3.org/2000/01/rdfschema#comment</uri>
  <plainLiteral xml:lang="en">This is a really good book!</plainLiteral>
</triple>
</graph>
```

Az RDFa (*RDF in Attributes*) formátum használatával lehetőség nyílik az adatelemeket értelmező jelölőket helyezni a HTML-formátumú weboldalak kódjába, amellyel mintegy felcímkézhetők a dokumentumban megjelenített információk. Alkalmasan megválasztott címkékészlet, valamint az RDFa használatával a nagyobb keresőszolgáltatások, mint például a Google, értelmezni tudják a weboldalakon közölt információt, és ezzel jóval informatívabb, részletgazdagabb formában teszik lehetővé a találati halmazban történő megjelenést (BP Digital, é.n.). Az RDFa formátumról a keresőoptimalizálásról szóló, 8. fejezetben még szó lesz.

A JSON-formátum a linked data-alapelvek megfogalmazása előtt született meg (W3C, 2020). Tisztán szöveges, programozási nyelvektől független formátum, struktúrája könnyen értelmezhető azok számára, akik járatosak a C, C#, C++ JavaScript és egyéb nyelvek használatában. A JSON név-érték párok rögzítésével ún. objektumok leírását teszi lehetővé.

```
{
  "name": "Manu Sporny",
  "homepage": "http://manu.sporny.org/",
  "image": "http://manu.sporny.org/images/manu.png"
}
```

Emberi szemmel könnyedén megfejthető a rövid adatszerkezet: egy személy nevét, weblapjának címét, illetve az őt ábrázoló digitális kép elérési útvonalát tartalmazza. A gépi értelmezés számára ezek az adatok azonban semmitmondók, ezért az objektumot leíró név-érték párokat URI-kkal kell egyértelművé tenni.

```
{
  "http://schema.org/name": "Manu Sporny",
  "http://schema.org/url":{"@id":"http://manu.sporny.org/"}, "http://schema.org/image":
  {"@id":"http://manu.sporny.org/images/manu.png"}
}
```

Mivel azonban az így nyert adatszerkezet túlságosan terjedős, ezért az ismétlődő bázis-URI-kat a JSON-dokumentum deklarációs részében kontextusként fel kell tüntetni:

```
{
  "@context": {
    "rdf": "http://www.w3.org/1999/02/22-rdf-syntax-ns#",
    "rdfs": "http://www.w3.org/2000/01/rdf-schema#",
    "xsd": "http://www.w3.org/2001/XMLSchema#",
    "bf": "http://id.loc.gov/ontologies/bibframe/",
    "bflc": "http://id.loc.gov/ontologies/bflc/",
    "madsrdf": "http://www.loc.gov/mads/rdf/v1#",
    "pmo": "http://performedmusicontology.org/ontology/",
    "datatypes": "http://id.loc.gov/datatypes/",
    "lclocal": "http://id.loc.gov/ontologies/lclocal/"
  },
```

A kontextus(ok) meghatározásával, valamint az URI-k hozzáadásával kis részben módosított JSON-adatszerkezetet, mivel ilyen módon alkalmassá válik az RDF-logika szerint megfogalmazott állítások hordozására, *JSON-LD*-nek (a *linked data* rövidítéséből) nevezik. Rob Sanderson megállapítása szerint ahhoz, hogy a legkülönfélébb módokon közzétett RDF-gráfok mások számára is használhatók legyenek, azaz ne csak LOD-ot, hanem LOUD-ot (linked open usable data – a rövidítés szellemesen az angol *hangos* szót adja ki) lehessen közzétenni, a JSON-LD-formátumot érdemes választani annak elérhetősége, valamint általános használhatósága okán (Sanderson, 2020).

A *HDT-formátumot* Javier D. Fernández és munkatársai mutatták be egy, a 2010-es évek első felében megjelent cikkükben (Fernández et al., 2013). Az elnevezés a formátumban alkalmazott hármass struktúra elemeinek – fejléc (Header), szótár (Dictionary), tripletek (Triples) – rövidítéséből származik. A fejléc tartalmazza a gráfra vonatkozó valamennyi leíró metaadatot, a szótár a gráfban előforduló valamennyi URI-t, szöveges értéket, illetve üres csomópontot listázza, míg a triplet-szakasz tömörítve tartalmazza a gráf tripletjeire vonatkozó információkat. A HDT alkalmazása, tömörített formátumról lévén szó, nagy adathalmazok kezelésekor tárhelyet, sávszélességet, illetve adatátviteli időt takaríthat meg, mivel a moduláris felépítésnek köszönhetően egyértelműen elkülöníti egymástól a szerkezeti elemeket. Ezen felül biztosítja az egyszerűbb adatműveletek hatékony végrehajtását, mivel hozzáférést biztosít a gráf fontos részleteihez, s ezzel lehetővé teszi azt, hogy a művelet a teljes adathalmaz átvizsgálása nélkül, gyorsabban történjen meg. Az ilyen tömörítés segítségével az eredeti adathalmaz mérete a tizenötödére csökkenhet: ezért a HDT-t a különféle adathalmazok archiválásakor is érdemes használni (Vander Sande et al., 2018).

Érdemes szólni továbbá az ún. *quadokról*, amelyek nem mások, mint egy kiegészítő provenienciainformációval négyeleművé bővített tripletek. Az alany-állítmány-

tárgy hármasan kívül az állítás negyedik tagja a kontextus (context), amely megmutatja, hogy az adott állítás melyik gráfhoz tartozik. A W3C ajánlásában (2014) az alábbi példát találjuk (a könnyebb áttekintés érdekében az eredetileg egy sorban közölt URI-kat széttördeltem):

```
http://example.org/#spiderman
http://www.perceive.net/schemas/relationship/enemyOf
http://example.org/#green-goblin
```

majd következik a kontextus URI-ja, amely mintegy megnevezi a gráfot:

```
http://example.org/graphs/spiderman
```

6.5. A linked data eszköztára

Tim Berners-Lee korábban (6.1. fejezet) említett négy ajánlása közül a negyedik arról rendelkezik, hogy az adatgazdák az általuk publikált adathalmazokban helyezzenek el hivatkozásokat más URI-kra, hogy az információt kereső gépi folyamat ezeket a linkeket követve más RDF-gráfokban talált állításokkal bővíthesse az eredményhalmazt (Berners-Lee, 2006). Noha valamennyi RDF-triplet egy-egy hivatkozást – még hozzá tipizált hivatkozást! – testesít meg. Haller és munkatársai (2020) tanulmánya alapján az az RDF-triplet tekinthető linknek, amelynek alanya, illetve állítmánya két külön adathalmazból származik, azaz, ha egy gráfban olyan URI-hármas van, amely egy eltérő névtérre mutat.

Haller és munkatársai (2020) e hivatkozásoknak, attól függően, hogy a külső névtérből származó URI az állítás melyik pozícióján foglal helyet, több típusát különítik el:

- *előfordulás-hivatkozás* (instance link), ha a külső URI a triplet harmadik pozíciójában, azaz tárgyként fordul elő;
- *osztályhivatkozás* (class link), amennyiben a külső URI, továbbá a triplet első pozíciójában – alanyként – előforduló URI valamilyen osztályt jelöl, és a triplet ezek között állít fel valamilyen (például hierarchikus, főosztály-alosztály) kapcsolatot;
- *előfordulás-minősítő hivatkozás* (instance typing link) abban az esetben, ha a triplet a saját adathalmaz valamilyen konkrét forrását – az alanyt – egy külső névtér-URI-val sorolja be egy osztályba, amely így a triplet tárgyát alkotja;
- *tulajdonsághivatkozás* (property link) az az eset, amikor a saját adathalmazban definiált tulajdonság valamilyen külső névtérben definiált, más tulajdonsággal áll

specifikált kapcsolatban (működése, funkciója tehát hasonló az osztályhivatkozásokéhoz);

- *előfordulás szerep-hivatkozás* (instance role link) épül fel akkor, ha a forrás valamely tulajdonságának leírása külső property-URI segítségével alakul tripletté.

A közzétett adathalmazokat 2007 óta regisztráló Linked Open Data Cloud szolgáltatás¹⁷ abban az esetben teljesíti a regisztrációs kérelmeket, ha a felhőbe felvetetni kívánt adathalmaz legalább 50 RDF-hivatkozással csatlakozik olyan halmazokhoz, melyek már a felhő részét képezik. Ezzel biztosítható az adathálózat szövetének koherenciája, és így maximalizálható a szemantikus web böngészésével visszakapható információ mennyisége.

A következőkben áttekintem azokat a segédeszközöket, amelyek nélkülözhetetlenek minden adatgazda számára, aki a rendelkezésére álló tudásvagyont linked dataként szeretné közzétenni a szemantikus világhálón.

6.5.1. Szótárkódolási sémák (névterek)

Az adathálózaton mindenekelőtt az egyes entitások konkrét előfordulásainak azonosítóra van szükség. Ezek az URI-k találhatóak meg a világhálón hozzáférhető számos *szótárkódolási sémában*, avagy *névtérben*: egyes névterek személyneveket tartalmaznak, mások testületi nevek vagy művek URI-jait szolgáltatják. Jelen értekezésben szótárkódolási sémaként tekintek mindazon azonosítóhalmazokra is, amelyek bizonyos kötött szótáras MARC-mezők választható értékeit reprezentálják, hiszen ezekben az esetekben egyazon entitás (például az információhordozó, a bibliográfiai forrás beszerzésének módja stb.) előfordulásai jelennek meg. A 6.2. fejezetben említett alapelv, a nem egységes elnevezések feltevése miatt várható, hogy egyetlen entitás-előfordulásnak, például Kosztolányi Dezsőnek több különféle névtérben is létezik egy-egy azonosítója, amelyek a 6.6. fejezetben tárgyalandó *adatgazdagítási* folyamat során összekapcsolhatók egymással, illetve azzal a – saját intézményi adatok publikálása céljából létrehozott – URI-val, amely ugyancsak Kosztolányi Dezsőt reprezentálja. Az azonosságok minél szélesebb körű explicit kifejezése a garancia arra, hogy az információkereső a lehető legtöbb tájékoztatást kapja az adott entitás-előfordulásról. Az URI-k ráadásul egyértelmű megkülönböztetéseket tesznek lehetővé a gépi információfeldolgozási folyamat számára, segítségükkel orvosolhatóvá válik például a homonimák, az azonos alakú, de eltérő jelentésű kifejezések problémája.

¹⁷ Hozzáférhető a <http://lod-cloud.net> címen [letöltés: 2022. május 19.]

Néhány példa a könyvtári, és szélesebb körben vett közgyűjteményi adatok publikálására (is) felhasználható névterek világából:

- *VIAF*¹⁸ - ha személynevek, testületi és földrajzi nevek, továbbá művek és kifejezési formák azonosítóit szeretnénk megtalálni, a VIAF szinte megkerülhetetlen szolgáltatás. Ez a hatalmas névadatbázis az OCLC kezdeményezésével és vezetésével jött létre, nemzeti könyvtárak – köztük az Országos Széchényi Könyvtár – és más partnerintézmények által épített authority állományok összeolvasztásával. A többé-kevésbé adatgazdag authority rekordok adattartalmát – beleértve a nyelvterületenként eltérő, de azonos személyt reprezentáló autorizált névalakokat is – egyetlen meghatározott, perzisztens azonosítóhoz rendelték, amely a kérés típusától függően HTML-, illetve RDF/XML, JSON, vagy akár MARC/XML-dokumentumhoz is tovább tudja irányítani a használót;
- *Nemzeti Névtér*¹⁹ – a jelenleg bétaverzióban működő szolgáltatás három entitás – földrajzi nevek, személynevek, testületi nevek – azonosítóit is rendelkezésre bocsátja. Feladata, hogy a különféle intézmények által kezelt adatvagyonot egyetlen felületen hozzáférhetővé téve kulturális aggregációs platformként működjön. E célját úgy teljesíti, hogy az egyes entitás-előfordulásokat igen széles látókörrel, történelmi nézőponttal kiegészítve, több intézményből származó, hiteles adatokra támaszkodva mutatja be. A Nemzeti Névtér a tervek szerint összekapcsolható lesz az integrált gyűjteménykezelő rendszerekkel, ezáltal a közgyűjteményekbe hiteles információkat tartalmazó authority rekordok kerülhetnek. Ezek az információk egyelőre MARC/XML-ben mozgathatók a rendszerek között, a jövőben a szolgáltatási platform várhatóan más formátumok felé is bővíteni fog (*A Nemzeti Névtér értelméről és hasznáról*, é.n.);
- A Library of Congress linked data-szolgáltatása²⁰ széles körben ajánl URI-kat a bibliográfiai adatok szemantikus publikálásához. Érdeemes megemlíteni jelentős méretű, igen sok entitás-előfordulást magába foglaló authority állományát, a National Authority File-t, és számos, a MARC 21 szótáras mezőinek értékeihez kapcsolódó azonosítót. Adatai a legtöbb RDF-szintaxisban hozzáférhető;

¹⁸ Hozzáférhető a <https://viaf.org> címen [letöltés: 2022. május 19.]

¹⁹ Hozzáférhető a <http://abcd.hu> címen [letöltés: 2022. május 19.]

²⁰ Hozzáférhető a <https://id.loc.gov> címen [letöltés: 2022. május 19.]

- *RDA Registry*²¹ – az RDA forrásleírási szabályzat rendelkezéseihez köthető értékszótárak elemeinek definícióit, illetve azonosítóit szolgáltatja. A világhálón N-Triples-t, RDF/XML-t, illetve JSON-LD-t képes továbbítani;
- *Who's On First*²² – igen részletes földrajzinév-adatbázis, böngészésre szánt felülete térképes megjelenítőt is tartalmaz. Az információt JSON-ra alapuló, speciális formátumban továbbítja (GeoJSON);
- *Wikidata*²³ – a világszerte ismert Wikipédia mellett létező, azzal szorosan összefonódó adatbázis, melyben az információt nem szócikkek, hanem strukturált adatlapok tárolják. A Wikidata alapegységei az *elemek*. Elemeket bárki szabadon létrehozhat, a tulajdonságok azonban kötött szótárból választhatók, amelyet csak kérésre bővítenek (Wikidata, é.n.) A Wikidata-elem, illetve az elemet szöveges formában leíró Wikipédia-szócikk között kapcsolat alakítható ki: számos szerkezeti elem, mely a szócikkekben előfordul – ilyen például a cikkekben található, alapvető adatokat tartalmazó infobox, vagy a Nemzetközi katalógusok sablon – egyedileg konfigurálható módon a Wikidatában tárolt elemleírás bizonyos részeit jeleníti meg.

6.5.2. Szótárak

A szemantikus web működtetéséhez a különféle entitás-előfordulások azonosítókkal való helyettesítése nem elegendő: az egyes elemeket összefűző kapcsolatok még hiányoznak a rendszerből. A forrásleíró keretrendszer lényege az, hogy két entitás-előfordulást egymással valamilyen módon összefüggésbe hozzon. Az értelemmel teli világhálón ezeket az összefüggéseket is http-alapú azonosítók reprezentálják, amelyek gondosan megszerkesztett együttesét *szótárnak*, *szemantikus elemkészletnek* vagy *ontológiának* nevezik.

Tom Gruber közismert definíciója alapján az ontológia nem más, mint egy „*megegyezésen alapuló fogalmi rendszer formális, egyértelmű leírása.*” John Frederick Sowa jellemzése alapján az ontológia „*valamilyen tárgykörben létező vagy feltételezett dolgok kategóriáinak elmélete*” (Ungváry, 2005), utalva ezzel a fogalom eredeti, filozófiai értelmezésére, mely szerint az ontológia a létezés természetének elmélete, annak meghatározása, hogy milyen típusú dolgok léteznek. Noy és McGuinness (é.n.) szerint az

²¹ Hozzáférhető a <https://www.rdaregistry.info> címen [letöltés: 2022. május 19.]

²² Hozzáférhető a <https://spelunker.whosonfirst.org/> címen [letöltés: 2022. május 19.]

²³ Hozzáférhető a <https://wikidata.org> címen [letöltés: 2022. május 19.]

ontológia számítógéppel feldolgozható értelmezéseket kínál egy bizonyos területen előforduló fogalmak és kapcsolataik számára. Ungváry Rudolf (i.m.) szerint a mesterségesintelligencia-kutatás területén már a hetvenes években ontológiának hívták a szoftverekkel megalkotott mesterséges világokat, a 90-es években pedig már olyan generikus hierarchiaszerkezeteket kezdtek el ontológiának nevezni, amelyeket formális logikai leírások, elsőfokú logikai kijelentések egészítenek ki.

Alemu és szerzőtársai (2012) úgy fogalmazzák, hogy az ontológiák egy bizonyos területről szóló tudást segítenek megszerezni – és aztán megosztani a humán és gépi ágensek között – fogalmak és a köztük lévő kapcsolatok segítségével. Ezt a meghatározást mindenképp ki kell egészíteni, mivel az ontológiák nem csupán a tudás rögzítésére alkalmasak, hanem segítségével a gépi információfeldolgozás automatikus következtetés-végrehajtással egészülhet ki. Így az egyes tulajdonságok hierarchikus viszonyai bejárhatóvá válnak, abban az értelemben, hogy az ontológia expliciten tartalmazza az egyes tulajdonságok fölé-, alá-, illetve mellérendeléseit. Például, ha valaki *barátja* valakinek, akkor az ontológiában rögzített formális tudás segítségével kikövetkeztethető – nem csupán emberi, hanem gépi úton is –, hogy egészen biztosan *ismeri* is. A következtetésben segít az informatikában (főként az objektumorientált programozási irányzatban) nagyon gyakran felhasznált sajátosság, az ún. generikus öröklődés, amely szerint, ha egy bizonyos entitásra (például a *személyre*) jellemző valamilyen tulajdonság (van *életkora*), akkor az az adott entitás minden konkrét előfordulására is igaz lesz.

Ontológiát – amely valamilyen, az alkalmazott nyelvtől függő kiterjesztéssel bíró, digitális állományként ölt testet – úgy kell építeni, hogy formálisan definiáljuk benne az egyes, leírni kívánt entitásokat (ezeket az ontológiák világában inkább osztályoknak nevezik), az ezek közötti kapcsolatokat, továbbá az egyes osztályok tulajdonságait is, figyelembe véve azok hierarchikus viszonyait. Technikai oldalról nézve az RDF, az RDF Schema, illetve az OWL ontológiaépítő nyelv elemeinek felhasználásával előállított szótárak hasonló szerepet töltenek be, mint az XML DTD-k, illetve XML-sémák: segítségével készülnek a valid dokumentumpéldányok, illetve érvényes RDF-gráfok.

A szótárépítés első lépéseként meg kell határozni az adott terület legfontosabb osztályait és ezek alosztályait is, amelyekbe majd az egyedi RDF-gráfokban leírt konkrét entitás-előfordulások kerülnek. Az osztályra egy sajátos webes forrásként kell tekinteni, mely ugyan kézzelfoghatóan nem létezik, mégis azonosíthatóvá, egyértelműen

elkülöníthetővé válik a saját URI-ja által. Osztályt az OWL segítségével az alábbi módon lehet kialakítani:

```
<owl:Class rdf:about="http://id.loc.gov/ontologies/bibframe/Work"></owl:Class>
```

alosztály definiálásakor pedig a fölé- illetve alárendelt osztályt is egy-egy URI jelképezi:

```
<owl:Class rdf:about="http://id.loc.gov/ontologies/bibframe/Cartography">  
<rdfs:subClassOf rdf:resource="http://id.loc.gov/ontologies/bibframe/Work"/></owl:Class>
```

Ha kialakult az egyes osztályok hierarchiája a szótárban, a következő lépés az azokat jellemző attribútumok, tulajdonságok rögzítése. Az RDF szótára a tulajdonságokat az `rdf:Property` segítségével, tipizálatlan módon teszi leírhatóvá. Az OWL segítségével már lehetőség van a differenciálásra: a rögzíthető attribútumok többsége ún. objektumtulajdonság (*ObjectProperty*), amelyek definiálásakor elsődlegesen két dolgot kell meghatározni: a tulajdonság értelmezési tartományát (*domain*), illetve azon értékek készletét, melyet a tulajdonság felvehet (*range*). Az értékészlet lehet valamilyen osztály tagja (az objektumtulajdonságok esetében), illetve literál (az ún. adattípus-tulajdonságok, *datatype property* esetében). Egy szótárban definiált tulajdonság, például a *gyűjtőkör* vonatkozhat könyvtárakra, múzeumokra, levéltárakra, cirkuszok jellemzésére azonban nem használható. A *lejátszási idő* tulajdonság vonatkozhat mozgóképekre és hangfelvételekre, értéke pedig valamilyen mértékegység szerint kifejezett időtartam lehet. Az alábbi példa a BIBFRAME-szótárban definiált `bf:hasItem` tulajdonság egyes paramétereinek rögzítését mutatja be:

```
<owl:ObjectProperty rdf:about="http://id.loc.gov/ontologies/bibframe/hasItem">  
<rdfs:domain rdf:resource="http://id.loc.gov/ontologies/bibframe/Instance"/>  
<rdfs:range rdf:resource="http://id.loc.gov/ontologies/bibframe/Item"/>  
<rdfs:subPropertyOf rdf:resource="http://id.loc.gov/ontologies/bibframe/relatedTo"/>
```

Az ún. *inverz tulajdonságok* explicit modellezése ugyancsak szükséges. Noha az emberi értelem könnyűszerrel megállapítja, hogy ha X az *édesanyja* Y-nak, akkor Y a *gyereke* X-nek, az elemkészlet szerkesztésekor az ilyen következtetési irányok megadására is tekintettel kell lenni, hiszen ezek a gépi adatfeldolgozás számára egyáltalán nem egyértelműek (Allemang & Hendler, 2011). Az alábbi példában saját alkotású URI-kkal szemléltetem ezt a megoldást:

```
<owl:ObjectProperty rdf:about="http://itdi.elte.hu/relacio/isTeacherOf">  
<owl:inverseOf rdf:about="http://itdi.elte.hu/relacio/isStudentOf">
```

E két rövid sor azt mutatja be, hogy az *X Y-nak a tanára* tulajdonság fordítottja az *Y X-nek a diákja*. Egy tripletnek elegendő csupán az egyiket tartalmaznia:

```
http://itdi.elte.hu/szemely/Kovacs_Jozsef
http://itdi.elte.hu/relacio/isTeacherOf
http://itdi.elte.hu/szemely/Barna_Lorant
```

a gépi információkeresési folyamat az ontológiában rögzítettek alapján új tudást képes előállítani: a *tanára* viszonyból következteti ki a *diákja* reláció résztvevőit. Vannak azonban olyan tulajdonságok (ezeket szimmetrikus tulajdonságoknak nevezik), például a *kapcsolódik hozzá*, melyek alanyát és tárgyát felcserélve az állítás változatlanul érvényes marad. Az ilyen tulajdonságokat a szótár kialakításakor ugyancsak jelölni kell.

Lehetőség van egy általános relációtípuson belül specifikusabb viszonyok meghatározására is, ún. altulajdonságok (subproperty) bevezetésével. Ha X a *tanára* Y-nak, akkor X *munkakapcsolatban áll* Y-nal, ez utóbbi tulajdonság tágabb terjedelmű az előzőnél. Az ontológia alapján új tudás állítható elő, ha egy triplet az *isTeacherOf* állítást tartalmazza.

```
<owl:ObjectProperty rdf:about="http://itdi.elte.hu/isTeacherOf">
  <owl:subPropertyOf rdf:about="http://itdi.elte.hu/worksWith">
```

Az *rdfs:label* tulajdonság értékével az URI-k emberi szemmel is olvasható címkékkel ruházhatók fel – egyetlen URI akár több nyelven is felcímkézhető, így az adatokat az emberi szem számára megjelenítő szoftvereszköz (például böngészőn keresztül használható felhőalkalmazás) a helyi beállításoknak megfelelő nyelvű címkét jeleníti meg, ha az az URI-hoz társítva rendelkezésre áll (Allemang & Hendler, 2011). E lehetőséget inkább a szótárkódolási sémák esetében használják ki.

```
<owl:Class rdf:about="http://bibliotek-o.org/1.1/ontology/CinematographerActivity">
  <rdfs:label xml:lang="en">Cinematographer</rdfs:label>
```

6.5.3. Könyvtártudományi szótárak

A fenti adatmodellezési eszközök segítségével a világ számos tudományterülete, köztük a könyvtártudomány is leírható. Ehhez meg kell határozni a diszciplínában előforduló legfontosabb relációkat, hogy azokat tripletbe foglalva valid kijelentéseket lehessen tenni a leírt bibliográfiai forrásokról, illetve az azokhoz kapcsolódó entitás-előfordulásokról. A nem egységes elnevezések feltevése (lásd a 6.2. fejezetet) a szótárakra is érvényes: előfordul, hogy egyazon relációt vagy osztályt több szótár is definiál: ezekben az esetekben tetszőlegesen (de körültekintően!) választhatunk közülük. Nem szabad azonban abba a hibába esni, hogy pusztán az osztályok vagy relációk névalakjának azonossága miatt tartunk

két szótárelemet egymás helyettesítőinek. Az egyes adatmodellek néha teljesen máshogy értelmezik ugyanazokat a fogalmakat, így az egyik modellből a másikra való, körültekintő óvatosság nélküli áttérés téves kapcsolatok felépülését, vagy épp ellenkezőleg, összefüggések elvesztését eredményezheti. A disszertáció 9.5.2. fejezetében részletesebben megvizsgálók egy ilyen esetet.

A szemantikus webes technológiát alkalmazó metaadatolás története a könyvtári területen a *Dublin Core*-ral kezdődik. Az eredeti, 15 elemet tartalmazó metaadatelemkészletet – cím, szerző, kiadó, nyelv, formátum stb. – webdokumentumok metaadatolására fejlesztették ki az 1990-es évek közepén; ezeket az elemeket a weboldalak kódjába kellett elhelyezni. Egyes repozitóriumszoftverek, mint például a Monguz Információtechnológiai Kft. által fejlesztett JaDoX, vagy a közismert DSpace, bármely digitális objektum (kép, hangfelvétel, mozgókép, szöveges dokumentum) metaadatait a Dublin Core-elemek használatával teszik rögzíthetővé. Az ezredfordulón a tizenöt elemhez URI-t alkottak (DCMI, é.n.), s a Dublin Core-ból vett példák hamarosan az RDF alapidokumentumainak állandó részeivé váltak (W3C, 2004). Így a DC tekinthető az első, könyvtári felhasználásra (is) alkalmas szótárnak; a közös cél miatt egyáltalán nem meglepő, hogy a két igen fiatal technológia (2000-ben a DC öt, az RDF pedig mindössze két esztendő volt) egymásra talált.

A Dublin Core Metadata Initiative napjainkban négy szótárat gondoz, ezek közül a legterjedelmesebbet DC Terms-nek nevezik. Összesen 55 elemet tartalmaz, s URI-jait ma is igen sok gráf alkalmazza adminisztrációs vagy leíró metaadatok rögzítésére – megfelelően ezzel az URI-k újrafelhasználására vonatkozó javaslatnak. Az ausztrál nemzeti könyvtár által üzemeltetett aggregátor portál, a Trove, a különféle forrásokból érkező adatok megjelenítésére a DC eredeti elemkészletén kívül a kibővített DC Terms szókészletből is több relációt használ (Trove, 2019).

Egyes elemeknek, melyek már az eredeti, tizenöt elemet számláló készletben is benne voltak, két, a DCMI által definiált URI-ja is van, s csak látszólag mindegy, melyik kerül az adatszerkezetekbe: a <http://purl.org/dc/elements/1.1/date>, illetve a <http://purl.org/dc/terms/date> azonosító igénybevételekor tekintettel kell lenni arra, hogy az utóbbinak kötött – literál típusú – értékészlete van, míg az előbbinél nincsen ilyen korlátozás (*DCMI Metadata Terms*, é.n.). Példaként láthatók a Creator metaadatelem meghatározása és ismérvei, amelyet a szótár az egyszerű tulajdonságok (`rdfl:Property`) osztályába sorol be:

```

<http://purl.org/dc/elements/1.1/creator>
  dcterms:description "Examples of a Creator include a person, an organization, or a
  service. Typically, the name of a Creator should be used to indicate the entity."@en ;
  dcterms:issued "1999-07-02"^^<http://www.w3.org/2001/XMLSchema#date> ;
  a rdf:Property ;
  rdfs:comment "An entity primarily responsible for making the resource."@en ;
  rdfs:isDefinedBy <http://purl.org/dc/elements/1.1/> ;
  rdfs:label "Creator"@en ;
  skos:note "A [second property] (/specifications/dublin-core/dcmi-
  terms/#http://purl.org/dc/terms/creator) with the same name as this property has been
  declared in the [dcterms: namespace] (http://purl.org/dc/terms/). See the Introduction to
  the document [DCMI Metadata Terms] (/specifications/dublin-core/dcmi-terms/) for an
  explanation."@en .

```

A következő, könyvtári célokra alkalmazható szótár létrehozására tett kísérlet – még mindig Tim Berners-Lee definitív írásának megjelenése előtt – 2005-ben történt. A MarcOnt megalkotói azzal a céllal láttak neki a munkának, hogy áthidalják vele a bibliográfiai információk tárolására akkoriban használt három legfontosabb formátum (DC, MARC 21, BibTex) közötti szakadékot. A bemeneti formátumokat először XML-be, majd XSLT használatával RDF-re konvertálták, amelyből előre definiált átalakítási szabályok segítségével született meg a végleges, MarcOnt-verzió – az átalakításban többször szerepelnek az egyik korai, általános személyi és kapcsolati adatokat leíró szótár, a FOAF relációi is. A MarcOnt 34 osztály és 43 tulajdonság segítségével igyekszik megőrizni a bemeneti formátumok gazdag információtartalmát. E szótár alkalmazására épült egy digitális könyvtári platform, a JeromeDL is (Kruk et al., 2007)

A következő példában egy objektumtulajdonság definíciója látható, amely a szerzői jogok osztályának tagjaival vehető használatba, értéke pedig egy olyan ágens, aki/amely a FOAF szótár megfelelő osztályának tagja:

```

<owl:ObjectProperty rdf:ID="hasOwner">
  <rdfs:label rdf:datatype="http://www.w3.org/2001/XMLSchema#string">hasOwner</rdfs:label>
  <rdfs:domain rdf:resource="#Copyright"/>
  <rdfs:range rdf:resource="http://xmlns.com/foaf/0.1/Agent"/>
  <rdfs:comment rdf:datatype="http://www.w3.org/2001/XMLSchema#string">Represents the owner of
  the copyright.</rdfs:comment>
</owl:ObjectProperty>

```

Frédéric Giasson és Bruce D’Arcus 2007-ben készítette el saját szótárát, amely a BIBO nevet kapta (*Bibliographic Ontology*). Elképzelésük szerint a BIBO-nak hármas célt kell szolgálnia: használható kell legyen idézetek, hivatkozások nyilvántartására, a dokumentumok osztályozására, és egyszerűen, bármely dokumentum jellemzőinek RDF-ben történő rögzítésére (D’Arcus & Giasson, 2009). Nagyon sok, könyvtári források leírásához használható relációt a BIBO definiált először, így – más, fejlettebb, összetettebb szókészletek hiányában – a korai adatpublikációs kísérletek (például a LIBRIS) nagy mértékben támaszkodtak rá (Malmsten, é.n.)

A lenti példában a *címzett* objektumtulajdonság leírása látható, RDFS-sémanyelvi címkék, és egy külső névtér, az `ns:` prefix-szel jelölt W3C Vocab Status Ontology felhasználásával. A tulajdonság értelmezési tartománya egy ugyancsak a BIBO-ban kialakított osztály, a *PersonalCommunicationDocument*, értékkészlete pedig – hasonlóan az előző példához – egy FOAF-ágens.

```
<owl:ObjectProperty rdf:about="&bibo;recipient">
  <rdfs:label>recipient</rdfs:label>
  <rdfs:isDefinedBy
rdf:datatype="&xsd:anyURI">http://purl.org/ontology/bibo/</rdfs:isDefinedBy>
  <rdfs:comment xml:lang="en">An agent that receives a communication
document.</rdfs:comment>
  <ns:term_status>stable</ns:term_status>
  <rdfs:domain rdf:resource="&bibo;PersonalCommunicationDocument"/>
  <rdfs:range rdf:resource="&foaf;Agent"/>
</owl:ObjectProperty>
```

Az RDA forrásleírási szabályzathoz kapcsolódó szótárkodolási sémákat a 6.5.1. fejezetben már röviden bemutattam, érdemes azonban megemlíteni, hogy ezek mellett URI-kkal definiálták a szabályzat kontextusában előforduló valamennyi kapcsolattípust is. Ezeket a relációazonosítókat ugyancsak az RDA Registry weboldal tartalmazza, amely így tehát egyszerre szolgál információkkal a szabályzat technikai megvalósítását szolgáló egyes entitás-előfordulásokról, valamint az összekötésüket lehetővé tévő, értelmezett kapcsolóelemekről. Az RDA-rendelkezések alapján leírandó entítások valamennyi tulajdonságának azonosítója megtalálható itt, amelyeket a bibliográfiai információk gráfban történő tárolásakor fel lehet használni. Ilyen reláció például a P30063 azonosítójú „*has note on title*”. Az elem címkéje mellett annak rövid definícióját, értelmezési tartományát, fölé- és alárendelt elemeit, esetleges alternatív megnevezéseit, valamint az RDA Toolkitben alkalmazott, kis mértékben eltérő elemcímkét, és meghatározást olvashatjuk. A P70004 azonosítójú, *has related work of place* elem esetében a hozzárendelhető érték – bár ezt explicit módon nem jelöli a Registry – a művek osztályából választható, amelyet az elem címkéje, illetve definíciója tesz egyértelművé.

Hasonló összetett struktúrában tették közzé az Open Metadata Registry-ben a MARC 21-ben előforduló, valamennyi értékshótáras mező kötött értékeihez tartozó URI-kat, illetve a MARC-szerkezetben hívójelekkel, almezőjelekkel, illetve indikátorértékekkel jelölt relációtípusoknak megfeleltetett azonosítókat.²⁴ Az így előálló, csaknem 11.000, sorszámot és kategóriajelölő betűt tartalmazó, önmagában kevésbé beszédes URI-val a MARC-rekord információtartalma nagyon alaposan átültethető valamilyen, RDF-logikát támogató

²⁴ Hozzáférhető a <http://www.marc21rdf.info> címen [letöltés: 2022. május 19.]

formátumra, és ezzel értelemmel dúsítható, ugyanakkor nem valósítható meg benne megfelelően az entitások elkülönítése, és nem visz közelebb a forrásleírás entifikálásához, ezen keresztül pedig a megnövelt hatékonyságú információkereséshez. A példában a 300\$a almezőnek megfelelő tulajdonság definíciója látható:

```
<rdf:Description rdf:about="http://marc21rdf.info/elements/3XX/M300_a">
  <rdfs:isDefinedBy rdf:resource="http://marc21rdf.info/elements/3XX" />
  <reg:status rdf:resource="http://metadataregistry.org/uri/RegStatus/1001" />
  <reg:name xml:lang="en">hasExtentInPhysicalDescription</reg:name>
  <rdfs:label xml:lang="en">Extent in Physical Description</rdfs:label>
  <skos:definition xml:lang="en">Number of physical pages, volumes, cassettes, total playing
time, etc., of of each type of unit.</skos:definition>
  <rdf:type rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-ns#Property" />
  <skos:scopeNote xml:lang="en">tag: 300 (Physical Description)&lt;br />ind1: # (no
caption)&lt;br />ind2: # (no caption)&lt;br />sub: a (Extent)&lt;br /></skos:scopeNote>
</rdf:Description>
```

Számos szótár készült az FRBR-ben megfogalmazott entitások figyelembevételével is. A funkcionális modell alapszintű, 13 osztályt és 48 tulajdonságot tartalmazó elemkészletét Ian Davis és Richard Newman készítette 2005-ben,²⁵ olyan osztályokat is bevezetve, melyeket az FRBR nem írt le: ilyen osztály például az *Endeavour* (~törekvés), amelyet a szótár alkotói a négy alapentitásnak megfelelő osztály fölérendelt osztályaként határoztak meg, ennek megfelelően minden egyes mű, kifejezési és megjelenési forma, valamint példány egyben *Endeavour* is.

Az alábbi példa a Mű entitás definíciójának egy részletét mutatja. Az owl:disjointWith reláció segítségével pontosan megadták azokat az osztályokat (vagy halmazokat), amelyek diszjunktak a művek halmazával.

```
<owl:Class rdf:about="http://purl.org/vocab/frbr/core#Work">
  <rdfs:label xml:lang="en">work</rdfs:label>

  <skos:definition xml:lang="en">An abstract notion of an artistic or intellectual
creation.</skos:definition>
  <rdfs:comment xml:lang="en">This class corresponds to the FRBR group one entity
'Work'.</rdfs:comment>
  <rdfs:isDefinedBy rdf:resource="http://purl.org/vocab/frbr/core#" />
  <rdfs:subClassOf rdf:resource="http://purl.org/vocab/frbr/core#Endeavour" />
  <rdfs:subClassOf rdf:resource="http://xmlns.com/wordnet/1.6/Work~2" />
  <owl:disjointWith rdf:resource="http://purl.org/vocab/frbr/core#Expression" />
  <owl:disjointWith rdf:resource="http://purl.org/vocab/frbr/core#Manifestation" />
  <owl:disjointWith rdf:resource="http://purl.org/vocab/frbr/core#Item" />
  < dct:issued>2005-07-15</dct:issued>
</owl:Class>
```

Silvio Peroni és David Shotton 2012-ben publikálta a FaBiO (FRBR-aligned Bibliographic Ontology) és CitO (Citation Typing Ontology) nevű szótárakat (Peroni &

²⁵ E szótár megtekinthető a <https://vocab.org/frbr/core> oldalon.

Shotton, 2012); előbbivel választ kínálva a BIBO-ban, illetve az FRBR Core-szótárban talált modellezési problémákra. A példában a könyvfejezet osztály definíciója történik meg:

```
:BookChapter
  a owl:Class ;
  rdfs:label "book chapter"@en ;
  rdfs:comment "A defined chapter of a book, usually with a separate title or number."@en ;
  rdfs:subClassOf [ a owl:Restriction ;
    owl:onProperty frbr:partOf ;
    owl:someValuesFrom :Book ],
  :Chapter .
```

A GNDO a Deutsche Nationalbibliothek munkatársai által kidolgozott szótár²⁶, amely a Gemeinsame Normdatei néven ismert authority adatbázisban tárolt adatok publikálása céljából született meg. A Gemeinsame Normdatei egy közösségi használatban lévő authority adatbázis, amely személynevek, testületi nevek, rendezvénynevek, földrajzi nevek és egységesített címek MARC 21-formátumú rekordjait tartalmazza. A folyamatosan bővülő állományt a DNB weblapján szabadon letölthetővé tették, ahol az adatok rekordtípusonként különválogatva, egy-egy nagyméretű, csomagolt állományban menthetők a felhasználó számítógépére MARC 21, illetve MARC/XML formátumban is.

A szemantikus web formátumaira épülő adatszolgáltatás érdekében a könyvtár munkatársai saját szótárat dolgoztak ki, amelyben 63 osztályt definiáltak – például testület, személy, család, nyelv, de olyan, szűkebb csoportokat is kialakítottak, mint például az isten, nyelv, karakter vagy morféma, történelmi esemény vagy korszak stb. Ezekhez az osztályokhoz csaknem 160 különféle objektumtulajdonság tartozik, amelyeket adattípus-tulajdonságok, illetve annotációs tulajdonságok egészítenek ki. Ezek segítségével igen komplex adatszerkezeteket, azaz gráfokat állítottak elő a különféle típusú authority rekordokból, amelyeket ugyancsak letölthetővé tettek a legtöbb, RDF-logikát támogató formátumban – RDF/XML-ben, Turtle-ben, JSON-LD-ben, sőt még HDT-ben is. A szótárelemek megnevezése beszédes, helyenként szinte mulatságosan hosszú URI-kat képeztek:

<https://d-nb.info/standards/elementset/gnd#preferredNameForThePlaceOrGeographicName>

Ugyanakkor előnye a szótárnak, hogy az egyes elemek definíciójába, amennyiben lehetséges volt, RDA-, illetve FOAF-osztályokat illesztettek, amelyeket azonossági, illetve alá-fölérendeltségi relációkkal társítottak a GNDO-osztályokhoz. A MARC-hoz való erős kötődést jelzi, hogy a legtöbb osztály, illetve tulajdonság definíciójában megadták az adott

²⁶ Hozzáférhető a <https://d-nb.info/standards/elementset/gnd> oldalon [letöltés: 2022. május 19.]

relációnak megfelelő MARC-hívójelet, sőt az almezőket is. A példában a testületi név osztály meghatározása látható az elemkészletben:

```
:CorporateBody a          rdfs:Class , owl:Class ;
  rdfs:isDefinedBy      : ;
  rdfs:label             "Körperschaft"@de , "Corporate Body"@en ;
  rdfs:subClassOf        :AuthorityResource , http://rdaregistry.info/Elements/c/C10005> ;
  owl:equivalentClass  foaf:Organization ;
  :marc21equivalent     "079 $v=kiz" , "079 $b=b" .
```

6.6. Az adatgazdagítás

A különféle adatszolgáltatók által előállított gráfok elszigeteltek maradnak, ha a gráfot alkotó tripletek kizárólag saját készítésű bázis-URI-kat tartalmaznak. Ahhoz, hogy a világ tetszőleges pontján előállított RDF-adatszerkezetekből igazi adathálózat, szemantikus web jöjjön létre, a különálló gráfoknak hivatkozásokat kell tartalmazniuk egymásra, miként azt a Tim Berners-Lee által megfogalmazott adatközzételi ajánlások is tartalmazzák. Az intézmények által gondozott gráfokban szereplő azonosítókat – lehetőleg minél többet – meg kell feleltetni a más gráfokban található, az intézményi adathalmazban jelölttel azonos entitás-előfordulást leíró URI-kkal, azaz párhuzamosságokat kell kimutatni a különféle adatszerkezetek között. A dokumentumok hálózatán, azaz a hagyományos világhálón ezt minél több, az adott tárgyra (például Arthur Koestlerre) vonatkozó weboldal hivatkozásának összegyűjtésével végzik el, ezáltal segítve a webet böngésző felhasználót, hogy minél több információt találhasson a híres, magyar származású szerzőről. A szemantikus web, azaz a tripletek világában az azonos entitás-előfordulást más és más gráfokban leíró URI-k kapcsolhatók össze egymással: ezt a munkafolyamatot nevezik adatgazdagításnak (*data enrichment*, *data enhancement* vagy *data reconciliation*). A *reconciliatio* latin kifejezés helyreállítást, kibékülést jelent, amely az adatgazdagítás kontextusában annak tudomásul vételét jelenti, hogy egyazon entitás-előfordulás több különféle néven is ismert lehet, és ezek a névkülönbségek kulturális különbségekből, nyelvészeti kérdésekből ugyanúgy eredhetnek, mint katalogizálási szabályokból és sajátosságokból, vagy egész egyszerűen gépelési hibákból. Az adatgazdagítás folyamatában e sokféleség értékének felismerése és elfogadása, és az egymással egyenrangúnak minősített névalakok egy helyre gyűjtése történik.

Az adatok gazdagítását egy adott entitás-előfordulást reprezentáló saját URI, és ugyanazt az entitás-előfordulást jelölő, de egy másik adathalmazból származó URI közötti azonosság kifejezésével kell elvégezni: ezt az azonosságot ugyancsak egy triplet írja le, mely a folyamat befejezése után a saját adathalmaz részévé válik, kiegészítve az adott entitás-előfordulásra vonatkozó állítások sorát. Ehhez leggyakrabban az OWL webontológianyelv

`sameAs` nevű relációját alkalmazzák, amely alkalmas két URI ekvivalenciájának formális kifejezésére. Azonban vannak eltérő modellezési elképzelések is. Az `owl:sameAs` reláció értelmében minden állítás, mely A entitás-előfordulásról szól, érvényes B entitás-előfordulásra is, és megfordítva. Ez bizonyos esetekben az egyes előfordulásokról tett állításokban ellentmondáshoz, belső következetlenséghez vezet. Wouter Beek és szerzőtársai (2020) az általuk vizsgált tanulmányokra hivatkozva úgy vélik, a téves azonossági állítások aránya a LOD Cloud-gyűjteményben 2,8-4% is lehet. Ezért az OWL-reláció helyett több esetben gyengébb szemantikájú relációkat (`skos:exactMatch`, `skos:closeMatch`, `skos:relatedMatch`, `umbel:isLike`) alkalmaznak az ekvivalencia szintjeinek differenciáltabb jelzésére.

Az egyes gráfokban található külső hivatkozások száma fontos minőségi mutató: azt mutatja meg, hogy az adatgazda milyen mértékben követte a Tim Berners-Lee által megfogalmazott ajánlást, és mennyire szőtte bele saját adatait a szemantikus hálóba. Emellett lényeges az is, hogy ezek a hivatkozások összesen hány adathalmaz felé mutatnak: ez utóbbi mutatószámot nevezi Kim Tallerás (2017) *outdegree*-értéknek (magyarra körülbelül *kifelé irányuló fokszámként* lehetne fordítani). Ez az érték azt mutatja meg, hogy egy adathalmazban szereplő forrásleírások milyen gazdag hivatkozási hálót tartalmaznak más adathalmazok felé, mennyiben biztosítják az átjárást további tripletek felé, amelyek újabb és újabb szegmensekkel bővítik az adott entitás-előfordulásra vonatkozó ismereteket.

Bizonyos adatkonverziós eljárásokban az eredeti szintaxis valamilyen RDF-logikájú formátumra alakítása egyszerre zajlik az adatgazdagítással: az ilyen típusú folyamatokat *aktív konverzió*nak nevezik. A kiinduló adatszerkezetben elkülöníthető entitás-előfordulást (például egy MARC-rekord 100-as mezőjében található személynevet mint karakterláncot) a konverziós eljárásban összevetik a külső adathalmazokban tárolt URI-k címkéivel (ezekről bővebben lásd a 6.5.2. fejezetet), majd az egyezés alapján kiválasztott URI-t a konvertált adatszerkezetbe helyezik az előre definiált kapcsolatazonosító segítségével.²⁷ Minél több külső adathalmaz elérési útvonala helyezkedik el az adatgazdagítási konfigurációban, annál több URI-t lehet társítani a saját adathalmazban található entitás-előfordulásokhoz.

²⁷ Egy ilyen elven működő konverziós segédeszköz hozzáférhető a következő címen:
<https://github.com/lcnetdev/marc2bibframe2#active-record-conversion> [letöltés: 2020.12.19.]

6.7. A linked data egyéb segédeszközei

A ShEx nyelv segítségével speciális validációs szabályokat lehet megfogalmazni az RDF-állítások tartalmára vonatkozóan. A ShEx-séma alapján könnyedén megállapítható egy gráfról, hogy az megfelelő-e. A validációt a ShEx nyelv szintaxisa alapján megalkotott korlátozások (constraints) irányítják, amelyeknek két fajtája ismeretes:

- a **csomóponti korlátozás (node constraint)** megvalósulhat például egy tripletben található literál értékeinek szabályozásában, például, hogy egy személy életkora nem lehet kisebb 13 évnél, és nem lehet nagyobb 20 évnél. Rögzíthetők bizonyos előre definiált értékek, vagy vonatkozhat korlátozás a literál adattípusára is;
- a **triplet-korlátozás (triple constraint)** hatóköre egy bizonyos kijelölt csomópontra, a fókuszcsomópontra (focus node) vonatkozó valamennyi állításra kiterjed. Szabályozhatja a csomópontra vonatkozó, előre specifikált állítmánnyal készülő tripletek számosságát, illetve – hasonlóan a csomóponti korlátozásokhoz – tartalmazhat megkötéseket bizonyos állítmányok után álló tárgyak adattípusára vagy szerkezetére vonatkozóan. A kardinalitás az alábbi karakterek segítségével fejezhető ki (W3C, 2019):
 - a + jel jelentése: egy vagy több állítás kell vonatkozzon a csomópontra a megadott állítmány felhasználásával;
 - a * jel jelentése: legalább egy állítás vonatkozzon a csomópontra a megadott állítmánnyal;
 - a ? jel jelentése: legfeljebb egy állítás vonatkozzon a csomópontra a megadott állítmánnyal;
 - az {n} kifejezés jelentése: pontosan n számú állítás vonatkozzon a csomópontra a megadott állítmánnyal;
 - az {m, n} kifejezés jelentése: legalább m, de legfeljebb n állítás vonatkozzon a csomópontra a megadott állítmánnyal.

A szerkezeti korlátozással elérhető például, hogy egy tulajdonság értékészlete csak egy bizonyos szótárkódolási séma URI-jaiból állhasson (ilyenek lehetnek például az RDA kötött szótárai).

A SHACL a ShEx-hez hasonlóan validációs célokra alkalmazható: segítségével egy RDF-gráf összevethető a tripletekre vonatkozó, előzetesen definiált szabályokkal. A SHACL használatával bizonyos alaki vagy szerkezeti követelményeket lehet rögzíteni az

egy csomópontokhoz (*node shapes*) vagy tulajdonságokhoz (*property shapes*). Az ilyen megkötéseket tartalmazó adatszerkezetet ugyancsak gráfnak nevezik – megkülönböztetik azonban a szabályokat definiáló alakzati gráfot (*shape graph*) és azt a gráfot, amely a validálandó adathalmazt tartalmazza (*data graph*) (PCC, 2020).

Egy *node shape* minden esetben egy bizonyos osztály – az ún. célosztály (*target class*), más néven a fókuszcsomópont (*focus node*) – valamennyi előfordulására vonatkozik, azok mindegyikére vonatkozóan kötelező érvényűvé teszi a meghatározott szabályokat. Előírható például, hogy egy bizonyos osztály tagjaira csak bizonyos tulajdonságokat lehessen alkalmazni – ekkor az osztály *zárt* –, a tulajdonságokra pedig olyan korlátozásokat lehet előírni, amelyek az adott tulajdonság kardinalitására (*minCount*, *maxCount*), adattípusára (*dataType*), vagy szerkezeti előírásokra vonatkoznak (*pattern*) – ez utóbbiak megadhatók RegEx-kifejezések segítségével is (W3C, 2017).

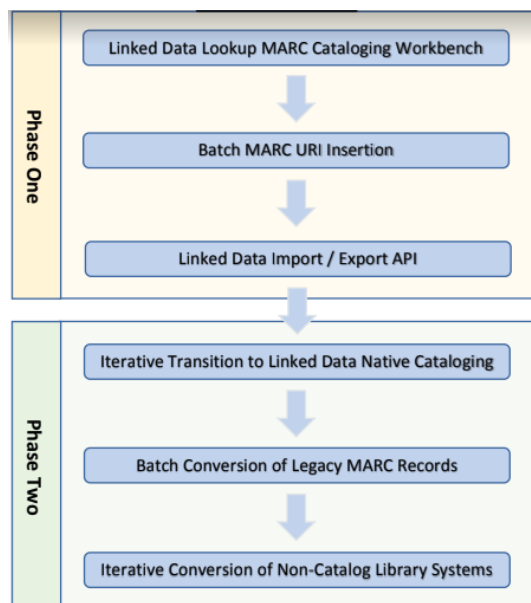
A FRESNEL szótár célja, hogy a benne foglalt osztályok és tulajdonságok segítségével más, a szemantikus weben közzétett tartalmak megjelenítését tegye szabályozhatóvá. E szótár sajátosságait Emmanuel Pietriga, illetve munkatársai ismertették a 2006-ban rendezett International Semantic Web Conference-en (Pietriga et al., 2006). Írásukból kitűnik, hogy a FRESNEL legfontosabb alkotóelemei a lencsék (*lenses*), illetve a formátumok (*formats*). Az előbbiek tartalmazzák azokat az RDF-relációkat, amelyeket a megjelenítésben feltétlenül ábrázolni kell (*fresnel:showProperties*), a formátumok pedig leírják a lencsére vonatkozó konkrét reprezentációs utasításokat. A svédországi közös könyvtári katalógus (LIBRIS XL) megjelenítő felülete a FRESNEL ismert implementációja: itt összesen három lencsét definiáltak, amelyekkel a bibliográfiai források, illetve más entitások, például személyek válogatott metaadatait jelenítik meg (Lindström, é.n.).

7. A linked data és a könyvtár

A linked data technológiája, illetve az előző fejezetben ismertetett informatikai segédeszközök jó része még ismeretlen a hazai könyvtáros szakemberek jelentős hányada előtt, s hasonló a helyzet a bibliográfiai leírás entitásközpontú megközelítésének kérdésével is. Disszertációm negyedik és ötödik fejezetében bemutattam az új katalogizálási szemlélet megjelenésének történetét, áttekintettem a bibliográfiai univerzum entitásainak leírását adó fogalmi modelleket, illetve megvizsgáltam az ezek gyakorlati alkalmazhatóságát biztosító RDA katalogizálási szabályzatot. A következőkben felvázolom az intézményekben megvalósítható informatikai fejlesztés előkészítéséhez, illetve megvalósításához szükséges elméleti, illetve gyakorlati ismereteket. Áttekintem a közgyűjtemények linked data-technológiára történő átállásának lépcsőfokait, amelyek a MARC-rekordok azonosítókkal való gazdagításától a teljesen RDF-alapokra helyezett informatikai szolgáltatásspektrumig vezetnek. Szót ejtek az adatok RDF-konverzióját megelőző alkalmazásprofil-készítési, formátum-megfeleltetési munkálatokról, az adatok transzformációjáról, publikációjáról, s ismertetek néhány olyan, komplex publikációs szolgáltatást, amelyek segítségével már számos intézmény hajtott végre sikeres konverziós, illetve adatközzételési projektet. Bemutatom továbbá a linked data-technológia információkeresésre gyakorolt hatását, és ismertetem, hogy milyen tulajdonságokkal, új funkciókkal bír egy RDF-alapú adatokra épülő, korszerű közgyűjteményi keresőfelület. A fejezetet a linked data-technológia alkalmazásával kapcsolatos lehetséges jövőképek felvázolása zárja.

7.1. A linked data-technológia bevezetésének lépései

E fejezetrészben az UC Davis University California, valamint a Zepheira informatikai cég munkatársai által koordinált, két éven keresztül futó BIBFLOW projekt eredményeit alapul véve mutatom be a linked data-alapú szolgáltatás felé vezető utat. A projekt tanulságairól összefoglaló tanulmányt készítettek az intézmény munkatársai (Smith et al., 2017). A dokumentumnak kimondottan nagy előnye, hogy érthetőbbé és átláthatóbbá teheti a döntéshozói pozícióban lévők szemében az intézmények előtt álló, kihívásokkal, ugyanakkor ígéretekkel is teli lépést: a linked data bevezetését az intézmény szolgáltatásai közé. A BIBFLOW zárójelentés kétszer három fázisra bontja a könyvtári munka-, illetve szolgáltatási környezet transzformálását, az URI-k MARC-ba illesztésétől egészen a teljes gyűjteményi ökoszisztéma új alapokra helyezéséig.



11. ábra – A BIBFLOW projekt szakértői által javasolt áttérési menetrend
(Smith et al., 2017)

7.1.1. Adatgazdagítás MARC-környezetben

Az első lépés, amely a linked data alkalmazása felé vezethet egy gyűjteményt, és amely semmiféle új katalogizálási szemléletet, sem pedig a MARC-tól eltérő adatformátum használatát nem igényli, az URI-k beillesztése a rekordok meghatározott pontjaira. Az amerikai közös katalogizálási együttműködés keretein belül *URIs in MARC* munkacsoport működik. Tevékenységének köszönhetően számos új adatmezővel, illetve almezővel bővült a MARC 21 adatsere-formátum elemkészlete, egyéb feladatai közül pedig az alábbiakat érdemes kiemelni (PCC, é.n.):

- magasabb szintű, nagy horderejű implementálások előtt meghozandó döntések előkészítése, amelyek egyebek mellett lehetnek az alábbiak:
 - alfanumerikus azonosítókat vagy URI-kat használjanak-e a rekordokban;
 - hogyan valósítsák meg egyazon entitás-előfordulás több azonosítóval való ellátását;
 - a bibliográfiai rekord mely pontjára helyezték el a mű-, illetve kifejezésforma-szintű entitásazonosítókat;
- irányelvek kidolgozása az azonosítók alkalmazására a bibliográfiai és authority rekordokban;

- mindazon entitások meghatározása, amelyek előfordulásaihoz azonosítót kell társítani [bár ajánlott, de ezeknek nem kell feltétlenül megegyezniük az LRM-ben, illetve az RDA-ban meghatározott entitásokkal, továbbiak is felvehetők – H.M.];
- az URI-k forrásait jelentő elemkészletek és szótárkódolási sémák kiválasztása.

Az URI-k *szabványos* helyre illesztésének lehetősége azonban csak a MARC 21-ben áll fenn, ezért azok a könyvtárak, amelyek HUNMARC-szabvány szerint végzik a feldolgozást, választásra kényszerülnek: vagy vállalják a MARC 21-re való teljes áttérés kihívásait – az ezzel járó adatjavítási és oktatási feladatokat –, vagy saját maguk (gyűjteménykezelő rendszerük fejlesztőivel együttműködve) helyezik el az azonosítókat egy általuk megállapított, nem szabványos helyre. Az itt nyert munkamegtakarítás azonban visszaüt akkor, amikor MARC 21-re írt RDF-konverziós segédeszközöket kell az egyedi viszonyokhoz alkalmazni az adatvesztés elkerülése érdekében. A szabványos megoldásoktól eltérő almezők alkalmazása az értekezés 10. fejezetében bemutatandó kísérleti konverziós projektben is okozott kisebb adatvesztési problémákat. Ugyanezen problémák nehezítették meg a 11. fejezetben tárgyalt konverziós feladat végrehajtását is.

Az adatgazdagítás történhet kézi munkavégzés segítségével, ekkor a forrásleírást végző könyvtáros a rekordokban fellelhető entitás-előfordulásokhoz egyenként keresi meg azokat a külső névterekből származó URI-kat, amelyek használatát például az RDA, az intézmény saját feldolgozási szabályzata vagy más normatív dokumentuma előírja. A retrospektív gazdagítás esetében a kezelendő rekordok nagy száma, vagy erre a feladatra dedikálható munkaerő hiányában ez az út csak igen nagy nehézségekkel járható. Érdemes inkább automatizálni a folyamatot, amihez az egyes névterek által szolgáltatott alkalmazásprogramozási interfészeket (API-kat) lehet igénybe venni. A VIAF például képes személyekről szóló információk átadására az ún. AutoSuggest API-n keresztül, testületi nevek esetében pedig az SRU Search API-t kell használni (Jin & Kudeki, 2019). A BIBFLOW projekt részeként a szakemberek a KUALI-OLE integrált gyűjteménykezelő rendszer katalogizálási modulját módosítva képessé tették a szoftvert arra, hogy automatikusan tudjon URI-kat átemelni a Library of Congress-ből, a VIAF-ból, illetve a Getty egyes tezauruszaiból (Smith et al., 2017). A kutatási zárójelentés arról számolt be, hogy az Ex Libris fejlesztőcég Alma nevű integrált könyvtári platformját is bővíteni tervezik olyan szolgáltatással, amely már a forrásleírás közben lehetővé teszi URI-k társítását a megtalált entitás-előfordulásokhoz.

7.1.2. Alkalmazási profil – retrospektív konverziós segédeszköz

Ha az intézmény úgy dönt, hogy adatszolgáltatási platformját a katalógusban tárolt adatvagyon szemantikus weben történő publikációjával is kiegészíti, akkor megfelelő linked data-alkalmazási profilt kell alkotnia. Ez nem azonos a korábban már tárgyalt RDA-alkalmazási profillal, amely, mint az ötödik fejezetben ismertettem, arra való, hogy segítségével kiválogassák azokat a metaadatelemeket, amelyeket az intézmény ténylegesen alkalmazni fog egy-egy bibliográfiai forrástípus leírásakor. A linked data-alkalmazási profil a specifikus metaadatigényeket kielégítő adatszerkezetek felépítését határozza meg, miközben szemantikai átjárhatóságot biztosít a nemzetközi szinten definiált szótárak és adatmodellek más alkalmazásaival (Coyle & Baker, 2009). Az alkalmazási profilt alapvetően az intézmény metaadatigényei alapján érdemes kialakítani, de tekintetbe kell venni az aktuális és a potenciális jövőbeni partnerek elvárásait is. Ennek megfelelően pedig több szerializációs szintaxisban megfogalmazott, vagy más-más szótárak elemeivel felépülő gráfokat is érdemes szolgáltatni. Az Alma integrált könyvtári platform esetében a konverziós lehetőségek például az alábbiak (*Linked Data*, é.n.):

- RDF/XML formátum a később ismertetendő BIBFRAME szótár relációival;
- RDF/XML formátum az RDA szótár relációival;
- JSON-LD formátum a DCTerms szótár relációival, szükség esetén más elemkészletekkel kiegészítve.

Az alkalmazási profil elkészítése úgy történik, hogy a hagyományos adatformátumban (leggyakrabban MARC-ban) tárolt adatelemek mindegyikéhez hozzá kell rendelni egy, tetszőleges szótárból választott relációjelölő URI-t, így lehetővé téve a hívójelek és almezőjelek által definiált egyes relációtípusok pontos tükrözését a gráfszerkezetben. A könyvtártudomány területén, mint az értekezés 6.5.2. fejezetében rámutattam, számos szótár hozzáférhető, így *elméletben* bármelyik alkalmas lehet arra, hogy segítségével ki lehessen fejezni a MARC-ban meglévő metaadatelemeket. A *gyakorlatban* azonban több dologra érdemes tekintettel lenni:

- megfelelően granuláris-e az adott szótár, azaz minden szükséges adatelem kifejezhető-e vele, amelyeket a MARC-rekord megkülönböztet, vagy összevonásokat kell alkalmazni – például van-e lehetőség különféle típusú megjegyzések rögzítésére?

- mennyire közismert, elterjedt az adott szótár? Ha olyan szótár, vagy szótárak olyan kombinációja kerül használatba az adatok konverziója során, amely(ek)nek alkalmazása igen ritka a gyakorlatban, az RDF általános használhatósága ellenére is újabb konverziós fázist kell közbeiktatni, ha az adatvagyonot valamilyen aggregációs szolgáltatásban szeretnék felhasználni. A könyvtári közösségnek meg kell állapotodnia lehetőség szerint egyetlen szótár használatában, így a ma meglévő, a nemzeti MARC-verziók okozta nehézségek is kiküszöbölhetővé válnak a jövőben.
- ha több entitást definiáló szótárak kombinációja kerül használatba, különösen nagy figyelmet kell fordítani arra, hogy az egyes osztálybesorolások nyomán ne keletkezzenek hamis állítások, ne jöjjenek létre téves kapcsolatok.

Az Országos Széchényi Könyvtár által 2010-ben végrehajtott adatközzététel például egy olyan alkalmazásprofil alapján valósult meg, amely a bibliográfiai, illetve authority rekordok adatait az akkor legismertebb szótáraknak, a FOAF-nak, a DC-nek, illetve az SKOS-nak feleltette meg (OSZK, é.n.).

7.1.3. Az adatok konverziója

Az alkalmazási profil elkészülte után kezdődhet meg az adatkonverziós folyamat, amelynek során a kiinduló adatformátumból valamilyen, RDF-logikát támogató adatszerkezet jön létre. A transzformációs munkát annak aprólékossága, pontosságigénye, de a kezelendő adatok többnyire igen nagy mennyisége miatt is számítógéppel, valamilyen XML-feldolgozóval végzik el, mivel a napjainkban ismert konverziós eljárások az igen sokoldalúan alkalmazható stíluslapokon alapulnak. Ezért a legjobb, ha a kiinduló adatállomány is XML-ben – logikusan MARC/XML-ben – áll rendelkezésre, amelyet a legtöbb integrált gyűjteménykezelő rendszer képes exportformátumként előállítani. Ugyanakkor bármely más, XML-környezetben is alkalmazható metaadatséma felhasználásával készült adatállomány is megfelel a célnak, ha ahhoz előzőleg linked data-alkalmazási profilt készítenek.

A stíluslap elkészítését általában nem könyvtáros szakemberek, hanem informatikusok végzik az intézmény által biztosított linked data-alkalmazási profil, megfeleltetés (*mapping*) alapján (ez történt a 10. fejezetben bemutatandó OSZK-projekt esetében is, ahol a profilt e sorok írója biztosította). Ez a stíluslap részletes utasításokat tartalmaz arra vonatkozóan, hogy a kiinduló XML-dokumentumpéldány címkéit melyik szótárrelációra kell átalakítani, illetve milyen osztálybesorolásokat kell alkalmazni az

entitás-előfordulások, illetve más metaadatértékek esetében. A következő példában látható XML stíluslap-részlet az 504-es MARC-mező adattartalmának átalakítását szabályozó utasításokat tartalmaz:

```
<xsl:template match="marc:datafield[@tag='504' or @tag='880']" mode="work504">
  <xsl:param name="serialization" select="'rdFXML'"/>
  <xsl:variable name="vXmlLang"><xsl:apply-templates select="." mode="xmlLang"/>
</xsl:variable>
  <xsl:choose>
    <xsl:when test="$serialization = 'rdFXML'">
      <bf:supplementaryContent>
        <bf:SupplementaryContent>
          <xsl:for-each select="marc:subfield[@code='a']">
            <rdfs:label>
              <xsl:if test="$vXmlLang != ''">
                <xsl:attribute name="xml:lang"><xsl:value-of select="$vXmlLang"/>
              </xsl:if>
              <xsl:value-of select="."/>
            </rdfs:label>
          </xsl:for-each>
          <xsl:for-each select="marc:subfield[@code='b']">
            <bf:count><xsl:value-of select="."/></bf:count>
          </xsl:for-each>
        </bf:SupplementaryContent>
      </bf:supplementaryContent>
    </xsl:when>
  </xsl:choose>
</xsl:template>
```

A sablon (*template*) kijelölő része tartalmazza, hogy a kiinduló dokumentumpéldány mely csomópontjait kell feldolgozni, jelen esetben az 504-es, illetve az adatelem párhuzamos nyelvi megfelelőit tartalmazó 880-as mezőt. A sablon feldolgozó része gondoskodik arról, hogy az adatmezők két almezőjének (\$a és \$b) értékei a bf:supplementaryContent metaadatelem értékei legyenek a kimeneti adatszerkezetben, illetve, hogy ezek az értékek a bf:SupplementaryContent osztály tagjaként definiálódjanak (Tóvári & Szabó, 2011). A második példa²⁸ egy bonyolultabb konverziós folyamatot ismertet:

```
<xsl:template match="marc:leader" mode="work">
  <xsl:param name="serialization" select="'rdFXML'"/>
  <xsl:choose>
    <xsl:when test="$serialization = 'rdFXML'">
      <xsl:variable name="workType">
        <xsl:choose>
          <xsl:when test="substring(.,7,1) = 'a'">Text</xsl:when>
          <xsl:when test="substring(.,7,1) = 'c'">NotatedMusic</xsl:when>
          <xsl:when test="substring(.,7,1) = 'd'">NotatedMusic</xsl:when>
          <xsl:when test="substring(.,7,1) = 'e'">Cartography</xsl:when>
          <xsl:when test="substring(.,7,1) = 'f'">Cartography</xsl:when>
          <xsl:when test="substring(.,7,1) = 'g'">MovingImage</xsl:when>
          <xsl:when test="substring(.,7,1) = 'i'">Audio</xsl:when>
          <xsl:when test="substring(.,7,1) = 'j'">Audio</xsl:when>
          <xsl:when test="substring(.,7,1) = 'k'">StillImage</xsl:when>
          <xsl:when test="substring(.,7,1) = 'o'">MixedMaterial</xsl:when>
          <xsl:when test="substring(.,7,1) = 'p'">MixedMaterial</xsl:when>
          <xsl:when test="substring(.,7,1) = 'r'">Object</xsl:when>
          <xsl:when test="substring(.,7,1) = 't'">Text</xsl:when>
        </xsl:choose>
      </xsl:variable>
    </xsl:when>
  </xsl:choose>
</xsl:template>
```

²⁸ A stíluslap-részletek forrásai azok a teljes állományok, amelyek a MARCEdit szoftver telepítésekor kerülnek a felhasználó számítógépére.

```

</xsl:choose>
</xsl:variable>
<xsl:if test="$workType != ''">
  <rdf:type>
    <xsl:attribute name="rdf:resource">
      <xsl:value-of select="concat($bf,$workType)"/>
    </xsl:attribute>
  </rdf:type>
</xsl:if>
</xsl:when>
</xsl:choose>
</xsl:template>

```

A rekordfej hatodik pozícióján lévő karakter értékétől függ, hogy a `bf:workType` metaadatelem milyen értéket fog kapni az átalakítás után: a stíluslap valamennyi lehetséges esetet tartalmazza. Az itt szereplő kódrészlet a hetedik karakter értékét vizsgálja, ez azért van, mert a MARC rendelkezései alapján a rekordfej karakterpozícióit a nulladikkal kezdjük számolni – az XML-feldolgozás szempontjából azonban a hetedik karakterhelyről van szó.

Ugyancsak a konverziós folyamat eredményeként jönnek létre az entitás-előfordulásokat (személyeket, testületeket, tárgyszavakat, vagy akár műveket, kifejezési formákat stb.) reprezentáló URI-k is a kimeneti adatszerkezetben, ennek alapját, az ún. bázis-URI-t a konverzió elindítása előtt meg kell adni. Aktív konverzió esetén (lásd a 6.6., illetve a 9.6.2. fejezeteket) a transzformációval egyidőben adatgazdagítási folyamat is lefut, a saját kialakítású URI-khoz más, külső adathalmazokból származó, ugyanazt az entitás-előfordulást reprezentáló azonosítók csatlakoznak.

A transzformáció végén előálló adatállományt publikálás előtt alapos minőségi vizsgálatnak kell alávetni. Az alkalmazási profilban foglaltak alapján meg kell győződni arról, hogy a kiinduló állományban található metaadatelemek megfelelően átalakultak-e a kijelölt szótárrelációkká, fennáll-e bármiféle szemantikai ütközés a relációjelölő, illetve az adattartalom között, illetve ellenőrizni kell, minden adatelem megjelenik-e a kimeneti adatszerkezetben, amelyet a bemeneti állomány is tartalmaz, vagy adatvesztés lépett fel. Bármilyen probléma esetén a konverziót a stíluslap megfelelő módosítása után újra le kell futtatni, illetve felmerülhet a kiinduló adatállomány módosításának szükségessége is.

A Monguz Kft. egy ReCatalogue nevű, böngészőből elérhető konverziós platformot alakított ki, melynek tényleges implementációi a disszertáció írásakor még nem ismertek, a felület működését azonban már több előadásban is ismertették (például Fülöp, 2019). A ReCatalogue felületére Z39.50 protokollon keresztül lehet beemelni MARC-rekordokat, tetszőleges adatforrásból. Az importált adatszerkezet elsőként alapos ellenőrzésen megy keresztül, mégpedig előre definiált, ún. prevalidációs szabályok alapján. E szabályok fő

funkciója, hogy megvizsgálják, hogy a MARC-rekord tartalmazza-e mindazon adatelemeket, amelyek szükségesek ahhoz, hogy az adatmodell által meghatározott entitásokat izolálni és azonosítani lehessen. A rendszer figyeli, hogy rendelkezésre állnak-e azok a kapcsolódási pontok, amelyek révén az újonnan létrejövő állítások kapcsolódhatnak a már meglévő gráfhoz, avagy szemantikai hálózhoz (például tartalmaz-e a rekord 240-es adatmezőt, amelyből a rekordban leírt műnek a címe megállapítható stb.). Amennyiben valamilyen szükséges információ hiányzik az entitások elkülönítéséhez, a modul erről üzenetet küld a felhasználónak és felszólítja az adatok pótlására. Ha azonban minden rendelkezésre áll, végbemegy a konverziós folyamat – ezt több mint száz, igen aprólékos, ugyancsak előre definiált szabály vezérli –, és az így keletkezett részgráf az entitás-előfordulásokat azonosító URI-k révén kapcsolódik a már korábban eltárolt szemantikai hálózhoz (a platform mögött álló ún. *data pool*hoz), gazdagítva ezzel az egyes entitás-előfordulásokra vonatkozó tudást. Későbbi fejlesztések eredményeként a rendszer képes lesz más szerkezetekbe (például BIBFRAME-re is) konvertálni az eredendően a Monguz Kft. saját adatmodelljében definiált entitások szerint tárolt metaadatközléseket, amelyek így könnyen és problémamentesen felhasználhatók a más szolgáltatások, más közgyűjtemények felé irányuló adatkommunikációban, partneroldali fogadásuk és kezelésük az univerzális formátumnak köszönhetően nem jelenthet problémát.

7.1.4. Az adatok publikációja

A közgyűjteményi adatok publikálása jól ismert, bár hazánkban aránylag ritkán alkalmazott gyakorlat. Amint az értekezés első oldalain utaltam rá, az adatpublikáció az informatikai szolgáltatások fejlesztőinek készül, s aki közzétételt végez, az a Z39.50 szervere elérhetőségeit, esetleg az OAI-csatornája címét és paramétereit hozza nyilvánosságra, ezeken keresztül az informatikus partnerek a megfelelő szaktudás birtokában hozzáférnek az intézményi adatbázishoz, s onnan többnyire XML-formátumú adatokat képesek átemelni.

A szemantikus publikáció ezzel szemben az RDF-logikát követő formátumok használatával történik. A linked data-technológia hajnalán a legtöbb adatgazda az RDF/XML-formátumot választotta, a gráfokat tartalmazó állományokat egyszerűen becsomagolt állományként letölthetővé téve (*bulk download*). Ez a lehetőség a mai napig

fennáll például a Magyar Elektronikus Könyvtár²⁹, továbbá a Digitális Képtár³⁰ folyamatosan frissülő metaadathalmazai esetében.

Igen gyakran járt út, hogy a gráfot valamilyen triplestore-ban tárolják: ezek olyan adatbázis-kezelő rendszerek, melyek a relációs adatbázisokkal ellentétben nem táblákban, hanem tripletek formájában tárolják az adatokat.³¹ A triplestore-ok a gráfadatbázisok részhalmazát alkotják (*RDF Stores*, é.n.), a lekérdezőnyelvük pedig általában a SPARQL. A nyelv használatával megfogalmazott lekérdezések az adatbázis-kezelő által biztosított felületen, az ún. SPARQL-végponton adhatók át a rendszernek, és itt jelenik meg az eredményhalmaz is. Ismert triplestore-ok a MarkLogic, az Apache Jena TDB, a Virtuoso, a Stardog, az Allegrograph és a Blazegraph (*DB-Engines Ranking of RDF Stores*, é.n.). Hazánk nagyobb közgyűjteményei, mint az Országos Széchényi Könyvtár³², vagy a Magyar Nemzeti Múzeum könyvtára³³, a Virtuoso szoftver segítségével tették elérhetővé adataikat.

A szemantikus webtechnológia könyvtári alkalmazásának kezdetén a közgyűjteményi dolgozók hűvös távolságtartással fogadták az RDF-alapú adatformátumokat, és az ezek segítségével történő adatközzételt, mert úgy vélték, a publikált adattömegnek semmiféle gyakorlati haszna nem lesz a felhasználók számára, értelmezése ellenben igen nehézkes, és az átlagos gyűjteményhasználótól nem elvárható. Rob Sanderson azonban felhívja a figyelmet a már korábban is tárgyalt tényre: a szemantikus adatpublikációval nem *közvetlenül* a használókat célozzuk (Sanderson, 2018). Amikor egy közgyűjtemény adatpublikációt végez, erre érdemes tekintettel lenni, és az adatsomagot valamilyen API-n keresztül hozzáférhetővé tenni azok számára, akik azt valamilyen szolgáltatásban szeretnék felhasználni (a korábban említett Alma platform is biztosítja ezt a lehetőséget (*Linked Data*, é.n.)). Megfelelő informatikai szaktudás birtokában elő lehet állítani olyan keresőfelületeket, adatvizualizációs megoldásokat és más szoftvereszközöket, melyek a mögöttük rejlő RDF-logika valamennyi előnyét ki tudják használni, és ezeket az előnyöket (az entitásalapú keresés lehetőségét, a kapcsolattípusok ábrázolhatóságát stb.) a felhasználók számára is megtapasztalhatóvá tudják tenni. A teljes folyamatot Sanderson egy

²⁹ Hozzáférhető a <https://mek.oszk.hu/html/lod.html> címen [letöltés: 2020.12.19.]

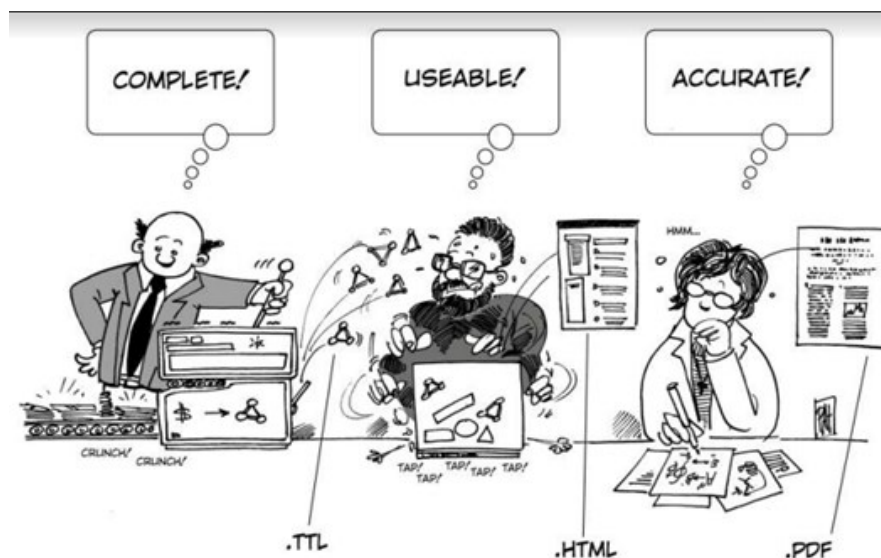
³⁰ Hozzáférhető a <https://dka.oszk.hu/html/export.html> címen [letöltés: 2020.12.19.]

³¹ Lehetséges azonban relációs adatbázisok tartalmát is gráfokká alakítani – például a Virtuosoval.

³² Hozzáférhető a <http://setaria.oszk.hu/sparql/> címen [letöltés: 2020.12.19.]

³³ További információk: <https://mnm.hu/hu/kozponti-konyvtar/nyilt-bibliografiai-adatok> [letöltés: 2021.06.02.]

karikatúra segítségével mutatta be 2018-ban, az EuropeanaTech szakmai konferencián tartott előadásában.



12. ábra – A tripletek nem közvetlenül a felhasználónak készülnek – szükség van az informatikus szakemberre is (Sanderson, 2018)

Az ábra első harmadában látható metaadatszakerítő, adattudós a közgyűjteményi adatvagyonból – amelyet vélhetően MARC-ban tárolnak, ezt jelképezi a dollárjel – RDF-gráfokat állít elő a fejlesztők számára, akik jól ismerik e gráfok felhasználásának módjait, és olyan felhasználói felületet készítenek, amelyekkel a végfelhasználók közvetlen interakcióba kerülnek. Nem is biztos, hogy a használó tudomást szerez arról, hogy a felület, amellyel dolgozik, újszerű adattárolási logikákra alapul.

7.2. Komplex publikációs szolgáltatások

A könyvtári szoftveralkalmazások fejlesztői (az angol terminológia szerint a *vendorok*) több esetben olyan térítéses szolgáltatásokat dolgoztak ki, amelyekkel az egyes intézmények számára jelentősen megkönnyítették a szemantikus weben történő megjelenést, mivel az egész folyamatot automatizálták az adatok konverziójától kezdve egészen a megjelenítésig. Négy ilyen szolgáltatásról érdemes megemlékezni.

- A *Library.Link Network* a Zepheira informatikai cég 2016-ban indított, globális infrastruktúra-szolgáltatása olyan könyvtárak és más kulturális örökségi intézmények számára, akik szeretnék növelni gyűjteményük láthatóságát a világhálón. A szolgáltatás üzemeltetése több partner együttműködésével zajlik, akik közül meg kell említeni az Atlas Systems-t, az Innovative-ot, illetve a SirsiDynix-ot,

ezek a cégek mind közismert szereplői a könyvtári szoftverek piacának. A Library.Link Network térítéses és ingyenes szolgáltatásokat is kínál partnereinek. 2017-ben már körülbelül 1100 könyvtár csatlakozott a hálózathoz, s összesen mintegy 29,5 millió MARC-rekord információtartalmát konvertálták át 118 millió konkrét entitás-előfordulás leírásává. Az intézmények által gondozott, ún. helyi gráfok (local graph) a rendszeren keresztül egymással összeköttetésbe kerülnek, így létrehozva egy magasabb szintű, összesített, intézményközi gráfot. A helyi gráfok az intézmények egységes szerkezetű Library.Link-weboldalain kereshetők.³⁴

- A *BLUECloud Visibility* szolgáltatást a SirsiDynix indította 2015 őszén. Segítségével a cég termékeit (Horizon, Symphony) használó könyvtárak a BLUECloud webalkalmazáson keresztül el tudják juttatni adataikat a Library.Link hálózatba.
- Az *ALIADA* (magyarul: szövetség) nevű projektről, illetve az ennek keretein belül fejlesztett, ugyancsak *ALIADA* nevű szoftverről bővebben is érdemes megemlíteni, mivel magyarországi vonatkozásai is vannak (Horváth 2015a, 2015b). A 2010-es évek első felében (2013-2015) futó európai uniós fejlesztési projektnek egy olyan keretrendszer megalkotása volt a célja, amelynek segítségével a közgyűjtemények – könyvtárak és múzeumok egyaránt – könnyedén, minimális humán beavatkozással publikálni tudják adataikat a szemantikus weben. A projektgazdák között megtalálható a spanyolországi Artium múzeum, a spanyolországi Tecnalía, illetve Scanbit cégek, valamint az @Cult nevű olasz informatikai fejlesztőcéggel, amely például az OliSuite nevű integrált könyvtári platformot is készítette. A projekt magyarországi résztvevője a Szépművészeti Múzeum volt. Az *ALIADA* szoftvert telepítés után lehet használatba venni.³⁵ Az alkalmazás Java nyelven íródott, az Apache Struts, Apache Camel, Apache Velocity, Silk, és Freelib-MARC4j nyílt forráskódú szoftverek felhasználásával. Az adatok tárolását a Virtuoso triplestore végzi, amely webes megjelenítő felülettel, továbbá SPARQL-endpointtal is rendelkezik, amelyen keresztül az adatok lekérdezhetők. Az adatkonverzió során képződő RDF-kijelentések állítmányai jórészt az FRBR_{oo} szótárából származnak – ezzel lehetővé téve a könyvtári és múzeumi adatok közzétételét is – az esetleges hiányzó relációkat más szótárakból pótolják, mint például a CRM, a SKOS és SKOSXL, a WGS84, a FOAF, DCTERMS, vagy az

³⁴ A Western Michigan Library adathalmaza például hozzáférhető a következő linken:

<http://wmich.library.link/> [letöltés: 2020.12.19.]

³⁵ A szoftvereszköz hozzáférhető: <https://github.com/ALIADA/aliada-tool/> [letöltés: 2020.12.19.]

OWL-TIME (Horváth, 2015b). A szoftver bemeneti formátumként MARC-ot, LIDO-t (a múzeumi közösség XML-sémája), illetve Dublin Core-t képes fogadni. A konverzióval egyidőben az alkalmazás adatgazdagítást is végez, így beágyazva a gyűjtemények adatait a nemzetközi szemantikai hálóba (Horváth 2015a). Az adatpublikációt időről-időre újra el kell végezni, követve a gyűjtemény folyamatos gyarapodását.

- A Share-VDE nevű szolgáltatást megvalósító többoldalú együttműködés főbb résztvevői a Casalini Libri olaszországi bibliográfiai ügynökség, továbbá az imént említett @Cult, akik a kulturális örökségi szektor egyes ágenseinek igényeit szolgálják ki integrált gyűjteménykezelő rendszerek, illetve szemantikusweb-technológiára alapuló szoftvermegoldások forgalmazásával. Rajtuk kívül még 16 észak-amerikai kutatókönyvtár vállalt részt a projektcélok megvalósításában, amelyek közül az alábbi három a legfontosabb:
 - hatékony munkakörnyezet kialakítása a linked data használatára könyvtári kontextusban, illetve azon kívül is;
 - az adatok gazdagítása további információforrásokkal, valamint kapcsolatokkal, amelyek a MARC-ban kifejezhetetlenek voltak; valamint a bibliográfiai, illetve authority adatok konverziója linked datává;
 - discovery felület kialakítása (összhangban az alkalmazott adatmodell hármas tagolódásával), megkönnyítendő az adatok emberi felhasználását.

A Share-VDE projekt legfőbb eredménye a partnerkönyvtárak gráffá (BIBFRAME-gráffá) konvertált adatállományára épülő keresőfelület³⁶, amely a személyek, kiadók, illetve tárgyszavak mellett művekre is képes keresni, továbbá ezekről információgazdag, külső hivatkozásokkal és más adatokkal (műveknél például nyelvi címváltozatokkal, személyeknél életrajzi adatokkal) bővített adatlapokat kínál a felhasználóknak. Az adatlapokról a használók továbbnavigálhatnak a konkrét gyűjteményi példányok felé, amelyeket címkés, MARC-, illetve gráf-formátumban is megtekinthetnek.

³⁶ Hozzáférhető a <https://share-vde.org>, egy újabb béta-verzió a <https://svde.org> oldalon [letöltés: 2022. május 19.]

7.3. A linked data lehetőségei az információkeresésben – adatgazdagítás, entifikáció és szerendipitás

A linked data használatának sokáig gyenge pontja volt az új technológia segítségével tárolt közgyűjteményi információk keresetősége. Sok intézményt eltántorított a váltástól az, hogy nem láttak olyan keresőfelületeket, amelyek funkciói meggyőzték volna őket az új technológia gyakorlati hasznáról, valamint az átállásba, és az új forrásleírási módszertan megtanulásába befektetett energia megtérüléséről. A szoftverfejlesztők azonban, ha lassan is, de elmozdultak az RDF felhasználásának irányába. Ez a lépés igen nagy jelentőséggel bír, hiszen az új technológia alkalmazása a közgyűjteményi keresőfelületek alapvető megújulását, egyszerűsödését hozhatja, továbbá, mint említettem, a gyűjteményhasználók leginkább az információkeresés területén tapasztalhatják meg a linked data alkalmazásának előnyeit.

Az Innovative fejlesztőcég 2019 áprilisában felmérést készített a felsőoktatási könyvtárak gyűjteményeit használó oktatók, kutatók körében, amely tulajdonképpen csak megerősítette a több éve nyilvánvaló tény: a közgyűjtemények által szolgáltatott adatokat a használók igen hitelesnek gondolják, azonban az információkereső felületek használatát túlságosan bonyolultnak vélik – ezért fordulnak a nagy, webes keresőszolgáltatásokhoz, jóllehet, az azok által szolgáltatott információkat már jóval kevésbé találják megbízhatónak (Innovative, 2019). Logikusnak és érthetőnek tűnik a következtetés, hogy a felhasználószám csökkenésének megállításához a könyvtári tudásanyagot minél nagyobb arányban reprezentálni kell például a Google találatai között. Ugyanakkor valós alternatíva, hogy a közgyűjtemény keresőfelülete egyszerűsödjön le, közelítsen a webes keresőkhöz, jóval egyszerűbben, informatívabban biztosítva a bibliográfiai források megtalálhatóságát.

A modern kori információkeresés másik kiemelt hívószava az egyszerűség mellett a *szerendipitás*. A kifejezést Horace Walpole író, történész és politikus alkotta egy, a XVIII. század közepén írt levelében, melyben egy perzsa népmesére utal, ahol a főhősök újra és újra váratlan, izgalmas felfedezéseket tesznek. A XIX. század közepéig azonban a szó nem terjedt el szélesebb körben, ekkortól azonban egyre többen és több alkalommal kezdték el használni a véletlenül tett, eredményes és örömteli felfedezésekkel kapcsolatban. Az International Encyclopedia of Information and Library Science című összefoglaló munka a szerendipitást „*az épp megtalált, ám nem keresett információ hasznossága, potenciálja felismerésére irányuló*” képességként definiálja. A Dictionary for Library and Information Science szerint a szerendipitás inkább az ilyen örömteli „rátalálás” – azaz a véletlenszerű

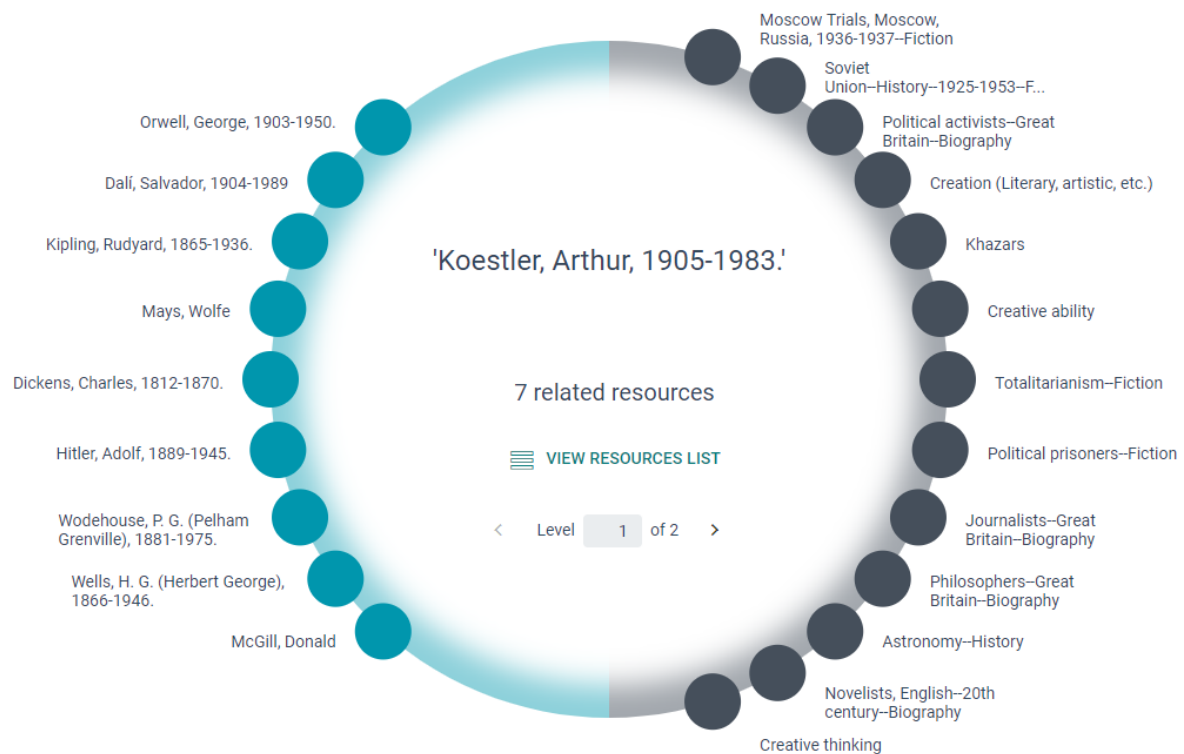
felfedezés – pillanataként értelmezhető, amelynek bekövetkezése azonban azon is igen nagy mértékben múlik, hogy az információt kereső ágens (ez esetben csak humán ágensről beszélhetünk) felismerje az éppen megtalált információ fontosságát, és relevanciáját a kutatott téma szempontjából (Carr, 2015). A linked data megjelenítését támogató információkereső felületeknek új feladatuk van, amelynek hatékony végrehajtásából aztán jelentős előnyt kovácsolhatnak: a felhasználók kérdéseire történő válaszadás mellett meg kell növelniük a szerendipitás-élmény bekövetkezésének valószínűségét, oly módon, hogy szabályozottan bár, de olyan kérdésekre is feleletet adnak, melyeket a használó *nem* tett fel.

7.3.1. Innovative Inspire

A szerendipitás élményét nyújtotta egy, ma már nem hozzáférhető, Inspire nevű discovery felület, amelyet az említett Innovative Solutions (egyebek mellett a Sierra, a Polaris és a Virtua nevű integrált könyvtári rendszerek fejlesztője) 2019 áprilisában mutatott be. Az első implementáció – a pennsylvaniai Cairn Egyetem könyvtárában – korlátozott ideig szabadon hozzáférhető, kipróbálható volt a világhálón.³⁷ Működésében nagy szerepet kapott a keresés eredményeül kapott információk vizualizációja, amely már nem MARC-rekordokra, hanem triplettekké konvertált adatállományra épült. A felületet megnyitva egyetlen keresőmező állt rendelkezésre, amelynek segítségével szerzőkre, címekre és fogalmakra, azaz tárgyszavakra lehetett keresni. A találati halmaz három részre tagolódott: az első rész (Resources) a keresett szerzőhöz kapcsolódó olyan források listáját tartalmazta, amelyek létrejöttében az illető közreműködött, vagy amelyek róla szóltak. A második rész (Related People) a keresett szerzőhöz kapcsolódó más ágenseket (személyeket és testületeket), míg a harmadik szakasz (Topics) az alkotóval összefüggésbe hozható LCSH-deszkriptorokat tartalmazott.

A találatok (kapcsolódó más szerzők, illetve deszkriptorok) a listás megjelenítés mellett egy könnyen áttekinthető vizuális eszköz, az ún. *context wheel* (kontextus-kerék) segítségével voltak megjeleníthetők, ahol az egyes csomópontokra kattintva további kapcsolatok tanulmányozására volt lehetőség. A keresési szempontok természetesen kombinálhatók voltak az ún. Workbench eszköz segítségével, amelynek végeredménye ugyancsak vizualizálható volt (Hubay, 2020a).

³⁷ A hozzáférés a <https://pbibu.na.innovativeinspire.com/> címen volt lehetséges.



13. ábra – A keresett személyhez kapcsolódó más személyek, illetve tárgyszavak megjelenítése az Innovative Inspire „context wheel”-jén
(képernyőkép, saját szerkesztés)

7.3.2. A Qulto szemantikus katalógus

Bizonyos felületek az RDF-kompatibilis formátumokban rejlő entifikációs lehetőséget használják ki, azaz a régóta ismert, és több intézmény által, több különféle technológiával sikeresen megoldott FRBR-esítési feladatot (ennek hagyományos technológiáit lásd az értekezés 5.3. fejezetében) valamilyen gráfszerkezetre alapozva valósítják meg. A Monguz Információtechnológiai Kft. által fejlesztett OPAC3 katalógusfelület³⁸ (illetve a mögötte rejlő adatbázis) megújításának alapja is egy új adatmodell, egy sajátos szótár volt, amelynek entitásait a szerzői jogok, illetve jogosultak figyelembevételével alakították ki. A Mű entitásszintjén megjelenő szerzői jogi jogosult jelen van minden alárendelt entitás esetében is, ugyanakkor feltétlenül szükséges a kifejezési formáknak megfelelő entitásszint elkülönítése is, hiszen itt új szerzői jogi jogosultak kerülhetnek a képbe (mivel például a fordítások nem külön művek, hanem egyazon mű kifejezési formái, így például a fordító személye a kifejezési formához és nem a műhöz

³⁸ A kísérleti felület megtekinthető a következő linken: <http://opac3.frbr.monguz.hu>

kapcsolódik). A Monguz ennek megfelelően alakította ki adatmodelljét, mégpedig a Debreceni Egyetem Egyetemi és Nemzeti Könyvtár (DEENK) katalógusából válogatott, magyar regényeket leíró bibliográfiai rekordokra alapozva. A szerzői jogi jogosultakat vizsgálva önálló műként határozták meg magukat a regényeket, az azokhoz mások által készített illusztrációkat, továbbá a megint mások által írt elő- és utószavakat. Kifejezési formaként határozták meg a műfordításokat, az egyszerűsített változatokat, illetve az irodalmi hangfelvételeket. Ezen az entitásszinten olyan szerzői jogi jogosultak jelennek meg, mint a fordító, az egyszerűsített változat írója, illetve a művet felolvasó személy. Ez a megközelítés csak részben vág egybe az LRM-ben foglaltakkal; mivel a Könyvtári Referenciamodell a műfordítást ugyancsak kifejezési formának tekinti, az egyszerűsített változatot azonban (például a gyermekek számára készítettet) már különálló műnek véli a befektetett szellemi-művészi munka mértéke okán (Riva, LeBoeuf & Žumer, 2017). Ha a művet – az értekezés 4. fejezetében írtak szerint – sajátos, újra meg újra felbukkanó mintázatként értelmezzük, akkor a Monguz megoldása logikusabb, intuitívabb információkeresést tesz lehetővé a Könyvtári Referenciamodellnél, hiszen az eredeti alkotás mintázata megőrződik az egyszerűsített változatban is, a szellemi-művészi munka – a befektetett energia mennyiségétől függetlenül – nem generál új mintát.

A találati lista az entitások absztrakciós sorrendjében közli az eredményeket: először a műveket, később a kifejezési formákat, majd a megjelenési formákat reprezentáló találatok érkeznek. Az egyes találatok megnyitásakor a felület megjeleníti az alárendelt entitásokat: művek, illetve kifejezési formák esetében a kiadások (megjelenési formák) metaadatait, illetve az azokhoz tartozó konkrét példányokra vonatkozó információkat (Fülöp, 2019). Szerzőre keresve ez a lista még kiegészül egy, magára a keresett személyre vonatkozó találattal, amely alatt minden, általa írt mű megjeleníthető.

Az egyes entitásszintekhez MARC-rekordok is generálódnak az alapvetően RDF-re, illetve lokálisan generált *qulto.eu* bázis-URI-val előálló tripletekre épülő szemantikus hálóból. A személy entitásszintű rekord a 024-es adatmezőben, hetes indikátorértékkel az adott személy előfordulásait tartalmazza közismert névterekben, a PIM névtérben, a Wikidatában, a Wikipédiában, a VIAF-ban, illetve a Nemzeti Névtérben; sőt, amennyiben a személy rendelkezik ISNI-azonosítóval, az is ebben az adatmezőben jelenik meg. A rekord tartalmazza továbbá a 100-as adatmező almezőit, illetve opcionálisan egy 979-es adatmezőt, amely a rekordhoz társított médiatartalom (például a szerző képe) adatait foglalja magába.

Sorrendben a következő a Mű entitásszintet reprezentáló MARC-rekord, amely jellemzően a 041-es (nyelv), 100-as (személynév kiemelt hozzáférési pontként), 245-ös (cím és felelősségi közlés), és 260-as (megjelenési adatok), továbbá 650-es (tárgyszó) mezőkből épül fel. A 100-as adatmező kiegészül egy \$4-es almezővel, amely a *writtenBy* közlést tartalmazza, amely a triplettekből emelkedik át. A 245-ös adatmezőhöz egy \$k almező adódik hozzá, amelynek értéke az entitásszintet nevezi meg: *literaryWork*. A 260-as mező értékei, azaz a megjelenés helye, illetve ideje, a Monguz Kft. definíciója szerint³⁹ a mű első kifejezési formájának első megjelenési formájára vonatkoznak. Érdekes még a 020-as adatmező (ISBN) nem releváns jelenléte a Mű entitásszinten, hiszen ez az adat mindig a megjelenési formák ismérve.

Művet reprezentáló MARC-rekord épül továbbá az illusztrációkból, illetve az elő- és utószókból mint önálló szellemi tartalomjegységekből. A jellemző adatmezők itt is a 020, 100, 245 és a 260. A 245-ös adatmező \$k almezője, mely ezen az entitásszinten is típusjelölő kifejezést tartalmaz, ezekben az esetekben a *visualWork*, illetve *literaryWork* értéket kapja, a művek címe pedig minden esetben azonos sémára épül: *Illusztrációk* vagy *Előszó XY: Z című művéhez*. A triplettekből származó, mű és mű közötti kapcsolatot a helyi szinten definiált 989-es adatmező jeleníti meg, a \$4 almező értéke ebben az esetben: *illustrationsFor* vagy *prefaceTo*. A 989\$w adatmező a kapcsolódó mű-rekord azonosítóját tartalmazza.

A kifejezésifforma-szintű rekordok az adatmodell szerint reprezentálhatnak műfordításokat, egyszerűsített változatokat, valamint irodalmi hangfelvételeket. A mű és kifejezési forma kapcsolattípusát rögzítő MARC-adatmező a 984-es, amelynek \$4 adatmezőjéből olvasható ki a konkrét relációtípus: *translationOf*, *retellingOf*, *recordingOf*. A kifejezésifforma-szintű rekordok 245\$k típusjelölő adatmezője a következő értékeket tartalmazhatja: *literaryTranslation*, *retoldWork (!)*, *literaryRecording*.

Némi értelmezési zavart okoz, hogy a felület Műtípus gyűjtőfogalom alatt két entitásszintet – művet, illetve kifejezési formát – reprezentáló rekordokat is kereshetővé tesz: az irodalmi műről, illetve a vizuális alkotásról művet leíró rekordok generálódnak, míg a másik három entitás esetében kifejezésifforma-szintű rekordok jelennek meg.

³⁹ Fülöp Endre szóbeli közlése alapján.

Műtípus

<input type="checkbox"/> irodalmi mű	<input type="checkbox"/> műfordítás
<input type="checkbox"/> egyszerűsített válto...	<input type="checkbox"/> irodalmi hangfelvétel
<input type="checkbox"/> vizuális alkotás	

14. ábra – „Műtípusok” keresése a Qulto Semantics felületén
(képernyőkép, saját szerkesztés)

A megjelenésforma-szintű rekordokat a Monguz Kft. entitásstruktúrájában *kiadásszintűnek* nevezik. Értelemszerűen ezek a rekordok emlékeztetnek a leginkább a hagyományos bibliográfiai rekordokra. A legfeltűnőbb különbség a 987-es, illetve 985-ös mező jelenléte, amely a megjelenési forma fölé rendelt kifejezési forma, illetve mű rekordjával köti össze az adott manifesztáció leírását. A 985-ös mező a kapcsolódó kifejezési formák rekordjaira, míg a 987-es a fölérendelt kifejezési formák és művek rekordjaira is rámutat.

985	\$4includes \$aBeszterce ostroma \$w4025
985	\$4includes \$aSzent Péter esernyője \$w4026
987	\$4includes \$e4025 \$aBeszterce ostroma \$w2193
987	\$4includes \$e4026 \$aSzent Péter esernyője \$w2211

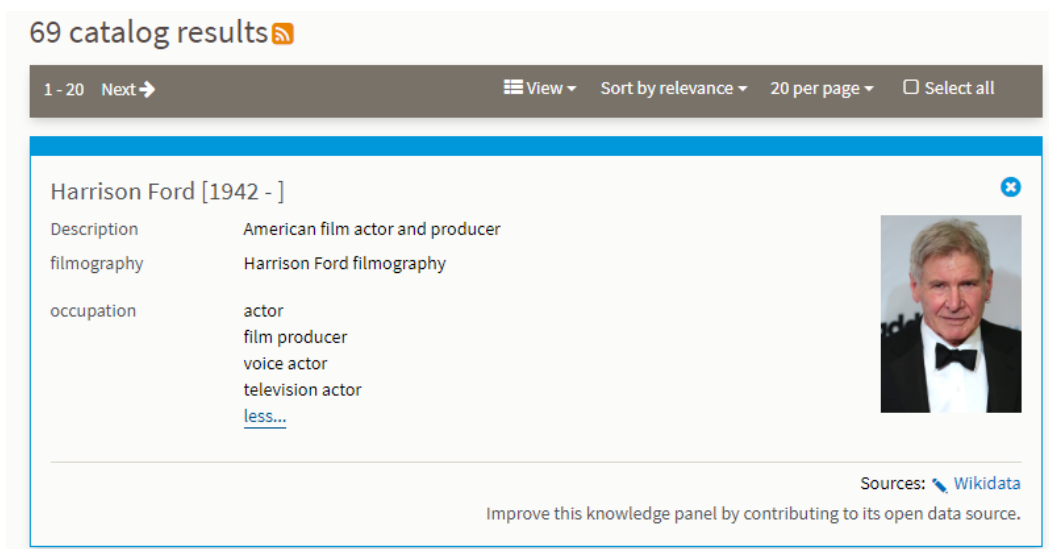
15. ábra – Entitáskapcsolatok MARC-alapú megjelenítése a Qulto Semantics-ban
(képernyőkép, saját szerkesztés)

7.3.3. Az adatgazdagítás előnyeinek kihasználása

A Stanford University-n futó, a Mellon Foundation által támogatott LD4P-projekt második fázisának keretein belül már hosszú ideje dolgoznak a Blacklight open-source szoftvermegoldásra épülő SearchWorks⁴⁰ közös egyetemi katalógusfelület megújításán. Az egyetem munkatársai által módosított és kibővített felület képes arra, hogy a Wikidata, illetve a Who's On First földrajzi névtér alkalmazásával a keresett személyekhez, illetve földrajzi nevekhez kapcsolódó, külső adatforrásokból átemelt információkat jelenítsen meg a találati lista fölött elhelyezkedő infoboxban. Ilyen dobozok a forrásleírásokhoz is csatlakozhatnak: a Blacklight módosított felülete – megint csak a Wikidatában tárolt adatok átemelése segítségével – az intézmény által készített leírásban nem szereplő információkkal

⁴⁰ Az eredeti Blacklight-felület a <https://searchworks.stanford.edu>, míg a módosított a <https://searchworks-ld.stanford.edu/> címen hozzáférhető [letöltés: 2022. május 19.]

képes bővíteni a találati képernyőt; hogy pontosan melyekkel, az tetszés szerint konfigurálható.



16. ábra – Tudáspanel megjelenítése a SearchWorks felületén elvégzett kereséshez kapcsolódóan. A külső adatforrás a Wikidata (Stanford University Libraries, 2019)

A külső adatforrásokból származó adatelemek szűrésekhez is felhasználhatók. A SearchWorks módosított felületén olyan adatelemek mentén szűkíthető a találati lista, melyeket nem tartalmaznak a helyben készült forrásleírások; kereshetők például olyan források, amelyek valamely közreműködője a Stanford University egykori diákja. Egy másik ígéretes fejlesztés a gyűjtemény böngészését támogató idővonal-funkció, amely a Wiki-univerzumból származó születési, illetve halálozási dátumokat felhasználva rendezi el a vizualizációban az alkotókat, illetve a gyűjteményben megtalálható műveiket.

Az egyetem munkatársai az alábbi célokat tervezik elérni a Blacklight kódjának módosításával, illetve továbbfejlesztésével (Stanford University Libraries, 2019):

- a keresési eredmények mellett a kontextuális információkat tartalmazó tudáspanel megjelenítése, amelyet az egymáshoz kapcsolódó adatok vezérelnek;
- az authority állományokat, továbbá a külső adatforrásokra mutató hivatkozásokat is felhasználó böngészési lehetőségek biztosítása a használók számára;
- sikertelen keresés esetén alternatív keresőkifejezések felajánlása, illetve gépelés közbeni keresési javaslatok tétele jelentésbeli kapcsolódás alapján, esetleg a földrajzi helyek keresésének finomítása az entitás-előfordulások URI-jainak biztosításával

(kiküszöbölve ezzel például az azonos alakú elnevezések okozta zajt a találati listákban);

- a keresőmotorok által végzett automatikus indexelés támogatása az oldalakon elhelyezett szemantikus jelölők (microdata) segítségével (Futornick, é.n.) – erről bővebben lásd a disszertáció 8. fejezetét.

7.4. A linked data szerepe a katalogizálás megújulásában

A fejezet zárásaként néhány lehetséges forgatókönyvet mutatok be a könyvtári forrásleírás jövőjével kapcsolatban. E jövőperspektívák között található optimista és pesszimista felfogások, innovatív, illetve konzervatív megközelítések is. A kérdés vizsgálatakor tekintettel kell lenni a metaadatelemezeket, illetve értékeiket hordozó formátum megváltozása, továbbá az új katalogizálási szemlélet bevezetése okozta kihívásokra is. A fokozatosság minél hatásosabb bemutatása, továbbá a teljesség érdekében azokat a lépcsőfokokat is ismertetem, melyek már közismertek.

- **Hagyományos szemléletű feldolgozás, hagyományos hordozóformátummal, hagyományos katalógusfelülettel.** A jelenlegi állapot. A munkatársaktól nem igényel semmilyen változást, az egyre gyorsabb tempóban elavuló technikai és technológiai környezethez való ragaszkodás azonban egyre kevesebb életteret hagy. Az intézmény a korszerűbb szolgáltatásokhoz, aggregációs platformokhoz nem fog tudni csatlakozni, hacsak nem teljesíti az azok által támasztott követelményeket. A katalógusfelület használata számos felhasználónak nehézkes lehet.
- **Hagyományos szemléletű feldolgozás, hagyományos hordozóformátummal, FRBR-izált keresőfelülettel.** Ebben az esetben az információkereső felület FRBR-izált megjelenítésének köszönhetően lehetőség van a találatok entitások mentén történő csoportosítására, az algoritmusok sajátosságai miatt azonban előfordulhat, hogy az automatikus entifikációs folyamat a MARC-rekordokban esetenként jelen lévő adathibák, vagy egyes, a csoportok kialakításához fontos adatmezők hiánya miatt nem releváns klasztereket képez, vagy nem végez el összevonásokat.
- **Hagyományos szemléletű feldolgozás, hagyományos ÉS korszerű hordozóformátummal, hagyományos keresőfelülettel.** Akik ezt az utat járják, vélhetően felismerték a linked data-technológia alkalmazásában rejlő könnyebb együttműködési lehetőséget az informatikai partnerekkel. Az intézmény képes arra, hogy szoftvereszközei segítségével a hagyományos adatformátum mellett

valamilyen RDF-szintaxist is szolgáltatni tudjon, amelynek eredményeképpen nem jön zavarba, ha valamely szolgáltatás fejlesztői ilyen adatokat igényelnek tőle. A gyűjtemény munkatársai és használói azonban nem részesülnek a technológia nyújtotta előnyökben, vagy amennyiben mégis, az olyan aggregációs platformon keresztül történik, ahol az egyes tagintézmények jelentősége csökken.

- **Hagyományos szemléletű feldolgozás, hagyományos ÉS korszerű hordozóformátummal, korszerű keresőfelülettel.** Természetesen lehetőség van arra, hogy az intézmény a falain belüli informatikai szaktudást kihasználva, vagy valamilyen terméket/szolgáltatást vásárolva úgy határozzon, hogy saját maga juttatja el a linked data-technológia előnyeit a felhasználókhoz. Az ilyen felületeken a megjelenítés az entitásokat követi, továbbá olyan informatikai megoldások is megtalálhatók rajtuk, amelyeket a MARC-kal nem lehetne megvalósítani.
- **Korszerű szemléletű feldolgozás, hagyományos hordozóformátummal, hagyományos keresőfelülettel.** A korszerű szemléleten az FRBR-ben – illetve ma már az LRM-ben – lefektetett alapelvek, illetve az ezeken alapuló RDA-szabályok követését kell érteni a forrásleírásban. A hagyományos formátum bizonyos, igen korlátozott keretek között lehetővé teszi az RDA szabályainak alkalmazását, illetve URI-k rögzítését az adatszerkezetbe (az ún. „hivatkozos MARC”), amivel elősegíti a későbbi konverziós eljárásokat (Wallis, 2018).
- **Korszerű szemléletű feldolgozás, hagyományos hordozóformátummal, FRBR-izált keresőfelülettel.** A megjelenítés pontosságán, a felület használhatóságán a választott FRBR-izálási technika mellett segíthet annak a ténynek a kihasználása, hogy a MARC-rekordok már URI-kat is tartalmazhatnak az erre kijelölt, szabványosított helyeken.
- **Korszerű szemléletű feldolgozás, korszerű hordozóformátummal, korszerű keresőfelülettel.** A kívánatos jövő. Ebben a forgatókönyvben a forrásleírási folyamat teljesen maga mögött tudja hagyni a MARC okozta korlátokat, és teljes mértékben megvalósíthatóvá teszi az FRBR/LRM elméleti, és az RDA gyakorlati útmutatásait, új dimenzióba helyezve ezzel a könyvtárak által nyújtott szolgáltatásokat. Az új szemléletű feldolgozást támogató felületek már ma is léteznek, azonban széles körben történő alkalmazásuk egyik nagy akadálya, hogy jelen pillanatban csak elszigetelten tudnak működni más könyvtári feladatokat támogató rendszermoduloktól. Napjainkban például még nincsen informatikai

eszköz arra, hogy a kölcsönzés gráfok segítségével legyen elvégezhető. A fejezet elején említett BIBFLOW-projekt résztvevői ezt a problémát úgy hidalták át, hogy valahányszor elkészül egy forrásleírás, és megtörténik az adatok tárolása (RDF-gráf formájában), egy csökkentett adattartalmú, csonka MARC-rekord is keletkezik a relációs adatbázisban, amelynek egyetlen feladata a többi rendszerfunkció támogatása (Smith et al., 2017).

Amíg a megfelelő mennyiségű, elérhető áron hozzáférhető szoftvereszköz nem áll rendelkezésre, az intézményeknek legalább a párhuzamos szolgáltatás lépcsőjéig, illetve az eredeti hordozóformátum RDA-kompatibilitásának maximalizálásáig érdemes eljutniuk. A párhuzamos szolgáltatás azt jelenti, hogy bár az adatok MARC-rekordokban tárolódnak, s az intézmény elsődleges keresőfelülete is MARC-alapú marad (optimálisan entifikált, azaz FRBR-izált változatban), a már ma is használható, térítésmentes megoldások valamelyikével gráfként is elérhetővé teszik az informatikai szakemberek számára az adatállományt. A házi katalógizálási szabályzatot és a MARC helyi implementációját pedig minden forrástípus esetében úgy érdemes módosítani, hogy az magába foglalja az RDA-mezőket is, így hozzájárulva az elkészülő gráf megfelelő minőségéhez és adatgazdagságához.

8. Keresőoptimalizálás a linked data segítségével

A szemantikus webben történő publikálás kiszélesíti az egyes gyűjtemények lehetőségeit. A könyvtárak, múzeumok és más intézmények által gondozott adatok az RDF-logikát támogató formátumok valamelyikére konvertálva értelmezhetővé válnak az automatikus feldolgozó folyamatok számára is, s a szemantikus formátumok új generációs szolgáltatások egész garmadáját inspirálhatják. Ezek a szolgáltatások aztán képesek kihasználni az egyes adatok között expliciten, minősítetten kifejezett kapcsolatokban rejlő lehetőségeket, s építenek az adatforrások összekapcsolhatóságára is. Az RDF-logika használata azonban fontos új eszköz lehet a keresőszolgáltatások találati listáinak kiemelkedő helyezéseikért folytatott küzdelemben, azaz a keresőoptimalizálási munkákban is.

Ahhoz, hogy egy adott weboldal megjelenhessen a keresők, például a Google találati között – azaz a látható, felszíni web része legyen – több feltételnek kell megfelelnie, e feltételeket Horváth (2018) összegzi. A legfontosabb ezek közül az *indexelhetőség*. A keresőmotorok listáiba akkor kerülhet be legnagyobb valószínűséggel egy weboldal, ha

- *állandó, saját URL-je van.* A könyvtári katalógusok találati listáiban megjelenő rekordok URL-jei nem ilyenek: lejáráó session-azonosítókat, változókat, esetenként – különösen, ha a rendszer az SRU-protokollt használja – keresési feltételek kódolt formáit tartalmazzák. A probléma megoldása, hogy minden egyes megjelenített katalógusrekordnak külön weboldalt kell készíteni, amely saját URL-lel azonosítható. Érdekes ezen felül minden ilyen weboldalnak egy tetszetős (*cool*) URI-t alkotni (a cool URI-król a 6.3. fejezetben tettem említést). Az egyszerű, rövid, humán ágens számára is értelmezhető URI-ról a használatát át kell irányítani a katalógusfelület által generált, tényleges URL-re – ezt a folyamatot linkfeloldásnak nevezik.
- *linkek mutatnak rá.* Törekedni kell rá, hogy a katalógusrekordok weboldalaira más oldalak hivatkozzanak. Ennek legcélszerűbb terepe a mindenki által könnyen, egyszerűen szerkeszthető közösségi enciklopédia, a Wikipédia, illetve társszolgáltatása, a Wikidata adatbázis. A Petőfi Irodalmi Múzeum keresőoptimalizálási projektjének részfeladatákként kísérletet tettem arra, hogy hivatkozásokkal kössem össze a PIM személynév-adatbázisának nyilvánosan hozzáférhető, kereshető rekordjait, illetve a személyeket reprezentáló Wikidata-elemeket. A Wikidata adatbázisból SPARQL-lekérdezéssel nyertem ki a Magyarországon született és élt személyek URI-jait, nevét, illetve születési-

halálózási dátumát, amelyet a Monguz Kft. munkatársainak segítségével vettem egybe a PIM gyűjteménykezelő rendszerének adatbázisával. Amennyiben egy Wikidata-elem, illetve egy személynév-authority rekord azonos személyt írt le, a Wikidata-elemet gazdagítani lehetett a PIM-rekord URI-jával, és viszont: a PIM-rekordba bekerült a Wikidata-elem URI-ja. A kísérlet eredményeként a Wiki-univerzumban körülbelül 6000, a Petőfi Irodalmi Múzeum online felületére mutató hivatkozás jött létre (Hubay, 2020b).

- *a tartalma érthető a keresőszolgáltatás számára.* Ennek érdekében a weboldal kódjába ún. szemantikus jelölőket kell elhelyezni, amelyek a keresőrobot számára értelmezik az oldal tartalmának egyes elemeit. A szemantikus jelölők szótárakból vett relációjelölőket, azaz URI-kat jelentenek. Keresőoptimalizálási célra azonban nem mindegyik szótár alkalmas; ha eredményt akarunk elérni, akkor a schema.org tematikus elemkészleteit, leíró sémáit kell használni. Ezt a készletet négy nagy informatikai cég – a Google, a Microsoft, a Yahoo és a Magyarországon kevésbé ismert, oroszországi eredetű Yandex – fejlesztette ki azzal a céllal, hogy a szótárelemek segítségével értelmezhetővé lehessen tenni a weboldalak tartalmát a keresőszolgáltatások számára. Az értelmezhetőség követelményének teljesítése önmagában véve nem jelent automatikus megjelenést a találati listákban, ugyanakkor a szemantikus jelölők felhasználásával jól strukturált metaadatok állíthatók elő a weboldalokról. Ha minden egyéb feltétel is teljesül, és az oldal a keresőindex része lesz, az eredményhalmazokban gazdagított információtartalommal (rich snippet) rendelkező találat jelenik meg róla.

A szemantikus jelölők elhelyezését a 6.4. fejezetben már említett RDFa szerializációs szintaxis teszi lehetővé, amelynek segítségével egy HTML-tag attribútumában lehet elhelyezni az adott adatelem értelmezését lehetővé tévő szótárrelációt: a példában (W3C, 2015) a Dublin Core-szótár két elemét:

```
<html>
...
<body>
...
  <h2 property="http://purl.org/dc/terms/title">The Trouble with Bob</h2>
  <p>Date: <span property="http://purl.org/dc/terms/created">2011-09-10</span></p>
...
</body>
```

A Petőfi Irodalmi Múzeum munkatársai eredményes kísérletet folytattak a személynév-authority rekordok weboldalainak szemantikus jelölőkkel való felcímkésével. A weboldalakat megjelenítő OPAC3 felület sablonjában az egyes adatelemek mellé, a property attribútum értékeként az alábbi schema.org-szótárelemek kerültek (Mohay & Hubay, 2018):

- a weboldal által leírt objektum típusa – `typeof`; értéke – `person`
- a személy neve – `name`
- foglalkozása – `jobTitle`
- születési helye, ideje – `birthPlace, birthDate`
- halálozási helye, ideje – `deathPlace, deathDate`
- testületi tagsági információk – `alumniOf`
- lakcím-információk: `address`
- díjak: `award`

A munka a jövőben tovább folytatódik; a PIM humáninformatikus szakemberei a személynév-authority rekordok mellett a múzeum tárgyleíró rekordjait is szeretnék felcímkézni, amelyhez a schema.org által kínált speciális CreativeWork-elemkészletek nyújtanak majd segítséget. Ezek felhasználásával a könyvek, kéziratok, eredeti fényképek, hangfelvételek, képzőművészeti alkotások metaadatelemei is megjelölhetők a megfelelő szótárelemekkel, ezáltal értelmezhetővé válnak a keresők számára.

9. A Bibliographic Framework (BIBFRAME) – az új MARC?

A BIBFRAME néven ismertté lett új bibliográfiai keretrendszer elméleti alapjai már az 1990-es évek közepén készen álltak. Olivia Madison tanulmányában (2005) utal egy 1994-ben készült munkadokumentumra, amelyet a Library of Congress adatmodellezéssel foglalkozó munkatársai alkottak meg – válaszul az IFLA által kidolgozott FRBR-modellre –, amely elveti az FRBR négyentitásos megközelítését, és csupán hármat tart meg ezek közül:

- a művet, amely a szakmai anyagban megfogalmazottak szerint intellektuális, művészeti vagy kreatív tartalmat jelöl,
- a megjelenési formát, amely azokat a fizikai jellemzőket foglalja magába, amelyeken keresztül a mű megvalósul, és
- a példányt, amely egy megjelenési forma egyetlen konkrét példája.

A Library of Congress szakmai munkacsoportja 2008-ban átfogó jelentést tett közzé a bibliográfiai számbavétel jövőjére vonatkozó elképzelésekről. Az áttekintés főbb csomópontjai a világháló megdőbbségre gyors térnyerése, a könyvtárhasználók számának csökkenése, a gyűjtemények által kezelendő bibliográfiai forrástípusok körének szélesedése, illetve a MARC-kal szembeni növekvő elégedetlenség voltak. A könyvtár szakemberei úgy vélték, hogy a bibliográfiai kontroll a jövőben a webre épülve kell végezni, ugyanakkor továbbra is jellemző lesz rá a nemzetközi együttműködés, és nemzetközi látókör, továbbá a decentralizált munkavégzés. Új elemként tartalmazza viszont a magánszektor, azaz a könyvtárhasználók bevonását a munkafolyamatba, ami által az adatok több forrásból fognak származni, így a bibliográfiai kontroll dinamizálható. A Library of Congress munkatársai szerint a változások minél előbb végrehajtandók annak érdekében, hogy a könyvtár megőrizhesse pozícióját a hiteles információk szolgáltatásának területén (Library of Congress, 2008).

Az időközben kidolgozott új forrásleírási szabályzat, az RDA megjelenése, továbbá annak entitásközpontú, a MARC-kal csak részben kompatibilis szemlélete még sürgetőbbé tette a korszerű könyvtári információszolgáltatásról való gondolkodást. 2011. május 13-án a Library of Congress nyilatkozatot adott ki *Bibliográfiai keretrendszerünk átalakítása* címmel, amelyben az alábbi feladatokat tűzték ki:

- meg kell határozni a jelenlegi metaadat-tárolási rendszerek azon elemeit, amelyeket meg kell őrizni, illetve továbbfejleszteni egy újabb formátumba. A vizsgálatba – más kezdeményezések mellett – a MARC 21-et is be kell vonni;
- kísérleteket kell folytatni a szemantikusweb-technológiával, hogy kiderüljön, milyen előnyöket biztosít ez a technológia a könyvtárosok közössége számára, továbbá hogyan kell módosítani a jelenlegi adatmodelleket az említett előnyök még hatékonyabb kihasználása érdekében;
- szorgalmazni kell a könyvtári metaadatok minél szélesebb körben történő újrafelhasználását, hogy a felhasználók minél több helyen megtalálhassák azokat, s lehetőségük legyen azok innovatív használatára;
- biztosítani kell a felhasználóknak az entitások mentén történő keresést, hogy azok segítségével pontosabb találati halmazokat lehessen nyerni a könyvtári információkereső felületeken, továbbá a világhálón;
- ki kell dolgozni a jelenleg létező metaadatok átvitelének tervét az új bibliográfiai rendszerekbe (Marcum, 2011).

2011. októberében a Library of Congress újabb nyilatkozatban rögzítette, hogy a könyvtáraknak és más, érdeklődő közösségeknek együtt kell dolgozni azon, hogy egy olyan hordozóformátum szülessen, amely tartalmazza a releváns bibliográfiai adatokat, továbbá biztosítja azok folyamatos cseréjét a könyvtári közösségen belül, és azon kívül is (*A Bibliographic Framework for the Digital Age*, 2011). Az 5.2. fejezetben említettek alapján a MARC ez utóbbi követelménynek nem felel meg, hiszen minden, külső partnernek előkészítendő adatátadást aprólékos konverziós fázisnak kell kiegészítenie.

A dokumentum szerzője, Deanna Marcum nem táplál illúziókat a folyamat sebességével kapcsolatban, és helyesen látja, hogy a MARC még igen sokáig, várhatóan még évekig a könyvtári munka alapját fogja képezni, egészen addig, amíg az intézmények nem készítik el a teljes átállás ütemtervét. Ezért nem lehet egyik pillanatról a másikra felhagyni a MARC szoftveres támogatásával, ugyanakkor az örökségként ránk maradt bibliográfiai információhalmaz nem lehet akadálya annak, hogy felelősségteljes döntések szülessenek a jövő generációk könyvtárai, illetve könyvtárosai érdekében.

Kiegészítésül a Library of Congress közzétette az új bibliográfiai keretrendszer általános tervét is, amelynek igen sokatmondó része a rendszerrel szemben támasztott követelmények leírása. Az intézmény által meghatározott főbb irányvonalak a következők:

- az új keretrendszer támogassa a leírási szabályzatok széles körét, azaz ne csupán az RDA-ban, hanem más rendelkezésekben (például DACS, VRA, CCO stb.) részletezett metaadatelemek rögzítését is tegye lehetővé;
- tudjon fogadni olyan adatokat, amelyek rendszerint kiegészítik a bibliográfiai források leírásait: példányadatokat, authority adatokat, osztályozási jelzeteket, megőrzési, technikai és jogi metaadatokat;
- legyen képes szöveges adatokat (literálokat) megjeleníteni, ugyanakkor álljon készen a linked data-technológia támogatására, mindenekelőtt az URI-k használatának biztosításával;
- vélhetően jobban le kell választania az adattárolási logikákat, a megjelenítő felületeket, továbbá a keresőürlapok szerkezetét a hordozóformátumról, mint ahogyan az ma megvalósul;
- elégítse ki a legkülönbözőbb típusú könyvtárak metaadatokra vonatkozó igényeit – a kisméretű közkönyvtáraktól kezdve egészen a nagy kutatókönyvtárakig;
- legyen kompatibilis a MARC-kal, amíg szükséges; biztosítsa a rekordokban tárolt valamennyi adat megőrizhetőségét, mivel az összes forrás újrakatalogizálása nem megoldható;
- álljon rendelkezésre a kétirányú konverzió lehetősége, megkönnyítve ezzel a fejlesztési munkákat (*A Bibliographic Framework for the Digital Age*, 2011).

A Library of Congress az új bibliográfiai keretrendszert az RDF segítségével kívánta megvalósítani, amellyel két célt kívántak elérni egyidejűleg: lehetőséget adtak az RDA rendelkezéseinek teljes körű alkalmazására, a feldolgozási folyamat entitásszemléletűvé tételére, valamint jelentős mértékben megnövelték az adatok újrafelhasználásának és a bibliográfiai adatokra épülő új szolgáltatások megszületésének valószínűségét.

9.1. A BIBFRAME első változata

A Library of Congress 2012. májusában tette közhírré, hogy szerződéses viszonyt létesített a Zepheira nevű informatikai céggel, amelynek munkatársai már akkor is jelentős tapasztalatokkal rendelkeztek a linked data-technológiával kapcsolatban. A könyvtár azt a feladatot szánta partnerének, hogy dolgozzon ki egy modellt (vagy modelleket), amelyek a későbbi megbeszélések, viták, finomítások alapját képezhetik, továbbá tekintse át a linked data-technológia könyvtári alkalmazásának körképét. A munka körülbelül fél évig tartott; a BIBFRAME alapidokumentumát, amely a *„Bibliographic Framework as a Web of Data:*

Linked Data Model and Supporting Services” címet viseli – és a szakirodalomban egyszerűen BIBFRAME Primerként szokás rá hivatkozni – 2012. november 21-i dátummal publikálták. A tanulmány bevezetőjében a Library of Congress leszögezi: az új hordozóformátum implementációja nem egyszerűen a MARC felváltását jelenti valami korszerűbbel. A BIBFRAME segítségével a világháló részeként definiálható újra a bibliográfiai kontroll munkafolyamata, amelynek segítségével a könyvtári közösség egy globális méretű adathálózat integráns részévé válhat (Library of Congress, 2012).

A Library of Congress – erre már Madison (2005) alapján korábban utaltam – saját entitás-kapcsolat modellel rendelkezett, amely azonban nem vágott egybe az IFLA által megjelentetett funkcionális modell felosztásával. A Library of Congress három entitásra alapuló adatmodellje jelenik meg a BIBFRAME Primerben is, ahol a szerzők Tom Delsey munkájára (2002) hivatkoznak, aki elvégezte a MARC 21 bibliográfiai és példány formátumának funkcionális elemzését. Vizsgálata eredményeként megállapította, hogy a rekordokban található adatok három fontos csoportba oszthatók:

- az alkotások szellemi tartalmához, lényegéhez kapcsolódó adatok;
- az alkotások bizonyos megjelenési formáihoz kapcsolódó adatok; ezek a megjelenési formák kézbevehetők, vagy letölthetők a hálózatról;
- a rekord metaadatai, ellenőrzőszámok, és egyéb annotációk.

Ebből adódóan a BIBFRAME alapstruktúráját kidolgozó Zepheira az alábbi osztályszerkezetet valósította meg:

- *Work* (avagy Creative Work) – a katalogizálandó bibliográfiai forrás fogalmi tartalma;
- *Instance* – egy mű egyedi, anyagi megtestesülése, fizikai vagy digitális formában. Egy Instance kizárólag egyetlen műhöz kapcsolódhat.
- *Annotation* – segítségével más osztályok (leggyakrabban Work és Instance) egyedi előfordulásaihoz kapcsolódó információk rögzíthetők, mint például borító, példányadat(!), értékelés, összefoglaló, vagy akár tartalomjegyzék. Az osztály tulajdonságai segítségével rögzíthető maga a kiegészítő információ (`bf:annotationBody`), annak forrása (`bf:annotationAssertedBy`), illetve tipizált kapcsolatjelölők állnak rendelkezésre az annotáció típusának kifejezésére (`bf:summaryOf`, `bf:tableOfContentsFor`, `bf:coverArtFor`).

- *Authority* – a Work-höz, illetve az Instance-hez, esetleg az Annotation-höz kapcsolódó személyek, családok, testületek, események, helyek, témák stb. leírását biztosítja. Keretet kíván biztosítani az egyes bibliográfiai ügynökségek által rögzített egységesített névalakok, illetve az ahhoz kapcsolódó URI-k tárolásához. Alosztályai az egyes, hozzáférési pontként használható névtípusokat jelenítik meg. Használható tulajdonságai közül legfontosabbak a `bf:AuthorizedAccessPoint`, amely az egységesített névalakot karakterlánc formájában tárolja, illetve a `bf:hasAuthority`, amelynek értéke egy URI, ahonnan az autorizált névalak származik. Ha szükséges, az autorizált névalak elemei a MADSRDF-szótár relációival is kifejezhetők, illetve további hivatkozások is rögzíthetők a `bf:referenceAuthority` tulajdonság segítségével – ez utóbbi lehetőség adatgazdagításként is értelmezhető.

A BIBFRAME-ben tehát nem szerepel a Kifejezési forma entitás; azt egy igen sajátos megoldással helyettesítik: jellemzőit belefoglalják a Work leírásába, ugyanakkor definiálnak egy `expressionOf / hasExpression` (kifejezése ennek / kifejezése a következő) relációpárt, amely – paradox módon – két mű között jelölendő ki. A Work osztály mindemellett több alosztállyal rendelkezik (például Text, Audio, MovingImage), amelyek ugyancsak az FRBR-i értelemben vett kifejezési formát írják le.

Hasonlóan érdekes a példányadatok annotációként való feltüntetése. A BIBFRAME első verziójában a konkrét példányt reprezentáló URI-t egy `bf:HeldItem` nevű osztályba kellett besorolni, majd azt a `bf:holdingFor` relációval a megfelelő Instance-hez kapcsolni, továbbá a birtokló intézmény (a `bf:Agent` osztály egy előfordulása) is rögzíthető volt a `bf:heldBy` reláció segítségével. A vonalkódok, polcjelzetek, specifikus példányszintű megjegyzések rögzítésére alkalmazható relációk értelmezési tartománya minden esetben a `bf:HeldItem` osztály volt (*BIBFRAME Vocabulary List View*, é.n.).

9.1.1. Rob Sanderson kritikája

2015 áprilisában részletes elemzés jelent meg a BIBFRAME-mel kapcsolatban (Sanderson, 2015), amely a szókészlet megfelelőségét vizsgálja a linked data jó gyakorlatainak szempontjából. A körülbelül ötven oldalas szöveg szerzői az alábbi részterületeket vizsgálták:

- a szótár megfelelően képezi-e le a választott terület (ebben az esetben a könyvtártudomány) forrásait a bevezetett osztályok segítségével;

- a szótár csak a választott területen belül definiál-e kifejezéseket, lehetővé teszi-e a más kezdeményezésekkel történő átfedés elkerülését, hogy a használóknak ne kelljen dönteniük az azonos kapcsolattípust kifejező relációk között;
- egy célra csak egyetlen kifejezést definiál-e a szótár – elkerülendő az adatmodell belső inkonzisztenciáját;
- URI-kkal történik-e a források azonosítása karakterláncok helyett;
- minden URI pontosan egy forrás azonosítását végzi-e;
- az URI-k „beszédese”-e – természetes nyelvű szöveget tartalmaznak-e, vagy mechanikus, például sorszámot;
- tartalmaznak-e az URI-k változó részeket, például dátummegjelölést;
- tartalmaznak-e az URI-k hashtaget (#);
- az URI-k felkeresésével használható információkat kap-e a felhasználó;
- található-e linkek más adathalmazok felé;
- található-e üres csomópontok, RDF-konténerek, illetve reifikáció az adathalmazban;
- a szótár újrahaznosít-e más elemkészletekben található relációkat;
- a szótárelemek elnevezése egységes-e;
- megfelelően, illetve megfelelő mértékben vannak-e definiálva az osztályok használatára vonatkozó megkötések (értelmezési tartomány, értékkészlet);
- definiálva vannak-e az inverz kapcsolatok;
- publikálva van-e a szótár, vagy csak a vele készült adathalmazok.

A tanulmány ezt követően javaslatokat fogalmaz meg, melyek közül a legfontosabbak az alábbi témák köré csoportosíthatók:

- az authority kezelése. A tanulmány szerzői az autorizált adatok modellezését nem tartják szükségesnek a linked data-ökoszisztémában, ahol mindent egyedi URI-k azonosítanak. Ezért törlésre javasolják a `bf:Resource` magas szintű osztályt, továbbá a `bf:Authority`-t, amelynek alosztályait helyettesíteni javasolják a FOAF-ból (Organization, Person, Agent), a `schema.org`-ből (Agent, GovernmentOrganization, Organization, Person), a VIVO-ból (Meeting, Agent), illetve a PROV-ból (Agent), sőt a Bibliographic Ontology-ból (Conference) származó osztályokkal;

- a leírás tárgyának alapos elkülönítése. Egyes tulajdonságok – mint például a `bf:changeDate`, `bf:creationDate`, `bf:descriptionSource` – az egyes forrásokat leíró MARC-rekordokra utalnak, azok keletkezési, módosítási dátumát, illetve azok forrását jelölik, s nem szerencsés, ha ezek a relációk keverednek a *magáról az entitás-előfordulásról* szóló állításokkal. A javaslat kidolgozói az adminisztrációs metaadatok elkülönítését, továbbá a vonatkozó BIBFRAME-relációk VOID-elemekkel történő helyettesítését javasolják;
- az annotációk kezelése. Azokban az esetekben, ahol az annotáció egy forrás értékelését (review), vagy összefoglalóját jelenti, érdemes volna a W3C által szabványosított Open Annotation szótár megfelelő elemeit használni: az `oa:Annotation` osztályt, illetve a két annotációtípus céljának megfelelő két tulajdonságot: az értékelésekhez az `oa:commenting`, míg az összefoglalókhöz az `oa:describing` tartozik;
- a példányok kezelése. A tanulmány szerzői helytelenítik, hogy a példányok a `Work`, illetve az `Instance` osztály előfordulásaihoz képest másodrendűek, és csupán az annotációk egy alosztályát alkotják. Ez definíciós szinten is problémát okoz, hiszen az annotáció definíciója – olyan forrás, mely kiegészítő információkat tartalmaz más BIBFRAME-forrásokról – nyilvánvalóan nem alkalmazható a példányokra;
- az azonosítók működése – a szerzők az egyes azonosítósémákat reprezentáló tulajdonságok helyett az `Identifier` osztály alosztályait javasolják kialakítani;
- a címek kezelésének átalakítása, ugyancsak alosztályok létrehozásával, amelyek az egyes címtípusokat (kulcscím, rövidített cím stb.) reprezentálják.

A BIBFRAME szótár legnagyobb gyengeségeit a tanulmányt kidolgozó szakemberek az értéként literált váró tulajdonságok nagy számában (118), valamint a külső adatforrásokból vett entitás-előfordulásokra (például személyekre, földrajzi helyekre, nyelvekre stb.) mutató hivatkozások hiányában látják, továbbá abban a tényben, hogy a BIBFRAME lényegében egyáltalán nem használ fel más szótárakból származó elemeket, még akkor sem, ha azok szemantikája teljesen kielégítené a definiált célokat: a `bf:label`-t tökéletesen helyettesítheti az `rdfs:label`, illetve a `bf:hasPart` helyett is állhatna a `dcterms:hasPart`.

Rob Sanderson egy 2015 decemberében tartott előadásában pontértékeket rendelt az egyes ajánlások megvalósulásához a BIBFRAME-ben. Az adatmodell körülhatároltságára vonatkozó ajánlások (4) közül a szótár mindössze egyet teljesít, az URI-k kezelésének területén megfogalmazott hat ajánlásból négynek felel meg, az információ használhatóságát

biztosító javaslatokból (6) kettő teljesül, s az újrafelhasználás területe is igen gyengén jelenik meg – hat szempontból egy valósul meg. A szótár a szakember értékelési szempontrendszer alapján 22 pontból nyolcat szerzett meg, ez 36%-os eredmény (Sanderson, 2015b).

Csak helyeselni lehet az Authority osztály megkérdőjelezését, amely a cédulakatalógus hagyományainak továbbörökítési szándékáról (vagy legalábbis az elszakadás nehézségeiről), az informatikai háttér és az új fogalmak csupán részleges megértéséről tanúskodik. Az egységes forrásazonosító megőrzi és továbbfejleszti a hozzáférési pont, és az abban tárolt kitüntetett névalak valamennyi funkcióját, a két megoldás egymás melletti alkalmazása felesleges redundanciát idéz elő a konvertált, vagy frissen előállított gráfszerkezetekben. Az adat-újrafelhasználás jegyében dicséretes a külső szótárrelációk meghatározása is az Authority, illetve Annotation osztály egyes elemeinek helyettesítésére, amellyel az adatok még inkább beágyazhatók a globális adathálózatba.

Ezen kívül érdemes tekintettel lenni az említett 118, értékül literált váró relációra. A linked data technológia egyik előnye a MARC-kal szemben, hogy lehetőség van azonosítókkal reprezentálni az egyes relációk értékeit (azaz a tripletek tárgyait), lecsökkentve ezzel a gépelési hibák mennyiségét. A külső adatforrások becsatornázásával kényelmesebbé, gyorsabbá, nem utolsósorban pedig pontosabbá válik a katalógizáló munka, csökken a téves adatbevitel szám, az adathalmaz pedig megint csak jobban beágyazhatóvá válik a nemzetközi szinten összekapcsolódó adatok globális méretű grájfjába.

9.2. A BIBFRAME 2.0

2015 októberének végén a BIBFRAME 1.0 korai tesztelőitől összegyűjtött javaslatok, valamint a Library of Congress által megvalósított pilot projektből leszűrt észrevételek alapján megkezdték a modell alapos átstrukturálását. Ebből a célból a könyvtár munkatársai hét, piszkozati (draft) állapotban lévő specifikációt tettek közzé véleményezésre a bibliográfiai keretrendszer weboldalán (*BIBFRAME Model & Vocabulary*, é.n.), amelyek az alábbi területeket érintették:

- a címek kezelésének változásai;
- a közreműködői funkciók és az authority-kezelés változásai;
- a példányadatok kezelésének módosulásai;
- az események kezelésének átalakulása;
- az azonosítók és megjegyzések kezelésének változásai;

- a különféle adminisztratív metaadatok kezelésével kapcsolatos módosulások;
- a különféle kategóriák kezelésének változásai.

A szakmai fórumokon, köztük a BIBFRAME-levelezőlistán érkezett visszajelzések feldolgozását és beépítését követően a szókészlet második verziójának hivatalos megjelentetése 2016. áprilisában történt meg, amelyen 2017. februárjában kisebb korrekciókat végeztek. Az új modellben a korábbi 53-mal ellentétben 186 osztály található, míg a tulajdonságok száma 289-ről 137-re csökkent. (A BIBFRAME első és második változatának teljes osztály- és tulajdonságlistája a disszertáció *1. mellékletében* olvasható). A kiemelten fontos változások az alábbiak:

- a példányadatok kezelése alapvetően megváltozott, mivel a modell második változatából kiiktatták az annotációk osztályát. A BIBFRAME 2.0-ban egy új magosztályt hoztak létre, Item megnevezéssel, amelynek tagjai a `bf:itemOf` relációval kapcsolhatók a megjelenési formák leírásaihoz. A példányok használhatóságának, illetve hozzáféréseinek paraméterei, fizikai vagy virtuális lelőhelyei, saját azonosítói, illetve az azokat birtokló intézmények mind leírhatók a megfelelő tulajdonságok, illetve osztálybesorolások alkalmazásával;
- a BIBFRAME 2.0-ban már nem használják az Authority magosztályt sem: a 2015-ben megfogalmazottak szerint a gráfok a jövőben nem authority rekordokra, hanem RWO-URI-kra mutatnak majd. Az Authority osztály alosztályai – az Agent, a Topic, a Temporal és a Place – innentől kezdve a Resource csúcsosztály alá rendelődnek, köztes szint nélkül. A `bf:authorizedAccessPoint` reláció, mely literálként tartalmazza az adott entitás-előfordulás autorizált névalakját, megszűnik, helyére egyszerű `rdfs:label` kerül, mely akár több nyelven is megadható. A `bf:hasAuthority` viszonyt, amely az entitás-előfordulást leíró authority rekordra mutató URI-t várt értéként, kiváltja a `bf:identifiedBy`. Az eddig `bf:referenceAuthority`-ként megkülönböztetett azonosító URI-k ugyancsak a `bf:identifiedBy` relációval kapcsolódnak az entitás-előforduláshoz (*BIBFRAME Authorities*, 2014). Az autorizált névalakok tárolását azonban nem hagyta veszni a Library of Congress – a 9.3.5. fejezetben rövidesen ismertetendő, házi szótárkiegészítéssel megőrizhetővé tették ezeket a névformákat is.
- a BIBFRAME második változata vezette be az Event (esemény) osztályt. A leírás szerint az esemény olyan valami, amely egy meghatározott helyen, illetve időben történik, s amelyet források dokumentálnak. Ilyen esemény lehet egy tudományos

konferencia, sportesemény, színházi előadás, koncert, beszéd stb. Az esemény természetesen képezheti művek tárgyát, erre jó példa a 2018. évi firenzei BIBFRAME Workshopról (esemény) készült, e sorok írója által készített beszámoló (mű), amely egy szakmai folyóiratban jelent meg. Amennyiben ezt a viszonyt gráfban szükséges ábrázolni, a mű, illetve az esemény között a `bf:subject` kapcsolatjelölő tulajdonságot kell alkalmazni, jelezve, hogy az adott mű tárgya a gráf egy másik szakaszában metaadatolt esemény. Azonban elképzelhető, hogy az eseményt videóra veszik, ekkor ez a mű nem egyszerűen *szól* az eseményről, hanem maga az esemény jelenti a szellemi tartalmat, amely a műben megjelenik. Ilyen esetben az esemény, illetve a mű leírása között a `bf:eventContent` tulajdonság, illetve inverze, a `bf:eventContentOf` reláció rögzítésével kell kapcsolatot létesíteni. E tulajdonságok a `relatedTo` általános kapcsolatjelölő reláció altulajdonságai. Előbbinek az értelmezési tartománya az Event osztály tagjaira, míg értékkészlete a Work osztály tagjaira korlátozódik, utóbbinál éppen fordítva: értelmezési tartományát a művek, míg értékkészletét maguk az események képezik. Továbbá egy esemény több műben is rögzítve lehet, e művek mindegyike a fenti két kapcsolatjelölővel kötődik az Event osztály egy konkrét tagjához; illetve előfordulhat, hogy az egyes rögzítések a teljes esemény csak valamely részletét tartalmazzák; ekkor a rögzítésekben megjelenő rész-események, továbbá a teljes esemény egymással rész-egész viszonyt alkot, amelyet a gráfban is reprezentálni kell, mégpedig három egyedi Eventként (*BIBFRAME 2.0 Event Model*, 2017).

- A modell első változatában a közreműködői funkciók nagyon korlátozott módon voltak csak leírhatók. A BIBFRAME 2.0-ban ezért bevezettek egy `bf:Role`, illetve egy `bf:Contribution` nevű osztályt, ahol az Agent osztályba tartozó entitás-előfordulások URI-jait összerendelhetővé tették a Library of Congress által kezelt, közreműködői funkciókat tartalmazó kötött szótár URI-jaival. E funkciójelölőket természetesen a Role osztály tagjaiként kell értelmezni, majd igény szerint felcímkézni. Szükség esetén az elsődleges felelősségi funkció megjelölhető a `bf:PrimaryContribution` osztálybesorolással.
- A legkülönfélébb azonosítók rögzítését a modell második változatában már nem külön tulajdonságok megadásával, hanem osztálybesorolással kell elvégezni. Az adott osztály (például `bf:Isbn`) egy konkrét előfordulását – `rdf:value`-ként rögzítve – a `bf:identifiedBy` reláció kapcsolja az azonosított entitás-előfordulás

leírásához. Az egyes azonosítótípusokat jelentő diszkrét osztályok száma meghaladja a negyvenet.

Rob Sanderson, akit a BIBFRAME 1.0 verziójának kritikája vonatkozásában említettem, megismételte elemzését az új adatmodell-változatra vonatkozóan is. A vizsgálat ugyanazon kritériumok mentén zajlott le, mint a korábbi szótárverzió esetében. Bár történtek előrelépések, az új verzió is csupán 12,5 pontot szerzett az elérhető 22-ből, amellyel 57%-osra növelte a linked data jó gyakorlatai teljesülésének részarányát. Még mindig kevés a külső adathalmazok újrafelhasználásának aránya, amelynek növelése érdekében igen nagy szükség volna például egy átfogó, valid adatokat tartalmazó műcím-névtérre, amelyet a bibliográfiai leírások első, legmagasabb absztrakciós szintjén fel lehetne használni. A VIAF-ban tárolt művek és kifejezési formák címei jó kiindulópontjai lehetnének ennek. Számos olyan adatelem van ezen kívül a BIBFRAME-ben, amelyek értékészletét az RDA Registry szótárkódolási sémái adhatnák (például a `bf:content`, `bf:media`, `bf:carrier` elemek). Ennek deklarálása jóval tisztább BIBFRAME-gráfokat⁴¹ eredményezne, jobban biztosítaná azok szabványosságát.

9.3. A BIBFRAME kiegészítései

9.3.1. A Zepheira és a bibfra.me

A marcr.org, illetve a bibfra.me oldalon egyaránt hozzáférhető szótárak lényegében alkalmazási profilokként születtek meg a BIBFRAME alapvető szókészletéhez. Mégsem lehet őket pusztán AP-ként definiálni, hiszen nem csupán bizonyos szempontok, illetve felhasználói igények szerinti elemválogatás valósul meg bennük, hanem hozzáadott relációkat is tartalmaznak, amelyek az alapszótárban nem találhatók meg. Minden ilyen szótár az ún. BIBFRAME Lite szókészletre épül, amely 28 osztályt, illetve 72 tulajdonságot tartalmaz, ezek szemantikája sok esetben azonos a BIBFRAME-ben, sőt a schema.org szótárban leírt más osztályokkal, így URI-jaik egymással megfeleltethetők. A tulajdonságok közül érdemes megemlíteni az Event osztály előfordulásainak jellemzésére használható, első pillantásra talán humoros *when*, *where*, *who*, *why*, illetve *what* megnevezésű relációkat,

⁴¹ Az értekezés 5.2. fejezetében beszéltem a MARC-formátum kifejezés szigorú értelemben vett helytelenségéről. Hasonló terminológiai probléma jelentkezik a „BIBFRAME-formátum” kifejezés esetében is: adatformátumon ugyanis a szemantikus web technológiájának kontextusában a 6.4. fejezetben tárgyalt szerializációs szintaxisokat kell érteni. Helyesebbnek tartom a BIBFRAME szótár relációival készült adatszerkezetre a „BIBFRAME-gráf” megnevezést használni.

amelyek az eseményekhez kapcsolódó egyes dátumtípusok (születés, halálozás, copyright stb.), földrajzi helyek, kapcsolódó ágensek, fogalmak leírására használhatók.

A Library elnevezésű szótár számos, a bibliográfiai források legkülönfélébb típusait reprezentáló osztállyal (festschrift, nonmusical, kit stb.), illetve ezek speciális ismérveit megjelenítő tulajdonsággal gazdagítja a Lite készletét. E speciális ismérvek vonatkozhatnak a felvétel technikájára (*recording technique*), a számítógépes állomány típusára (*type of computer file*), vagy akár Braille-formátumú zenelejegyzésekre is (*Braille music format*).

A Relation szótárban található a különféle közreműködői funkciók és egyéb kapcsolattípusok (származéka, fordítása stb.) modellezésére használható tulajdonságok, csaknem 350, amelyek részben megfeleltethetők a MARC-szabványban definiált relátoroknak, illetve az RDA entitás-előfordulások jellemzésére használt szótárelemeinek. Ilyen például a `<http://bibfra.me/vocab/relation/engraver>` URI-val jelölt elem, amelynek értékészlete az Agent osztály. Az RDA Registry-ben e tulajdonságnak összesen öt elem feleltethető meg, amelyek az egyes ágenstípusok konkrét megjelölésében különböznek egymástól (*engraver agent, engraver person, engraver family* stb.)

Az Archive kiegészítő szótár levéltári anyagok leírására használható osztályokat, illetve tulajdonságokat definiál, a BIBFRAME Lite elemei mellé (például Finding Aid, Box, Folder, EAD Identifier stb.) A ritka anyagok (rare materials) relációi a Lite és a Library szótárra épülnek, s csak minimális mértékben gazdagítják azokat. Az Early Printing Trade nevű szókészlet az amerikai könyvkereskedelem területének, illetve ágenseinek relációit teszi modellezhetővé.

9.3.2. Art and Rare Materials

A Stanford University-n futó Linked Data For Production projekt első fázisa (2016-2018) keretében számos olyan szókészlet készült el, amelyek bizonyos speciális bibliográfiai forrástípusok leírásához nyújtanak segítséget, azaz biztosítják azokat az URI-kat, amelyek olyan speciális metaadatok leírására szolgálnak, amelyeket a BIBFRAME szótára nem tartalmaz. E kezdeményezések a Mellon Foundation által finanszírozott projekt alfeladataiként valósultak meg, s azokat a forrástípusokat célozták, melyek igen gyakran előfordulnak a könyvtárak gyűjteményeiben: művészeti, térképészeti anyagok, mozgóképek, rögzített zenei anyagok, illetve ritka, egyedi dokumentumok.

Az ArtFrame néven (később Art and Rare Materials, ARM) ismert szótár fejlesztésének történetét, adatmodelljének főbb jellegzetességeit, valamint implementációját

O’Keefe és szerzőtársai mutatták be írásukban (O’Keefe et al., 2019). A szókészlet kidolgozói között ott található a számos kultúrkörből, illetve időszakból csaknem 12.000 műalkotást regisztráló Columbia University, amelynek metaadat-leírásai sokáig egy Art Properties nevű metaadatséma alapján készültek el. Az Art Properties fő megalkotója az Avery Architectural & Fine Arts Library volt, s nem csupán a sémát, hanem annak MARC 21-megfeleltetését is előállították, így a két metaadat-formátumban tárolt leírások átjárhatóvá váltak. Az Art Properties kidolgozása során több külső elemkészlet, szótárkódolási séma elemeit is felhasználhatóvá tették a metaadatelemek értékeinek megadásakor: e szótárak közül érdemes megemlíteni a Library of Congress, illetve az amerikai közös katalogizálási program, a PCC által használt nemzeti authority állományt, az LCNAF-ot, a Getty által üzemeltetett földrajzinév-tezauruszt (Getty Thesaurus of Geographic Names, TGN), továbbá az ugyancsak a Getty gondozásában lévő művészeti-építészeti tezauruszt (Art & Architecture Thesaurus, AAT).

Az új szótár fejlesztési munkái az Art Properties, illetve a BIBFRAME elemeinek vizsgálatával, továbbá más, művészeti objektumokat leíró adathalmazok elemzésével kezdődtek. Ennek során megállapítást nyert, hogy a legtöbb hasonló intézmény a linked data-technológián alapuló állományokat a CIDOC által kidolgozott múzeumi adatmodellre (a CRM-re), továbbá az Europeana európai digitális könyvtár modelljére, az Europeana Data Modelre (EDM) építve állítja elő, továbbá említést érdemel a VRA-RDF néven ismert szótár is. A művészeti alkotások metaadat-modellezéséért felelős ArtFrame Groupban, illetve a Cornell University által koordinált Rare Materials Ontology Extension Groupban tevékenykedő szakemberek hamar felismerték, hogy a műalkotások, továbbá az egyedi, ritka források leírására használható metaadatelemek halmaza igen sok átfedést mutat, ezért egyesítve erőiket, egyetlen közös szótár kidolgozása mellett döntöttek, elvetve ezzel a két diszjunkt modell fejlesztésére irányuló törekvéseket. Innentől kezdték a készülő adatmodellt ARM-nek nevezni. Mivel a cél egy kiegészítő szókészlet előállítása volt, ezért a kidolgozók kizárólag akkor definiáltak új szótárelemeket, ha azok más helyeken nem voltak megtalálhatók, vagy használatuk az ott megjelenő szemantikai korlátozások miatt nem volt lehetséges.

A tevékenységek, illetve az elkészült szótár dokumentációja a <https://ld4p.github.io/arm/> oldalon tekinthető meg. A legterjedelmesebb az ún. mag-elemkészlet, csaknem ötven osztályt, és nagyjából azonos mennyiségű objektumtulajdonságot definiál. A fontosabb osztályok:

- *Marking* (jelzés), amely olyan szimbólumot, vagy más rájegyzést jelöl, amelyet a leírandó objektum előállítása során, vagy azt követően helyeztek el az objektumon, és amely információt tartalmaz az objektum eredetéről, annak készítőjéről, hitelességéről (*authenticity*), vagy az állapotában bekövetkezett valamilyen változásról. A *Marking* osztály számos alosztályt definiál, amelyek a következők:
 - *Autograph* (kézzel írt bejegyzés, többnyire egy személynév);
 - *Binder's ticket* (a könyvkötő jelvénye) – egy apró vésett vagy nyomtatott címke, amelynek elhelyezése a XVIII-XIX. században volt divatban, s a könyvkötő nevét, illetve címét tartalmazta, s többnyire valamelyik előzéklap sarkában volt megtalálható;
 - *Inscription* (felirat) – az objektumon található kézírásos rájegyzéseket foglalja magába;
 - *Label* (címke) – az objektum azonosításához szükséges adatokat tartalmaz, s többnyire különálló papír- vagy bőrdarabon, esetleg szövetdarabon vagy más anyagon található, amelyet az objektumhoz rögzítettek;
 - *Seal* (pecsét) – azonosítási vagy dekorációs céllal valamilyen puha anyagba (viasz, agyag stb.) nyomott monogram, írás, szám vagy szimbólum, ábra;
 - *Stamp* (bélyegző);
 - *Watermark* (vízjel).

A *Marking* osztály konkrét tagjai egyetlen tulajdonságpár segítségével kapcsolódnak a leírt objektumokhoz: *marked by* (jelöli ez) – *marks* (jelöli ezt).

- Az *Enclosure* (burkolat) osztály alosztályai a következők:
 - *Binding* (kötés);
 - *DustJacket* (védőborító);
 - *SlipCase* (egyik oldalán nyitott tok, amelybe általában könyveket csúsztatnak);
 - *WrapAroundBand* (~körbekötő szalag) – többnyire promóciós célokra használt, a védőborítóval megegyező hosszúságú, de annál jóval keskenyebb szalag, mely körbeveszi a kötetet;
 - *Wrapper* (~csomagolás) – olcsó, többnyire ideiglenes, papírból készült borító a késői XVIII. – korai XIX. századból.

A mag-elemkészleten kívül több kisebb relációjegyzék is készült, ezért az ARM-et helyesebb nem önmagában álló szótárként, hanem összetartozó szótárak családjaként meghatározni. Ezek a kisebb szótárak a különféle bibliográfiai forrásoknak vagy bibliográfiai ágenseknek kiosztott díjak adatait teszik leírhatóvá (Award Ontology), bemutathatóvá teszik egy objektum birtoklástörténetét (Custodial History Ontology), illetve lehetőség van az objektumok bizonyos releváns mérőszámainak rögzítésére (Measurement Ontology). A készítőik tervei és szándéka szerint olyan szótárelemek készültek, melyeket a bibliográfiai univerzum kontextusán kívül is könnyedén, hatékonyan fel lehet majd használni.

A relációkon kívül az Art and Rare Materials projekt munkatársai néhány kötött szótárkódolási sémát is definiáltak, azaz a metaadatelemekhez tartozó értékek listáit, amelyek az alábbiak:

- arrangement;
- handwriting type;
- origin;
- status;
- typeface.

Az ARM fejlesztése kis mértékben bár, de hatással volt a BIBFRAME szókészlet egyes elemeinek specifikációjára is – erre jó példa a `bf:colorContent` elem, amely a leírt entitás-előfordulás színhasználatát tipizálja, lehetséges értékei például fekete-fehér, többszínű stb. Ennek használata a Library of Congress javaslata alapján a művek, illetve a megjelenési formák szintjén releváns.

Az LD4P mellett párhuzamosan futó Linked Data for Libraries Labs projekt keretein belül fejlesztett VitroLib katalogizálási felület használatával lehetségessé vált az ARM-szótárak működésének gyakorlati tesztelése, egy, a Cornell University-n üzembe helyezett VitroLib-implementáció segítségével. Három forrástípus, az állókép (`bf:StillImage`), a szöveg (`bf:Text`), valamint az objektum (`bf:Object`) alkalmazási profilját, azaz leíró űrlapját valósították meg a rendszerben. A szükséges relációk a BIBFRAME, illetve az ARM szótárból származtak, illetve megjelölték az egyes relációkkal használható értékszótárakat is, így például az `arm:hasStylePeriod` tulajdonság értékei a Getty Art & Architecture Thesaurusból származhatnak. A profilok segítségével előállítható gráfokban

előforduló értékeket a SHACL RDF-validációs séma (lásd a 6.7. fejezetet) segítségével szabályozták (O’Keefe et al., 2019).

9.3.3. *bibliotek-o*

A bibliotek-o szótárcsomag fejlesztésének kiindulópontja Rob Sanderson korábban ismertetett kritikája, illetve egy azzal párhuzamosan készült elemzés volt, melyet az LD4L projekt Ontológia Munkacsoportja készített el; ez az elemzés ugyancsak hiányosságokat, fejlesztendő területeket mutatott ki a BIBFRAME 1.0 változatával kapcsolatban (Kovari, Folsom & Younes, 2017). A jelentés megfogalmazásával párhuzamosan a szakemberek kidolgoztak egy ideiglenes szókészletet, amely az LD4L projektben lefektetett modellezési alapelveket kívánta átültetni a gyakorlatba. A projekt első fázisának végén a munkatársak arra számítottak, hogy a BIBFRAME 2.0 végre megszünteti a feltárt hiányosságokat, azonban ez nem történt meg; több modellezési kérdésben a két szótár építői eltérően gondolkodtak. A bibliotek-o elemkészlet tehát a fenti gondolkodásbeli eltérések formába öntése. A kidolgozók abban bíztak, hogy a bibliotek-o a bibliográfiai források jóval gazdagabb modellezési környezetét nyújtja, mint a BIBFRAME. Ez azonban azzal a kompromisszummal jár, hogy minden konverzió, amely a bibliotek-o szótárelemeit BIBFRAME-re cseréli, szemantikai veszteséget fog okozni a pontosan fedésben lévő szemantikával rendelkező kifejezések hiányában.

A fejlesztési irányelveket az alábbiakban összegezték:

- külső szótárakhoz való igazodás, továbbá azok elemeinek felhasználása az adatsere és az interoperabilitás elősegítése érdekében (például a `dcterms:subject` reláció használata a `bf:subject` helyett). Az egyes újr felhasznált relációkat szótáranként csoportosítva a <https://bibliotek-o.org/overview/overview.html#> oldalon lehet megtekinteni. Igen sok elemet átemeltek például az RDA szótárából, de jelen van a Lingvoj, a Dublin Core, a schema.org, sőt a CIDOC-CRM szókészlet néhány eleme is);
- olyan kifejezések definiálása, melyek kellőképpen széles szemantikával bírnak ahhoz, hogy külső adatforrások számára is használhatók legyenek;
- az OWL és az RDF elemkészletének (például értelmezési tartomány, értékészlet) használata olyan mértékben, hogy biztosítható legyen a relációk kifejezőképessége, amellyel azonban nem szabályozzák túl a szótárat, illetve a vele modellezendő adatokat;

- az objektumtulajdonságok, strukturált adatok és ellenőrzött szótárak előnyben részesítése a strukturálatlan karakterláncokkal, literálokkal szemben;
- a legegyszerűbb, „legáramvonalasabb” modell megalkotása, amellyel az adatok megfelelően, hűen modellezhetők;
- a kapcsolatok, attribútumok kifejezésének egyszerű megvalósítása;
- összetett helyett atomi szintű adatrepresentációs módszerek kidolgozása.

A bibliotek-o talán legizgalmasabb leleménye a különálló Activity osztály, amelyet a szótár úgy definiál, mint egy ágens valamilyen tevékenysége, közreműködése, amely hatással van egy forrás létezésére, vagy módosítja annak állapotát. A BIBFRAME-mel ellentétben, amely a közreműködői kapcsolatokat a Contribution osztály tagjaival használható `bf:role`, illetve `bf:agent` relációkkal írja le, a bibliotek-o egyetlen szuperosztályt definiál, s azon belül körülbelül 140 alosztályt határoz meg, melyek mindegyike egy-egy tevékenységi kört fed le. A tevékenységek a BIBFRAME-ben bevezetett három alapentitás közül bármelyikhez csatlakozhatnak, a bibliotek-o-ban definiált tulajdonság (`bib:hasActivity`) és inverze (`bib:isActivityOf`) segítségével. Ebben a modellben az ágensek a `foaf:Agent` és nem a `bf:Agent` osztály elemeiként írandók le, s ugyancsak egyetlen tulajdonságpár, a `bib:hasAgent`, illetve a `bib:isAgentOf` segítségével kapcsolódnak az egyes tevékenységekhez. A tevékenységek helyszínének leírásához ugyancsak egy külső szótár, a PROV-O egyik osztálydefinícióját, a `prov:Location`-t használják fel, mely egyszerre teszi lehetővé valós földrajzi helyek, továbbá olyan absztrakt helymeghatározások használatát, mint például egy számítógépes mappa, vagy egy dokumentum valamely sora, illetve oszlópa.

Szintén újrafelhasználással teszik leírhatóvá a bibliográfiai forrásokhoz kapcsolódó annotációkat. (A BIBFRAME 1.0 erre külön osztályt definiált, mely aztán a 2.0 verzióban már nem jelent meg). A bibliotek-o erre a célra a W3C által karbantartott Web Annotation Ontology-t használja, s minden forrásentításhoz kapcsolódó annotációt az `oa:Annotation` osztályába sorol, majd a forrásentítást, illetve az arról szóló annotációt az `oa:hasTarget` relációval kapcsolja össze. Érdeemes megjegyezni, hogy az Annotation Ontology ennek a tulajdonságnak nem tartalmazza az inverzét, azt a bibliotek-o-n belül kellett definiálni, ez lett a `bib:isTargetOf`.

A BIBFRAME Work-ök, illetve Instance-ok osztályán belül a bibliotek-o számos további osztályt definiál az egyes bibliográfiai forrástípusok leírásához. A Work-ön belül a

BIBFRAME 2.0 változatában jelenleg 11 alosztály található, amelyhez a biblioték-ot továbbiakat is társít, s ugyanez a helyzet az Instance-ek osztályával, amelyen belül jelenleg öt alosztály szerepel. A kiegészítéseket mindkét esetben az RDA-ban definiált tartalomtípusok, médiumtípusok, illetve hordozótípusok alapján készítették el.

9.3.4. *Performed Music Ontology*

A Performed Music Ontology (PMO), azaz az előadott zene felvételét tartalmazó bibliográfiai források leírásához használható szókészlet is a BIBFRAME kiegészítéseként jött létre a Stanford University, a Library of Congress, a Music Library Association, az Association for Recorded Sound Collections, és az amerikai közös katalogizálási program (PCC) munkatársainak együttműködésével. A PMO a BIBFRAME Esemény (Event) osztályát tagolja az alábbi alosztályok bevezetésével:

- pmo:Audition
- pmo:Ceremony
- pmo:Concert
 - pmo:BenefitConcert
- pmo:CommandPerformance
- pmo:ConcertSeries
- pmo:ConcertTour
- pmo:Performance
- pmo:FirstPerformance
- pmo:LivePerformance
- pmo:OpenMicPerformance
- pmo:MasterClass
- pmo:RecordingSession
- pmo:Rehearsal

továbbá több BIBFRAME-tulajdonságot speciális, zenei forrásokra vonatkoztatható altulajdonságokkal bővít. Példának okáért egy pmo:Performance osztályba sorolt esemény és az esemény hangfelvételét tartalmazó bibliográfiai forrás (bf:Audio) között a pmo:hasRecording – pmo:recordingOf hoz létre kapcsolatot.

A modellezés során meg kellett fontolni továbbá, hogy az események alkalmával zeneművek előadására kerül sor, amelyek BIBFRAME-művekként jelölhetők a gráfszerkezetben. Ebben az esetben az adott esemény (vegyük példának ismét a Performance típusú eseményt) és az ott elhangzó zeneművek (bf:Work) közötti kapcsolatot a

`pmo:hasPerformance` – `pmo:performanceOf` relációpárossal biztosítható. A BIBFRAME-mű kifejezési formája a `bf:Audio` osztály egy konkrét példányaként szerepel a modellben.

A Performed Music Ontology osztályai, illetve tulajdonságai, illetve a PMO használatához szükséges szótárkódolási sémák elemei (lejátszási sebesség, kódolási formátum, hangnem stb.) fontos részei azoknak az alkalmazásprofiloknak, amelyeket a később ismertetendő Sinopia, illetve BIBFRAME Editor szerkesztőeszközök használatához fejlesztettek (Lorimer, 2019).

9.3.5. A *bf:lc*, a Library of Congress speciális relációkészlete

A Library of Congress kísérleti céllal definiálta kiegészítő szótárát 2017 februárjában. A *bf:lc*-szótár összesen 24 osztályt és 38 relációt tartalmaz. Ez utóbbiak közül érdemes megemlíteni a különféle *MarcKey*, illetve *MatchKey*-tulajdonságokat, melyek az egyes *authority* mezők értékét tárolják el, amennyiben a BIBFRAME-gráfot MARC-rekordokból konvertálták.

A 100-as MARC-mező adattartalmából például a következő gráfrészlet generálódik a konverziós folyamat során:

```
<http://bibframe.example.org/5226#Agent100-14> a bf:Agent,  
    bf:Person ;  
    rdfs:label "Flanagan, Terry." ;  
    bflc:name00MarcKey "1001 $aFlanagan, Terry." ;  
    bflc:name00MatchKey "Flanagan, Terry." ;
```

a 700-as adatmező tartalma pedig az alábbi módon képződik le:

```
<http://bibframe.example.org/5226#Agent700-25> a bf:Agent,  
    bf:Person ;  
    rdfs:label "Schulz, Charles M. (Charles Monroe), 1922-2000." ;  
    bflc:name00MarcKey "7001 $aSchulz, Charles M.$q(Charles Monroe),$d1922-2000." ;  
    bflc:name00MatchKey "Schulz, Charles M. (Charles Monroe), 1922-2000." .
```

Látható, hogy a különféle *MarcKey*-relációk, amelyek a *name00*, *name10*, *name11*, továbbá *title00*, *title10*, *title11*, *title30*, *title40* prefixeket vehetik fel, utalva az eredeti rekord adatmezőinek hívójeleire, az adott MARC-mező teljes tartalmát megőrzik, a hívőjelet, az indikátorértéke(ke)t, az almezőjele(ke)t, és az almezőkbe bevitt karakterláncokat is. A *MatchKey*-relációk, melyek szintén a fent felsorolt prefixekkel együtt állhatnak, csupán az almezők értékeit tartalmazzák, hívőjelet, indikátorértéket, valamint almezőjelet nem.

Az autorizált névalakoknak, főleg azok teljes, MARC-rekordokból származó formájának szerepe a gráfban vélhetően az ellentétes irányú (BIBFRAME-ről MARC-ra)

konverzió segítésére korlátozódik, hiszen az entitás-előfordulások minden kétséget kizáró azonosítása a gráfon belül (s tágabb értelemben véve az egész szemantikus világhálón) az URI-kkal történik, s a gráfon belüli hivatkozások is ezek segítségével valósulnak meg.

Érdemes szólni róla, hogy a BIBFRAME alapszótárában nem volt olyan reláció, amely alkalmas lett volna a MARC-rekord rekordfejének 17. karakterhelyén található, a leírás jellegéről szóló közlés kifejezésére. A Library of Congress kiegészítésében e hiány pótlására definiálták a `bflc:encodingLevel` tulajdonságot, amelynek értéke kötött szótárból választható a MARC 21 szabvány szerint (minden érték URI-ja hozzáférhető az `id.loc.gov` oldalról). A tulajdonság értékei kizárólag a `bflc:EncodingLevel` osztályból származhatnak, ezért minden gráfban megtalálható a választott URI-val jelölt egyedi entitás-előfordulás osztályba sorolását elvégző kódrészlet, valamint a megjelenítést támogató `rdfs:label` címkereláció:

```
<http://id.loc.gov/vocabulary/menclvl/f> a bflc:EncodingLevel ;  
  rdfs:label "full" .
```

Bizonyos közreműködői funkciók (és a hozzájuk tartozó ágensek) elsődlegesnek címkézhetők, így a gráfban is megkülönböztethetővé válik a 1XX, továbbá a 7XX adatmezőkből származó felelősségi közlés. A `PrimaryContribution` osztályt (a BIBFRAME `Contribution` osztályának alosztályát), valamint az ehhez rendelt három tulajdonságot (`primaryContributorName00MatchKey`, `primaryContributorName10MatchKey`, `primaryContributorName11MatchKey`) ugyancsak a `bflc` szótár vezette be. Ilyenkor az ágens leírása kiegészül egy újabb triplettel:

```
bflc:primaryContributorName00MatchKey "Flanagan, Terry." .
```

valamint a konkrét közreműködői funkciót mint entitás-előfordulást be kell sorolni a `PrimaryContribution` osztályba:

```
bf:contribution [ a bflc:PrimaryContribution,  
                 bf:Contribution ;  
                 bf:agent <http://bibframe.example.org/981245#Agent100-12> ;  
                 bf:role <http://id.loc.gov/vocabulary/relators/ctb> ] ;
```

Lényegében szükségtelenül fejezik ki az adatszerkezetben, hogy az adott entitás-előfordulás egyszerre tagja a `PrimaryContribution`, továbbá a `Contribution` osztálynak. A `bflc` szótárban ugyanis előbbi az utóbbi alosztályaként szerepel, így a hierarchikus reláció akkor is kikövetkeztethető a gépi feldolgozó folyamat által, ha az az egyes gráfokban expliciten nem fejeződik ki.

9.4. A BIBFRAME-kompatibilis gyűjteménykezelő rendszer

A 2017-ben, Frankfurtban rendezett első Európai BIBFRAME Workshopon a közösség vizsgálni kezdte az új bibliográfiai keretrendszer és az integrált könyvtári rendszerek kompatibilitási paramétereit. A személyes találkozót követően az európai implementori közösség vezetői Tiziana Possematót, az @Cult munkatársát kérték fel arra, készítsen vázlatos specifikációt a BIBFRAME-nek való megfelelés követelményeiről. A dokumentum (*BIBFRAME Expectations for ILS tenders*, 2018) az alábbi szinteket különbözteti meg:

- a rendszer képes jól formázott, a BIBFRAME-konverzióra megfelelően előkészített MARC-rekordok előállítására. Ebbe beleértendő az RDA szabályainak követése, a rekordok URI-kkal történő gazdagítása, illetve a bibliográfiai és authority rekordok között kialakított gazdag kapcsolati háló is. Ezen a szinten a rendszer még teljes egészében a MARC-ra épül;
- a szállító biztosít valamilyen konverziós segédeszközt, amely MARC-ból, MODS-ból, illetve egyéb formátumokból BIBFRAME-gráfot képes előállítani. Ezen a szinten a konverziós segédeszköz még nem a rendszer részeként, hanem autonóm, kiegészítő szoftverként hozzáférhető. A gráfok tárolására szolgáló triplestore a rendszer alapját biztosító relációs adatbázishoz képest másodlagos (slave) szerepet tölt be;
- a szállító az integrált könyvtári rendszerhez csatlakozó konverziós segédeszközt biztosít; ebben az esetben a tömeges adatkonverziót követően a BIBFRAME-adathalmaz folyamatos frissítésére nyílik lehetőség, követve a rendszerben elvégzett módosításokat;
- az integrált gyűjteménykezelő rendszer BIBFRAME-re alapuló szerkesztőfelületet biztosít, amelyen az adatbevitel elvégezhető. A tripletekben tárolt adatok egy API-n keresztül konvertálódnak a relációs adatbázisok számára megfelelő formátumba, így az IKR egyéb programmoduljai számára is felhasználhatóvá válnak, ezzel támogatva a hagyományos funkciók működését, például a kölcsönzését;
- a rendszernek saját BIBFRAME-re alapuló adatbeviteli interfésze van, az adatbevitel tehát entitásorientált. Ebben az esetben a triplestore átveszi az elsődleges (master) szerepet a relációs adatbázistól, de minden művelet mindkét adatbázisban végbemegy;

- a rendszer ökoszisztémája teljes egészében a linked data-technológiára épül, így a feldolgozás és az adattárolás RDF-alapokon történik, nincsen járulékos relációs adatbázis;
- a rendszer API-n keresztül kapcsolatot tud fenntartani a hagyományos funkcionalitású rendszerekkel, hogy azokból adatokat legyen képes átemelni;
- a szállító a BIBFRAME előnyeit egyértelműen prezentáló, entitás-kapcsolat modellre épülő keresőfelületet biztosít.

9.5. Átjárhatóság más metaadat-rendszerek felé

9.5.1. Az Europeana és az EDM adatmodell

Az európai digitális könyvtár, az Europeana, helyesebben az azt üzemeltető Europeana Foundation 2008-ban jött létre, központja a holland nemzeti könyvtárban (Koninklijke Bibliotheek, Hága) található. Feladatának tekinti a sokszínű és gazdag európai kulturális örökség, az értékes közgyűjteményekben felhalmozott kulturális kincsek eljuttatását a szélesebb közönséghez. Munkáját többszintű központi irányítással végzi: a legfőbb döntéshozói az Europeana Governing Board és a Management Board nevű testületekbe tömörülnek, mellettük azonban más szervezetek is működnek, köztük a stratégiai és finansziális célok megvalósításáért dolgozó Members Council, valamint a kulturális örökségi területen dolgozó, kreatív és technológiai feladatokkal foglalkozó szakembereket (az Europeana tagintézményeinek képviselőit) tömörítő Network Association.

Ez a hatalmas, határokon átívelő gyűjtemény napjainkban körülbelül ötvenmillió rekordot tartalmaz, amelyek összesen mintegy 3500 különféle adatszolgáltatótól (adatpartnertől, Data Partner), az ún. GLAM-szektor tagjaitól származnak. Az összegyűlt hatalmas adatmennyiséget az Europeana az ún. EDM-formátumban tárolja, amely élesen elkülöníti egymástól a valóságban létező kulturális örökségi objektumot, illetve annak valamilyen digitális reprezentációját, például fényképét, szkennelésen és szövegfelismerésen átesett változatát stb. Egy különleges osztály, az `ore:Aggregation` feladata, hogy ezeket az önálló entitásokat egymással összerendelje. Egy ilyen összerendelés rámutat az adatszolgáltatóra (`edm:dataProvider`), a szolgáltatótól érkező webes reprezentáció(k)ra (`edm:webResource`), valamint leírja, hogy az adott reprezentáció(k) melyik kulturális örökségi objektumot mutatja(k) be (`edm:providedCHO`). A CHO-hoz természetesen hozzárendelhetők annak saját leíró metaadatai, amelyek szemantikus

adatformátumokban, akár a Dublin Core-szótár relációinak segítségével is kifejezhetők. Létezik azonban egy további szint, amelynek megértéséhez figyelembe kell venni, hogy az egyes kulturális örökségi objektumokat a világon több néven is ismerik. Leonardo da Vinci híres festményét például *Mona Lisaként*, *Giocondaként*, továbbá *Lisa Gherardini portréjaként* is szokás emlegetni. Ezért az Europeana adatmodellje bevezette a *proxy* (*helyettesítő*) osztályát, amely az adatszolgáltatók (dataProviderek) által megvalósított szolgáltatási eseményekben (aggregation) más és más módon megnevezett, illetve leírt, de valójában egyetlen kulturális örökségi objektum (providedCHO) helyett álló szurrogátumokat foglalja magába.

Intenzív kutatások folynak abban az irányban, hogy az Europeana Data Modelt, valamint a BIBFRAME-et átjárhatóvá, azaz egymásba konvertálhatóvá tegyék, ezért érdemes ezekről a kísérletekről néhány szót ejteni.

Az EDM, valamint a könyvtári dokumentumok metaadat-formátumainak illesztési alapelveit egy 2012-ben elkészült dokumentum tartalmazza (Angjeli et al., 2012). Ebben szerepel egy igen érdekes fogalom, a kiadásé (edition), amely a kézbevert konkrét dokumentumtól függetlenül létező absztrakt entitás, és egyesíti magában az FRBR-ben bemutatott Mű, Kifejezési forma és Megjelenési forma entitásokat. Az illesztés kiindulópontja az, hogy könyvtári dokumentumok esetében a kiadás a kulturális örökségi objektum, és ennek a más-más intézményekben található példányai a digitális reprezentációk. Ebből adódik az első megoldás a BIBFRAME szótára segítségével megadott leíró metaadatok EDM-re való átültetésére. Sofia Zapounidou, Michalis Sfakakis és Christos Papatheodorou, a Ión Egyetem munkatársai több írásukban is foglalkoztak a BIBFRAME, illetve az EDM adatmodelljének megfeleltetésével, hogy ezáltal átjárhatóvá tegyék a különféle gyűjteményi ágazatok által biztosított adathalmazokat. 2014-ben írt konferenciáikkukban a BIBFRAME első változatával dolgoztak (Zapounidou, Sfakakis & Papatheodorou, 2014b): ebben a verzióban a Work osztály magában foglalja az FRBR-i értelemben vett mű és kifejezési forma valamennyi tulajdonságát, az Instance osztályhoz pedig a megjelenési forma tulajdonságai tartoznak. Az EDM-értelmezés szerint e két (szigorúbban véve három) entitás adatai mind a kulturális örökségi objektum (providedCHO) egyes tulajdonságainak feleltethetők meg, a szolgáltatási eseményben megjelenő digitális reprezentációk – webes források – adatai pedig példányadatokként jelenhetnek meg, amelyeket a BIBFRAME 1.0-ban a `bf:HeldMaterial` osztály segítségével kellett formalizálni, s ezeket a `bf:hasAnnotation` reláció felhasználásával az Instance

osztályhoz kötni. A `bf:HeldMaterial` osztály jellemzéséhez felhasználható volt a `bf:electronicLocator` reláció, mellyel az adott digitális reprezentáció URI-ját lehetett a gráfban megjeleníteni.

A második megoldás számol azzal, hogy egy bibliográfiai forrás leírásával keletkező metaadatművek nem feltétlenül egységesek, hanem egymástól szerkezetükben, tartalmukban, mélységükben jelentősen eltérhetnek a világ gyűjteményeiben – azaz az átjárhatóság kialakítása során tekintetbe veszi az EDM-proxy-kat. Ebben az esetben a BIBFRAME-ben Workként és Instance-ként leírt adatok nem a tényleges kulturális örökségi objektum, hanem egy-egy helyettesítő szurrogátum adatai lesznek, amelyek a fentebb leírt módon kapcsolódnak az Europeanában nyilvántartott kulturális örökségi objektumokhoz. Az Annotation-ből származó jellemzők változatlanul az EDM-értelmezésű digitális reprezentációk adatai lesznek.

A harmadik lehetőség abban nyújt többet, hogy segítségével jobban elkülöníthetőek maradnak a BIBFRAME Work-ök adatai. Ez a megoldás, amelynek részleteit az EDM-FRBRoo átjárás lehetőségeit bemutató munkacsoporti jelentés ismerteti (Doerr et al., é.n.), azon alapszik, hogy a BIBFRAME-művek adatai nem az EDM szerint definiált kulturális örökségi objektum adataiként tárolódnak, hanem egy általános, `InformationResource` osztály használatával, amely típusmegjelöléssel is ellátható; ez a típusmegjelölés pedig a BIBFRAME-művek típusa is lehet. E felosztás szerint az EDM-CHO adatai a BIBFRAME-Instance adataival egyeznek meg; az `InformationResource` pedig magában foglalja (incorporates) a kulturális örökségi objektumot.

A negyedik megoldás szerint, mely lényegében a harmadik és második kombinációja, a BIBFRAME-mű `edm:InformationResource`-ként jelenik meg, az Instance-adatok pedig nem a kulturális objektumhoz, hanem azok szurrogátumaihoz, a proxy-khoz csatlakoznak. A szolgáltatási esemény és a helyettesítő páronként meghatározza egymást.

A görög kutatók a témáról írt másik publikációja (Zapounidou, Sfakakis & Papatheodorou, 2014a) speciális megfeleltetési esetek rövid leírását adja, olyanokét, mint:

- ha az egyik BIBFRAME-mű egy másik BIBFRAME-mű része, az EDM-gráfban ez a Dublin Core-szótár `dc:hasPart` relációjával képezhető le, két kulturális örökségi

objektum rész-egész viszonyaként: *ProvidedCHO* – `dcterms:hasPart` – *ProvidedCHO*;

- ha az egyik BIBFRAME-mű egy másik BIBFRAME-mű fordításaként jött létre, akkor ez ugyancsak leképezhető az EDM-ben, a `bf:hasTranslation` reláció azonban elveszíti speciális szemantikáját, mivel azt a `dcterms:hasVersion` kapcsolatjelölő helyettesíti: *ProvidedCHO* – `dcterms:hasVersion` – *ProvidedCHO*;
- ha két BIBFRAME-megjelenési forma között reprodukció típusú származékos kapcsolat áll fenn, az ugyancsak kifejezhető két objektum származékos viszonyaként, itt a `bf:reproduction` relációt az általánosabb jelentésű `edm:isDerivativeOf` helyettesíti: *ProvidedCHO* – `edm:isDerivativeOf` – *ProvidedCHO*.

9.5.2. Interoperabilitás az RDA adatmodelljének irányába

A BIBFRAME adatmodellje nem különíti el a Kifejezési forma entitást: annak jellemzőit a művek tulajdonságaival egybevonva teszi leírhatóvá. Ez jelentős eltérés az RDA forrásleírási szabályzat rendelkezéseit követő szótárban definiált entitásokhoz képest, és nem eléggé körültekintő konverzió esetén adatvesztést, illetve nem valós viszonyok létrejöttét eredményezheti. E probléma okait és hatásait ugyancsak Sofia Zapounidou, Michalis Sfakakis, valamint Christos Papatheodorou tanulmánya alapján lehet a legjobban bemutatni (Zapounidou, Sfakakis & Papatheodorou, 2019).

Az eltérés legplasztikusabban az ún. származékos kapcsolatok modellezésében jelentkezik. Barbara Tillett (2001) szerint a származékos kapcsolatok a bibliográfiai kapcsolatok egy sajátos típusát jelentik, amely családba az azonossági (equivalent), a leíró (descriptive), a rész-egész (whole-part), a járulékos (accompanying), a folytatólagos (sequential), illetve a közös jellemzőkre alapuló kapcsolat (shared characteristic) is tartozik. Tillett a származékos kapcsolatokat is két csoportra osztja annak alapján, hogy az eredeti művön végzett bármilyen változtatás, módosítás új művet hozott-e létre, vagy csupán az eredeti mű új kifejezési formáját.⁴² Az első csoportba tartoznak egy mű (komolyabb módosítás nélküli) többedik kiadásai, a fordítások, az illusztrált kiadások, a rövidített kiadások, a felújított kiadások (expurgated edition), míg a második csoportot a

⁴² A Barbara Tillett által megkülönböztetett bibliográfiai kapcsolatokról lásd a 4. ábrát a 4.2. fejezetben.

szabadfordítások, az összefoglalók, absztraktok, kivonatok, paródiák, műfajban bekövetkezett változtatások (forgatókönyvek, librettók, dramatizált, regényesített változatok stb.) alkotják. Az eredeti mű és annak kisebb-nagyobb módosítással létrejött származékai Smiraglia és Leazer szerint ún. *bibliográfiai családot* alkotnak; a család tagjai tehát minden esetben egyetlen közös ősrre vezethetők vissza, és nagyjából azonos eszmei-szellemi tartalommal bírnak (Smiraglia & Leazer, 1999).

Az RDA szabályzatban rögzítettek szerint származékos viszony áll fenn:

- egy mű két kifejezési formája között (például egy eredeti nyelvű és egy más nyelvre fordított változat);
- két mű között (ez a speciális eset az RDA kidolgozói szerint akkor áll fenn, amikor nem határozható meg pontosan, hogy a származék előállításakor az eredeti mű melyik kifejezési formáját használták fel);
- két mű egy-egy meghatározott kifejezési formája között.

A BIBFRAME szótár esetében származékos kapcsolat csak művek között létezhet, abban az esetben, ha ismert mind az eredeti, mind pedig a származékos kifejezési forma is. A két entitás-előfordulás az általános `bf:relatedTo` reláción kívül rögzíthető a `bf:hasDerivative / bf:derivativeOf` viszonypár alkalmazásával, illetve ezen belül is találni konkrét származékviszony-típusokat leíró relációkat, azonban valamivel kevesebbet, mint az RDA-szótár esetében. A két modell közötti átjárás kiépítésekor tehát számolni kell azzal, hogy az RDA-ban modellezhető kapcsolattípusok egy részének specifikussága a gráf BIBFRAME-konverziója során elvész. Így például a rövidített változatra utaló viszonyjelölőt, vagy éppen a dramatizált változatra mutató relációt csak általánosabb BIBFRAME-relációval van mód helyettesíteni (Zapounidou, Sfakakis & Papatheodorou, 2019).

Az RDA-BIBFRAME konverziós folyamat során egyetlen RDA-mű, amelynek n számú kifejezési formája van, a BIBFRAME-gráfban n számú mű leírásává képződik le, amelyeket a `bf:hasExpression` relációval kell összekötni a bibliográfiai kapcsolat megőrzése érdekében. Ha azonban kettő vagy több RDA-művet kell BIBFRAME-re konvertálni, amelyek egymással tetszőleges származékos kapcsolatban állnak – és az egyiknek n , a másinak m számú kifejezési formája van –, akkor az RDA-kifejezési formákból képződő BIBFRAME-művek *mindegyike* között felépül a származékos kapcsolat. Ez pedig nyilvánvalóan helytelen, hiszen a származék az eredeti mű mindössze egyetlen kifejezési

formájának felhasználásával születik. A megoldás a görög kutatók szerint a következő: csak olyan származékos kapcsolat menthető át BIBFRAME-be művek közötti kapcsolatként, amely az RDA-gráfban kifejezési formák között állt fenn. Azokat a viszonyokat pedig, amelyek az RDA-gráfban két mű között húzódtak, a BIBFRAME-gráfban létrejövő téves kapcsolatok elkerülése érdekében figyelmen kívül kell hagyni.

A kutatók az elméleti állítások igazolására összesen 235 MARC-rekordból álló halmazt válogattak össze gazdag kiadástörténettel rendelkező művek (mint a Száz év magány, a Bűn és bűnhődés, az Iliász vagy éppen a Faust) megjelenési formáit reprezentáló bibliográfiai rekordokból –. A MARC-rekordokból képzett MARC/XML-t a kísérlet első fázisában RDA-gráffá konvertálták, s ezt a halmazt RDA Goldnak nevezték el. A második fázisban már ezzel dolgoztak tovább: az összehasonlíthatóság érdekében két BIBFRAME-gráfot is készítettek, eltérő módszertannal: az egyikben figyelembe vették az RDA-művek, illetve kifejezési formák között ábrázolt kapcsolatokat, a másikban kizárólag a kifejezési formák kapcsolatait transzformálták, tekintettel az elméleti hipotézisben megfogalmazottakra.

A MARC/XML-ből RDA-gráfra történő átalakítás során 48 műről és 195 kifejezési formáról keletkeztek RDF-tripletek, melyekhez összesen 236 megjelenési forma metaadatai csatlakoztak. A kutatók előzetes vélekedése igazolódott: abban a gráfban, melyet csupán a kifejezési formák közötti kapcsolatok alapján készítettek, az RDA-kifejezési formákkal megegyező számú mű-leírás készült, míg a megjelenési formák száma megegyezett. Az RDA-művek közötti kapcsolatok tekintetbe vétele azonban zajossá tette a másik BIBFRAME-gráfot, s jóval több kapcsolat keletkezett, mint amennyi valóban fennáll az egyes entitások között.

A kutatás adatai – a kiinduló MARC/XML-állomány, továbbá a BIBFRAME- és RDA-gráfok – szabadon letölthetők és tanulmányozhatók a http://libdata.tab.ionio.gr/models/si-mapping/si_project.html webcímen. Érdemes megemlíteni, hogy a kísérletet a kutatók megismételték, ezúttal ellenkező irányban: BIBFRAME-gráfokat igyekeztek RDA-gráfokká alakítani. A munkafolyamat során keletkező dokumentáció a projekt weblapján ugyancsak szabadon megtekinthető, az eredményeket és tanulságokat részletező publikáció a *Cataloging & Classification Quarterly* 2020. évi 7. számában jelent meg (Sfakakis, Zapounidou & Papatheodorou, 2020).

9.5.3. Átjárás a *schema.org* irányába

Osma Suominen a 2018. évben Firenzében rendezett európai BIBFRAME-workshopon számolt be egy projektről, amelynek során BIBFRAME-gráfokat konvertáltak *schema.org* gráfokká, majd ezeket publikálták (Suominen, 2018). A körülbelül egymillió MARC-rekordot tartalmazó, a finn nemzeti bibliográfia (Fennica) adatait leíró kiinduló állományból a Library of Congress által készített, hivatalos BIBFRAME-konverter (*marc2bibframe2*) segítségével készült az eredeti gráf, amelyet SPARQL CONSTRUCT lekérdezés segítségével alakítottak át a kimeneti szerkezetre.⁴³ A BIBFRAME-ben Work-ként, illetve Instance-ként leírt entitások a *schema.org* adatmodelljében egyaránt Creative Work-nek minősülnek, s egy BIBFRAME-Work, illetve BIBFRAME-Instance közötti kapcsolat a *schema.org* szótárában található két tulajdonsággal, a *schema:exampleOfWork*-kel, illetve inverzével, a *schema:WorkExample*-lel írható le; előbbi a *bf:Instance* felől a *bf:Work* felé mutat, míg utóbbi a *bf:Work* felől irányul a *bf:Instance*-re.

Az átalakító lekérdezés után az adathalmaz egy gazdagító CONSTRUCT-on is keresztülment, amelynek során a gráf külső adatforrások azonosítóival bővült a tárgyi authority adatok, testületi nevek és bizonyos RDA-szótárkódolási sémák értékeinek esetében. A kész adathalmazt végül N-Triples és HDT-formátumban tették hozzáférhetővé (Suominen, 2017).

9.6. Konverzió a MARC és a BIBFRAME között – segédletek és eszközök

9.6.1. Összehasonlító segédeszköz

A Library of Congress munkatársai a konverziós specifikációkra építve előállítottak egy olyan segédeszközt, amely egymás mellé helyezi az eredeti MARC-rekordot, és mellette megmutatja annak BIBFRAME-gráffá konvertált megfelelőjét. Ezek összevetése hasznos tanulságokkal szolgál a szótár mélyszerkezetével, működésével, fogalomértelmezésével kapcsolatban. Az eszköz használata nagyon egyszerű: a Library of Congress rendszeréből származó rekordazonosítót (Bib ID) vagy ellenőrzőszámot (LCCN) kell megadni a megtekintendő MARC-rekord azonosításához, továbbá beállítható, hogy a BIBFRAME-gráf megjelenítése Turtle vagy RDF/XML-formátumban történjen-e. A Search gombra kattintás után a képernyő bal oldalán a MARC-rekord látható, míg jobb oldalon a kívánt formátumú

⁴³ A lekérdezés hozzáférhető a következő címen: <https://github.com/NatLibFi/bib-rdf-pipeline/blob/master/sparql/bf-to-schema.rq> [letöltés: 2020.12.22.]

gráf szerepel. Mentési lehetőség nincsen, azonban a kívánt részletek, vagy akár az egész rekord/gráf a vágólapra másolva, szöveges változatban eltárolható.

9.6.2. Aktív konverzió

Aktív konverzióknak azt a folyamatot nevezzük, amikor a kiinduló MARC/XML-ből úgy állítanak elő BIBFRAME-alapú RDF/XML-t, hogy a konverziós stíluslap átalakítási szabályainak érvényesítése mellett adatgazdagítás is történik, azaz bizonyos relációkkal felépülő tripletek tárgyaiként előforduló literálokra alapozott keresés fut le a specifikációban meghatározott helyeken, majd az ott megtalált URI-k bekerülnek a konverzió eredményeként előálló gráfba. Az alábbi példában:

```
<backend syntax="xml" name="rdf-lookup">
  <xslt stylesheet="xsl/marc2bibframe2.xsl"/>
  <rdf-lookup debug="1">
    <namespace prefix="bf" href="http://id.loc.gov/ontologies/bibframe/" />
    <namespace prefix="bflc" href="http://id.loc.gov/ontologies/bflc/" />
    <lookup xpath="//bf:contribution/bf:Contribution/bf:agent/bf:Agent">
      <key field="bflc:name00MatchKey"/>
      <key field="bflc:name10MatchKey"/>
      <key field="bflc:name11MatchKey"/>
      <server url="http://id.loc.gov/authorities/names/label/%s" method="HEAD"/>
    </lookup>
  </rdf-lookup>
</backend>
```

az XPATH-kifejezés által meghatározott útvonalon – jelen esetben a közreműködők leírásánál – az automatikus folyamat megkeresi mindazokat a tripleteket, amelyek a bflc kiegészítő szótár három relációja, a name00MatchKey, a name10MatchKey, illetve a name11MatchKey segítségével készültek a MARC 100/700-as, 110/710-es, 111/711-es adatmezői értékeinek felhasználásával, majd az ott talált karakterláncot a server URL-ben meghatározott elérési útvonalba behelyettesítve (a \$s helyére) keresést végez. Találat esetén a fellelt URI hozzákapcsolódik a BIBFRAME-gráfhoz.

Az aktív konverziós állomány⁴⁴ tanúsága alapján automatikus adatgazdagítás megy végbe az alábbi pontokon:

- valamennyi bf:Agent esetében, amennyiben rendelkezésre áll a MARC-rekordból származtatott MatchKey-érték. Az adatgazdagításhoz használt URI ebben az esetben: <http://id.loc.gov/authorities/names/label/>

⁴⁴ Hozzáférhető a következő címen: <https://github.com/lcnetdev/marc2bibframe2/blob/master/record-conv.xml> [letöltés: 2020.12.23.]

- tárgyszavak esetében, ha a triplet a `bf:Temporal`, `bf:Topic`, `bf:Place`, vagy `bf:GenreForm` tulajdonságok segítségével készült. Ekkor az alkalmazott keresési útvonal: <http://id.loc.gov/authorities/subjects/label/>
- a közreműködői funkciókat leíró tripletek esetében, ha azok a `bf:Role`, vagy `bflc:Relation` property-k segítségével állnak elő. Keresési útvonala: <http://id.loc.gov/vocabulary/relators/label/>.

A konverzió az IndexData által fejlesztett YAZ szoftvereszköz segítségével futtatható.

9.6.3. MARCEdit

Vélhetően az egyik legkényelmesebb, leghatékonyabban és leggyorsabban használható konverziós segédeszköz a Terry Reese által fejlesztett MARCEdit nevű szoftver (letölthető a <http://marcedit.reeset.net> címen) beépített BIBFRAME-átalakítója. A MARCEdit egy, az Oregon State University-n tervbe vett, nagyléptékű adatbázis-tisztítási munka támogatására született meg a 2000-es évek első felében, és azóta a világ számos országában több tízezer használja ezt az igen sokoldalú segédeszközt. A szoftver fejlesztője nyomon követi a könyvtárinformatika legfontosabb trendjeit, ezért az eszköz nem csupán a hagyományos feladatokat (MARC-rekordok kezelése, szerkesztése, csoportos feldolgozása, tisztítása, külső keresések és letöltési lehetőség átviteli protokollok használatával stb.) támogatja, hanem különböző konverziós feladatokra, adatgazdagításra, illetve akár SPARQL-végpontok távoli böngészésére is használható. A BIBFRAME Testbed nevű funkció használata lehetővé teszi a szoftver segítségével megnyitott állományok – legyenek azok akár bináris MARC, vagy MARC/XML-formátumban – BIBFRAME-gráffá konvertálását, néhány egyszerű paraméter megadásával.

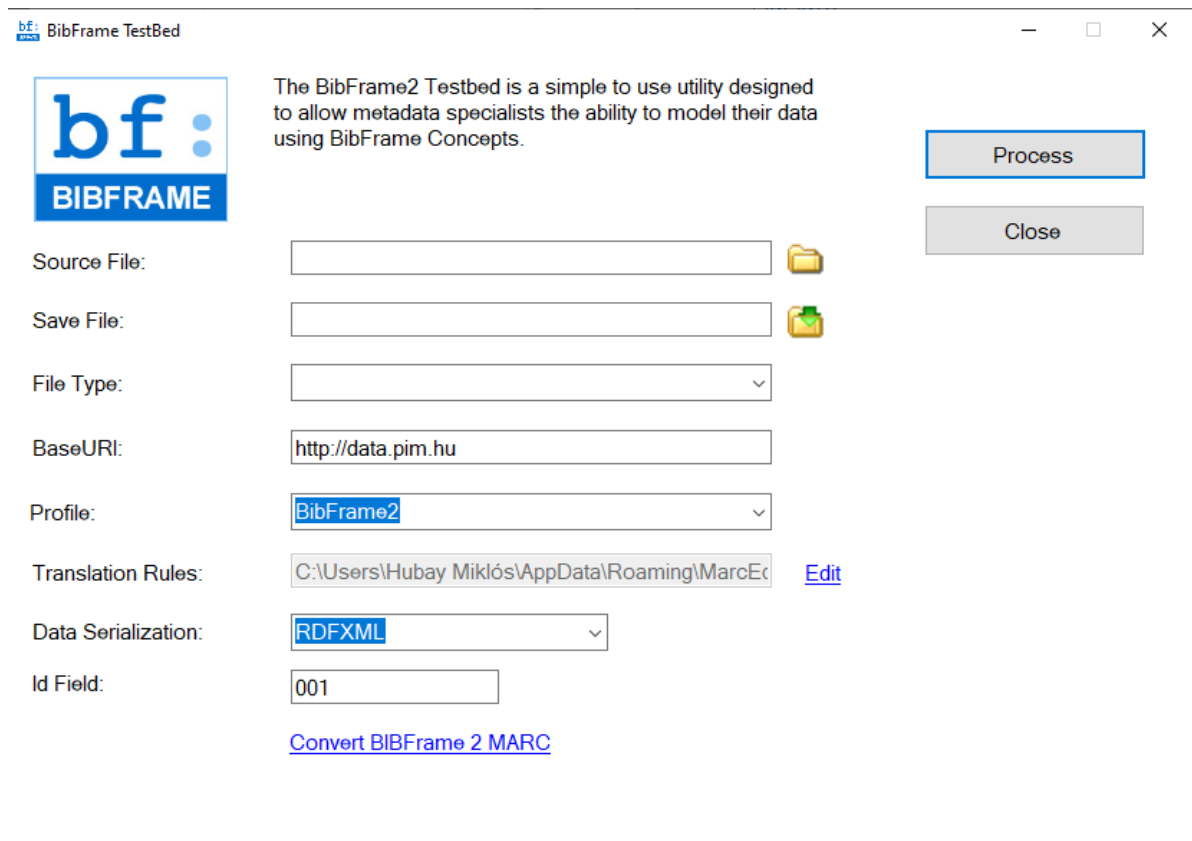
A konfigurációs űrlap mezői az alábbiak:

- *Source File*: az adatokat tartalmazó, konvertálandó forrásállomány elérési útvonala. A file formátuma *.mrc, illetve *.xml lehet;
- *Save File*: a konverzió során elkészülő állomány elérési útvonala;
- *File Type*: a bemeneti állomány típusa. A támogatott formátumok, illetve sémák a következők: (bináris) MARC, MARC/XML, EAD, MODS, továbbá Onix és FGDC. Ez utóbbi betűszó a Federal Geographic Data Committee (szövetségi földrajziadat-bizottság) által kidolgozott, speciális formátumot jelenti, amelynek MARC 21-re,

továbbá Dublin Core-ra való mappelésének részleteit Adam Chandler és szerzőtársai írták le (Chandler, Foley & Hafez, 2000);

- *BaseURI*: a gráfban használni kívánt, saját bázis-URI (tehát az azonosítók állandó része). Megalkotásakor érdemes figyelembe venni az ismert URI-névkonvenciókat (például W3C, 2008);
- *Profile*: a BIBFRAME-szótár alkalmazandó verziójának kiválasztása. Lehetséges értékei: BibFrame1, illetve BibFrame2. Az adatok minél szélesebb körű felhasználhatósága érdekében a második szótárverzió használatát érdemes előnyben részesíteni;
- *Translation Rules*: a részletes konverziós szabályokat leíró stíluslapok elérési útvonala. A szoftver a BIBFRAME második verziójára történő konverzióhoz a Library of Congress által készített XSLT-ket használja, melyek a MARCEdit telepítéskor automatikusan a felhasználó számítógépére kerülnek. A BIBFRAME első verziójára történő konverziót vezérlő, deklarációkat, illetve FLWOR-mondatokat tartalmazó XQuery-állományok eléréséhez hálózati hozzáférés szükséges, mivel azok a MARCEdit szoftver honlapjáról érhetők el. A file-ok elérési útját általában nem szükséges szerkeszteni, de amennyiben mégis, lehetőség van a módosításra;
- *Data Serialization*: a konverzió során alkalmazandó szerializációs szintaxis. Kétféle RDF/XML választható, ezen kívül N-Triples, illetve JSON, valamint exhibitJSON-formátum előállítására is van lehetőség;
- *Id Field*: lehetőség van annak megadására is, hogy a konverziós folyamat mely MARC-adatmezőben keresse a rekordazonosítókat, az ugyanis nem minden esetben található a 001-es hívójelnél.

A 2020 májusában bejelentett BIBFRAME-MARC-irányú konverziós specifikációk (erről bővebben a 9.6.4. fejezetben) nem sokkal később ugyancsak bekerültek a szoftver lehetőségei közé; így az már gráfok rekordokká alakítását is lehetővé teszi. Az opció a MARC segédeszközök között található, s beállítása jóval egyszerűbb, mint az ellentétes irányé. A bemeneti állománynak *.rdf vagy *.xml kiterjesztésűnek kell lennie.



17. ábra – A MARCEdit szoftver BIBFRAME-konfigurációt paraméterező ablaka
(képernyőkép, saját szerkesztés)

9.6.4. Konverzió BIBFRAME-ről MARC-ra

2020. május elsején Sally McCallum egy, a BIBFRAME-levelezőlistára küldött levelében jelentette be, hogy elkészültek a BIBFRAME-gráfok MARC-rekordokká történő átalakításához szükséges konverziós specifikációk, és a szükséges szoftvereszközök. Az első látásra talán logikátlanak tűnő lépés megtételére már több éve készült a Library of Congress: a fejlesztések a BIBFRAME tesztelésének második fázisában, a 2017. évtől indultak meg az intézményen belül (McCallum, 2018). Hogy a könyvtár ilyen munkákba fogott, arra az ad magyarázatot, hogy a Library of Congress tisztában volt azzal, hogy az előrelátható jövőben nem lehetséges a MARC kivonása a mindennapi gyakorlatból, hiszen a legtöbb könyvtárinformatikai szoftver nincs felkészülve a gráfok megfelelő kezelésére. S még ha el is jut egy intézmény odáig, hogy a katalogizáló programmodult és a teljes intézményi adatbázist az RDF-re és más, korszerű informatikai technológiákra alapozza, még mindig ott van a többi programmodul, például a kölcsönzés, melyek transzformációja napjainkban még nem megoldott. Ezek a részrendszerek továbbra is csak MARC-rekordokkal működtethetők (erről a 7.4. fejezetben is említést tettem). A harmadik fontos

indok a két adatmodell jelentős különbségében rejlik: a BIBFRAME és a MARC közelítéséhez, az adatok minél minőségibb, teljesebb körű megőrzéséhez szükséges, hogy a transzformáció mindkét irányban többször végigmenjen, s ezzel ellenőrizhető legyen a folyamat, illetve a kidolgozott konverziós specifikációk megfelelősége. Így fény derülhet bizonyos, előre nem látható problémákra, s lehetőség adódik ezek kiküszöbölésére.

A konverziós táblázatokat a Library of Congress-szal szerződéses viszonyban álló IndexData készítette el, továbbá rendelkezésre áll egy, a korábbiakban ismertetethez hasonló összehasonlító segédeszköz, amely a Bib ID vagy LCCN alapján megtalált BIBFRAME-gráfot MARC/XML-ként vagy hagyományos MARC-ban teszi megtekinthetővé.

9.7. BIBFRAME-re épülő katalogizáló felületek

9.7.1. A BIBFRAME Editor

A Library of Congress 2014 óta aktívan használja a BIBFRAME szótárat katalogizálási kísérleti projektjében (BIBFRAME Pilot). A projektben résztvevő feldolgozó könyvtárosok – 2018-ban körülbelül hatvan fő – a kezükbe kerülő bibliográfiai forrásokat először RDF-alapokon dolgozzák fel, azaz elkészítik a művekre, megjelenési formákra, valamint példányokra vonatkozó metaadatleírás-halmazokat. Ezt követi a munka második fázisa, amelynek során MARC-ban is előállítják a szükséges bibliográfiai, valamint authority rekordokat.⁴⁵ A projekten kívül maradó feldolgozók (2018-ban körülbelül 200 fő) által előállított, eredendően MARC-ban születő leíró rekordokat BIBFRAME-gráffá konvertálják, s ugyanígy tesznek az authority rekordokkal, például az egységesített címet tartalmazó authority rekordok ugyancsak BIBFRAME-műveket reprezentáló gráfokká alakulnak a konverziós folyamatban. Az egyéb authority rekordok az LCNAF-ba kerülve a Library of Congress által üzemeltetett URI-szolgáltatás (id.loc.gov) részét képezik.

Az így előálló adathalmazra épül a projektben használt BIBFRAME Editor működése, amely a <http://bibframe.org/bfe/index.html> címen hozzáférhető, és szabadon, regisztráció nélkül kipróbálható. Fő fejlesztője Kevin Ford volt, jelenlegi karbantartásáért, frissítéséért Kirk Hess és Jodi Williamschen felel (Miller, 2020a). A BIBFRAME Editor Workspace Browse füle segítségével az előzőleg a rendszerbe került metaadatleírás-

⁴⁵ A kétszer egymás után elvégzett forrásleírás jelentősen lecsökkenti a feldolgozó munka időhatékonyságát. Ez a probléma ugyancsak sarkallta a Library of Congress munkatársait a BIBFRAME-ről MARC-ra történő konverzió mielőbbi megvalósítására (McCallum, 2018).

halmazok, azaz bibliográfiai forrásokat reprezentáló gráfok tekinthetők meg, ezek művek, illetve megjelenési formák leírásait adják. Amennyiben szükséges, az elmentett tételek módosíthatók is.

Új bibliográfiai forrás leírása az Editor fülre kattintva kezdeményezhető, a megfelelő bibliográfiai forrástípus kiválasztásával. Egyelőre 11-féle beviteli űrlapból, azaz speciális BIBFRAME-alkalmazási profilból lehet választani, amelyek az alábbiak:

- Monograph
- Notated Music
- Serial
- Cartographic
- Sound Recording: Audio CD
- Sound Recording: Audio CD-R
- Sound Recording: Analog
- Sound Recording: Cassette
- Moving Image: BluRay DVD
- Moving Image: 35mm Feature Film
- Rare Materials.

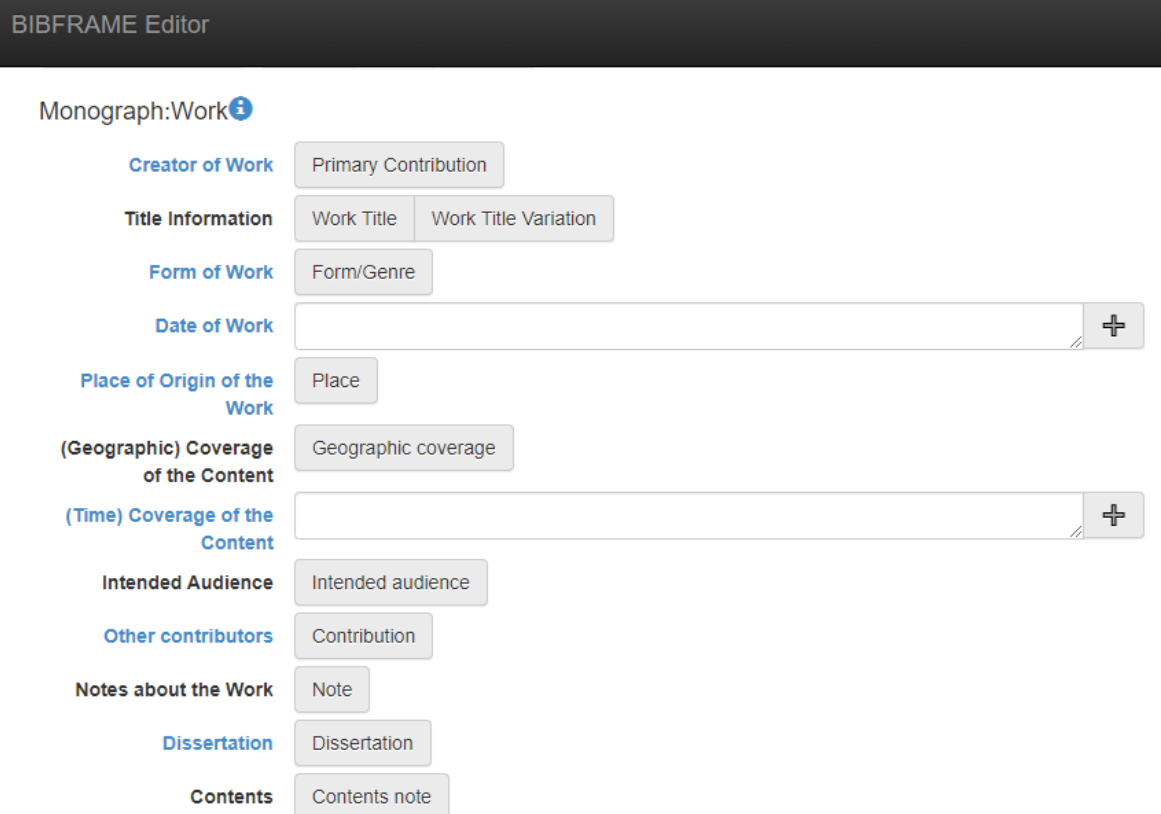
Minden űrlap bővíthető: a képernyő alján elhelyezkedő Add Property mezőben további metaadatok vehetők fel, de csak olyanok, amelyekhez rendelkezésre áll a megfelelő relációazonosító a BIBFRAME 2.0-ban, a BFLC-ben, sőt akár a MADSRDF vagy a PREMIS szótárban. Nem lehet túlfeszíteni a szótárak biztosította kereteket.

A bibliográfiai forrástípusok mindegyikéhez megadható az alkalmazandó entitásszint, amely a Work vagy az Instance lehet. A megjelenő űrlapon három beviteli lehetőség adódik:

- egyszerű beviteli mezők, amelyek szöveges értékeket tartalmazhatnak, az ismételhetséget a mező után látható pluszjel jelzi (például Edition Statement, Copyright Date, Dimensions);
- olyan beviteli mezők, melyek értéke kizárólag valamely szótárkódolási séma eleme lehet, ezek az értékek legördülő menü segítségével adhatók meg (Mode of Issuance, Media Type, Carrier Type, Base Material stb.);
- olyan adatok, ahol a rögzítendő érték több adatelemből áll – ekkor a megfelelő nyomógombra kattintva, a megjelenő előugró ablakban lehet rögzíteni a szükséges

értékeket (például egy megjelenési forma leírásakor a Title Information, amely Instance Title, Variant Title, Parallel Title, illetve Collective Title lehet – ezek értékei a megfelelő nyomógombra kattintva adhatók meg.)

A forrásleírási munkában segítséget nyújt, hogy számos beviteli mező címkéjére kattintva a (mára már elavult, 3R-projekt előtti) RDA Toolkit nyílik meg, amely az adatelem helyes rögzítéséhez szükséges ismereteket, szabályokat közli.



The screenshot shows the BIBFRAME Editor interface for a 'Monograph:Work'. The interface is organized into a list of categories on the left, each with a corresponding button on the right. The categories and their buttons are:

- Creator of Work**: Primary Contribution
- Title Information**: Work Title, Work Title Variation
- Form of Work**: Form/Genre
- Date of Work**: [Empty input field with a plus sign icon]
- Place of Origin of the Work**: Place
- (Geographic) Coverage of the Content**: Geographic coverage
- (Time) Coverage of the Content**: [Empty input field with a plus sign icon]
- Intended Audience**: Intended audience
- Other contributors**: Contribution
- Notes about the Work**: Note
- Dissertation**: Dissertation
- Contents**: Contents note

18. ábra – Mű metaadatelemeit tartalmazó űrlap a BIBFRAME Editor felületén
(képernyőkép, saját szerkesztés)

Az adatmezők kitöltése után a BIBFRAME Editor előnézeti lehetőséget kínál: a bevitt adatok Turtle, RDF/XML, illetve JSON-LD-formátumban is megtekinthetők, illetve egy, a JSON-LD-t kiindulópontnak használó adatvizualizációs eszköz segítségével meg is jeleníthetők.

A Library of Congress munkatársa, Jodi Williamschen, a Firenzében rendezett European BIBFRAME Workshopon (2018) az alábbi újdonságokra hívta fel a közösség figyelmét a BIBFRAME Editorral kapcsolatban:

- a személyek authority adatai most már nem csupán a névelemek, hanem LCCN (Library of Congress Control Number) alapján is kereshetők, illetve ennek alapján beemelhetők az Editorral készített forrásleírásba;
- URL alapján lehetőség van külső helyről származó BIBFRAME-művek leírásának importálására, illetve szerkesztésre történő megnyitására;
- LCCN vagy a Voyager integrált könyvtári rendszer által osztott BibID alapján lehetőség van MARC-rekordok beemelésére;
- az elkészült leírás másolható (klónozható), így jelentős mennyiségű munkát lehet megtakarítani.

Matt Miller, a NDMSO munkatársa 2020 májusában ismertette a BIBFRAME Editor felhasználói felületének megújítására vonatkozó elképzeléseket. Az átalakítást a BIBFRAME-katalogizálási kísérleti projektben résztvevők egyre növekvő száma, továbbá a projekt során felhalmozódott tapasztalatok indokolták. A NDMSO – a Samhaeng koppenhágai céggel szerződéses viszonyban – javaslatokat dolgozott ki a felhasználói élmény javítására. Az egyelőre demonstrációs környezetben megtekinthető felület valóban modernebb hatást kelt, mint elődje, részben az alkalmazott színeknek köszönhetően. A két legfontosabb entitás, a Work, valamint az Instance egyetlen képernyőn szerkeszthető, ami meggyorsítja a munkát, a képernyő bal oldalán pedig lehetőség van az űrlap bizonyos adatmezőire ugrani egy tartalomjegyzék-sáv segítségével. Ezen kívül számos designbeállítási lehetőséget is nyújt a rendszer, egyebek mellett a betűk típusa és mérete tekintetében, továbbá módosítható a szoftver színösszeállítása, és szabályozható a diakritikus jelek bevitelének módja is.

Work

Creator of work

Name
Taunton Press ×

Relationship designator

Title information

Title type
Work title

Preferred title for work
Fine woodworking

Form/Genre

Periodicals ×

Date of work

January/February 2012

Identifiers

Identifier
ISSN

ISSN
0361-3453

Contribution

Name
Taunton Press ×

Relationship designator

Classification numbers

Classification number
Library of Congress classification number

Library of Congress classification number
TT186 RT

Classification item number

19. ábra – Megnövelt felhasználói élményt kínáló, továbbfejlesztett BIBFRAME Editor-felület (Miller, 2020b)

9.7.2. A BIBFRAME Profile Editor

A Library of Congress által üzemeltetett weblapon a BIBFRAME-alapokra helyezett forrásleíró eszköz mellett lehetőség van maguknak a beviteli űrlapoknak a megformálására is. Az űrlapok alkalmazási profiloknak minősülnek, azaz arra szolgálnak, hogy segítségükkel az egyes entitásokhoz hozzárendelhetők legyenek a leírandó metaadatelemek, és a rendelkezésre álló szótár(ak)ból ezekhez URI-kat lehessen társítani. Ezeket az

alkalmazási profilokat a BIBFRAME Profile Editor nevű segédeszközzel lehet elkészíteni, illetve módosítani a már korábban előállítottakat. A <http://bibframe.org/profile-edit/> webcímen hozzáférhető eszköz e disszertáció írásának idején összesen 33 különféle profilt tartalmaz, mint például:

- adminisztrációs metaadatok;
- ágensek;
- osztályozási jelzetek;
- azonosítók

és természetesen az egyes bibliográfiai forrástípusok leírásához szükséges művek, megjelenési formák, illetve példányok űrlapjai is megtalálhatók. Jodi Williamschen (i.m.) a profilokat három típusra osztja:

- olyan profilok, melyek a leírandó bibliográfiai forrástípustól függetlenek, például az adminisztrációs adatok;
- a forrástípusok specifikus profiljai, amelyek általában a három magentitás – Work, Instance, Item – leírásához szükséges metaadatelemeket tartalmazzak;
- olyan profilok, melyek a forrástípusok űrlapjainak egyes almenüit aktiválva jelennek meg, ilyen például az azonosítók leírására használható profil.

Amikor új profil készül, az első lépés az adminisztrációs metaadatok megadása. Az alábbi adatmezők kitöltését kell elvégezni:

- ID – egyedi azonosító, mely a profilt azonosítja;
- Description – a profil rövid szöveges leírása;
- Author – a profilkészítő neve;
- Title – a profil megnevezése;
- Date – a profil készítésének dátuma;
- Remark – megjegyzések a profillal kapcsolatban;
- Source – hivatkozás valamely forrásra, amely további információkat tartalmaz a profillal kapcsolatban;
- Adherence – a profilban követett szabályok leírása.

Profile	
ID*	<input type="text" value="eg., profile:MyRDAProfile"/>
Description*	<input type="text"/>
Author*	<input type="text"/>
Adherence	<input type="text"/>
Title*	<input type="text"/>
Date	<input type="text"/>
Remark	<input type="text"/>
Source	<input type="text"/>

20. ábra – A BIBFRAME profilszerkesztő felület részlete az adminisztrációs metaadatelemek beviteli mezőivel (képernyőkép, saját szerkesztés)

A következő feladat a profilban leírt entitások sablonjainak megalkotása, amelyet az Add Resource Template feliratú gombra kattintva lehet elvégezni. A konfigurálás itt is adminisztrációs adatok megadásával veszi kezdetét:

- ID – az adott forrás egyedi azonosítója, amely nem azonos a forrás URI-jával;
- Resource Label – a forrástípus megnevezése, például BIBFRAME Work;
- Guiding statement for the use of this resource – kiegészítő, eligazító szöveg a forrásról, amely megjelenik a felhasználói felületen;
- Resource URI – a forrás URI-ja;
- Author – a forrás-template alkotójának neve.

Végül az egyes források leírására használható tulajdonságok rögzítése következik, amely lépés lényegében a beviteli mezők definiálását jelenti. Egy metaadatelemet csak akkor lehet hozzáadni a profilhoz, ha ahhoz rendelkezésre áll a hozzáférhető szótárakban a megfelelő reláció. Ezek az elemkészletek a BIBFRAME 2.0, a BFLC hivatalos kiegészítő szótár, az RDF, illetve az RDFS alapszótára, a schema.org, a PREMIS, továbbá a MADSRDF.

Ezen a szinten is találhatóak adminisztrációs metaadatok, amelyek között az alábbiakat kell rögzíteni:

- a metaadatelem URI-ja (a reláció kiválasztásával automatikusan beemelődik);
- a metaadatelem címkéje, amely a felhasználók számára megjelenik;
- a metaadatelem típusa (például literál);
- eligazító közlés a tulajdonságról – ide kerül a metaadatelem leírásának módját szabályozó RDA Toolkit-fejezetpontra mutató hivatkozás;
- a metaadatelem ismételtetésére vonatkozó beállítás;
- a metaadatelem kötelezőségére vonatkozó beállítás.

The screenshot displays the BIBFRAME profile editor interface. It is organized into several sections:

- Property URI***: A text input field with a warning icon.
- Type***: A dropdown menu set to "literal".
- Mandatory**: A dropdown menu set to "False".
- Value Constraint**: A section containing:
 - Value Language**: A text input field.
 - Language Label**: A text input field.
 - Editable**: A dropdown menu.
 - Validate Pattern**: A text input field.
 - + Add Default**: A blue button.
- Value Data Type**: A section containing:
 - URI**: A text input field with a "Select Data Type" button.
 - Label Hint**: A text input field.
- Templates**: A section with a "+ Add Template" button.
- Values**: A section with a "+ Add Value" button.

On the right side of the interface, there are additional fields:

- Property Label***: A text input field.
- Guiding statement for the use of this property**: A text input field.
- Repeatable**: A dropdown menu set to "True".
- Language URI**: A dropdown menu.
- Remark**: A text input field.
- Label**: A text input field.
- Remark**: A text input field.

21. ábra – A BIBFRAME profilszerkesztő felület részlete az egyes metaadatelemek paraméterezésére szolgáló beviteli mezőkkel (képernyőkép, saját szerkesztés)

A Value Constraint szakasz beállításával rögzíthető, ha egy beviteli mezőnek alapértelmezett értéke van: megadható az ehhez az értékhez tartozó címenév, továbbá URI, amennyiben az érték valamilyen szótárkódolási sémából származik – e szótárkódolási séma elérhetőségét a Values szakaszban kell rögzíteni, például az RDA médiatípusok esetében. Ugyanitt van lehetőség letiltani valamilyen beviteli mezőnek a szerkeszthetőségét is. A mezőbe rögzítendő érték adattípusa (azaz az értékekre kötelezően alkalmazandó osztálybesorolás) a Value Data Type szakaszban rögzített URI-val szabályozható.

9.7.3. A Sinopia

A *sinopia* kifejezés eredetileg a freskók festését megelőzően elkészülő, vörösesbarna pigmenttel készülő előzetes, vázlatos rajzot jelenti. Ez a megnevezése ugyanakkor egy szintén előzetes, ideiglenes, kísérleti linked data-alapú katalogizálási felületnek is, amelyet a LD4P projekt második fázisában kezdtek el fejleszteni; a rendszer mögött található szoftverkomponensekről Jeremy Nelson (2019) készített összefoglalót. A felület ingyenes és gyors regisztrációt követően szabadon kipróbálható.⁴⁶ Kidolgozásának indokait Schreur (2019) ismerteti:

⁴⁶ Hozzáférhető a <http://sinopia.io> címen [letöltés: 2022. május 19.].

- segítségével ki lehessen alakítani egy folyamatosan bővülő linked data-tárat (data pool), amely BIBFRAME-alapú alkalmazási profilok/úrlapok használatával jön létre;
- szülessen egy olyan kiterjedt, könyvtárakat tömörítő szakmai közösség, amely képes linked data-alapú katalogizálást és adat-újrafelhasználást végezni egy felhőalapú tesztkörnyezetben;
- váljon lehetségessé olyan irányelvek, technikák, illetve munkafolyamatok kidolgozása, amelyek a linked data létrehozását és felhasználását szabályozzák;
- legyen megoldható a könyvtári metaadatok és azonosítók webes integrációja, a Wikidata adatbázissal történő együttműködésben.

Bejelentkezés után két fontos funkcióhoz biztosít hozzáférést a rendszer: a Linked Data Editor menüpontra kattintva a megfelelő úrlap (azaz BIBFRAME-alkalmazási profil) kiválasztása után elkezdhető a forrásleírás – hasonlóan a BIBFRAME Editorhoz, itt is hivatkozások jelennek meg a használandó RDA-fejezetpontokra. Egyes adatelemek megadásában külső adatforrások is segítenek: így például a műfajt (form of work) leíró kifejezések a Getty, a MESH, illetve a Library of Congress által biztosított szótárkódolási sémákból is származhatnak. Bizonyos nyelvek speciális karaktereinek, írásjeleinek bevitelére külön karaktertáblák vehetők igénybe. A kész leírás táblázatként, továbbá Turtle és N-Triples formátumban is megjeleníthető, majd az intézményi affiliáció rögzítése után eltárolható. A Sinopiában ugyancsak rendelkezésre áll profilszerkesztő modul, ennek működése – sőt kinézete is – igen hasonló a Library of Congress által üzemeltetett szolgáltatáséhoz.

Lehetőség van RDF-állományok importjára is: a felület Turtle, N-Triples, N-Quads, N3, valamint TriG⁴⁷ szerializációs szintaxisokat képes befogadni. A használandó bázis-URI megadása után megtörténik a gráf importja, amelyet aztán a Sinopia felületén lehet tovább szerkeszteni.

⁴⁷ A W3C által 2014-ben definiált RDF-szerializációs szintaxis, a Turtle kiterjesztése (W3C, 2014b).

Title Information*

Work Title + Add another Work Title

Property: <http://id.loc.gov/ontologies/bibframe/title>
Class: <http://id.loc.gov/ontologies/bibframe/Title>

▼ Preferred Title for Work* 🗑️

Property: <http://id.loc.gov/ontologies/bibframe/mainTitle>
Enter a literal

 ä en 🗑️

+ Add Part number
+ Add Part name
+ Add Note

Classification numbers 🗑️

Dewey Decimal Classification + Add another Dewey Decimal Classification

Property: <http://id.loc.gov/ontologies/bibframe/classification>
Class: <http://id.loc.gov/ontologies/bibframe/ClassificationDdc>

+ Add DDC Classification number
+ Add Dewey Edition

Library of Congress Classification + Add another Library of Congress Classification

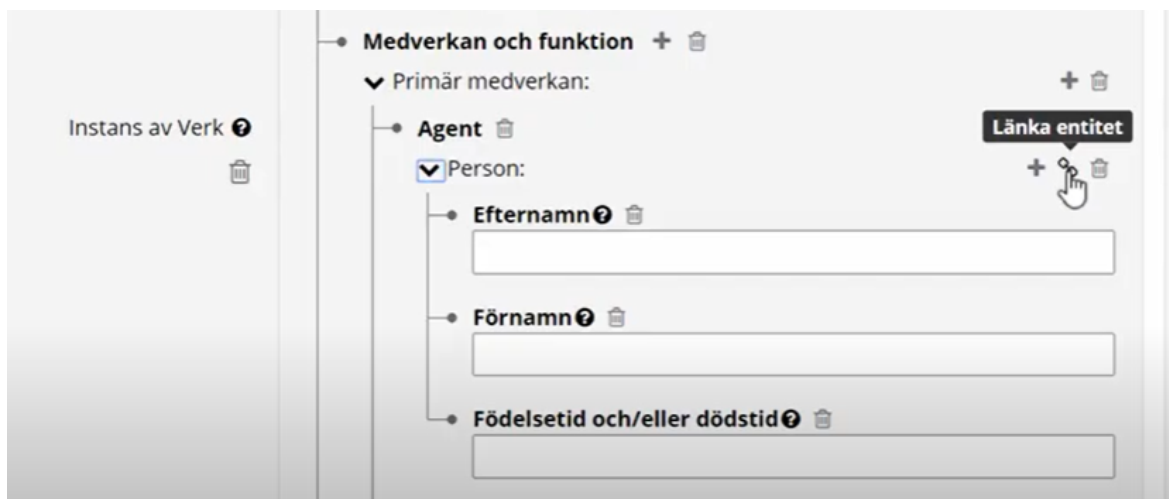
Property: <http://id.loc.gov/ontologies/bibframe/classification>
Class: <http://id.loc.gov/ontologies/bibframe/ClassificationLcc>

+ Add LC Classification Number
+ Add Classification item number

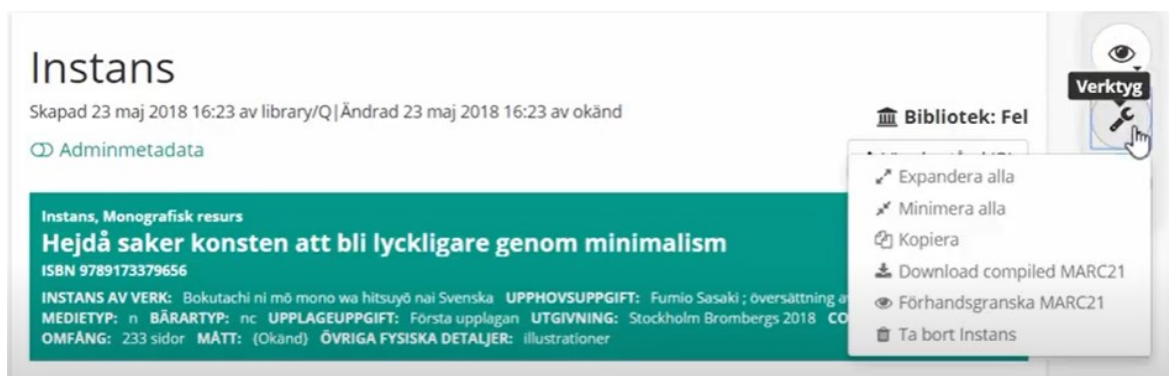
22. ábra – Metaadat-beviteli mezők a Sinopia felületén (képernyőkép, saját szerkesztés)

9.7.4. A LIBRIS XL

A LIBRIS Svédország könyvtárainak közös katalogizáló felülete – és az első, éles körülmények között alkalmazott, linked data-alapú forrásleírási eszköz. Az entitásalapú űrlapok használatával feldolgozott bibliográfiai források adatai már RDF-ben, gráfként tárolhatók, a végeredmény pedig MARC 21-be exportálható, hogy a közös katalógust gyarapító könyvtárak lokális rendszereibe is beilleszthető legyen. Egy 2019 végi változásjegyzék tanúskodik arról, hogy a MARC-export a LIBRIS felületére bejelentkezett intézmények egyedi exportbeállításai alapján szabályozható, azaz az egyes könyvtárak sajátosságai érvényesíthetőek a LIBRIS-ből történő adatkihozatal során (KB, 2019).



23. ábra – A közreműködő adatainak rögzítése a LIBRIS egy leíró űrlapján (Libris katalogisering av bok, 2018)




24. ábra – A LIBRIS felületén elkészült leírás letöltése és megjelenítése MARC 21-ben (Libris katalogisering av bok, 2018)

Az űrlapok „mögött” rejlő adatmodellt KB Base Vocabulary-nak nevezik,⁴⁸ amelynek prefixe az RDF-állományokban a *kdv*. Alapját a BIBFRAME 2.0 szótár adja, de készítéséhez felhasználták a BFLC-t, több Dublin Core-szótárat, a BIBO-t, az SKOS-t, a MADS-ot, sőt a FOAF-ot és a schema.org-ot is. A körülbelül 300 osztályt, illetve nagyjából 400 tulajdonságot definiáló adatmodell – ellentmondva a linked data jó gyakorlatának – minden szótárelemhez saját URI-t definiál, noha ezeket a legtöbb esetben meg lehet feleltetni más szótárakból vett osztályoknak, illetve tulajdonságoknak, mint az a 25. ábrán is látszik:

⁴⁸ Hozzáférhető a <http://id.kb.se> oldalon [letöltés: 2022. május 19.].

birth date | birthDate

URI (link to resource)	https://id.kb.se/vocab/birthDate 
Domain	Person
Is defined by	KB Base Vocabulary
Label	en birth date sv födelsedatum
Equivalent property	http://schema.org/birthDate http://www.loc.gov/mads/rdf/v1#birthDate

25. ábra – Egy tulajdonság legfontosabb adatainak megjelenítése a LIBRIS szótárának információs weboldalon (képernyőkép, saját szerkesztés)

Érdekesség, hogy az adatmodellt helyenként a MARC adatelemeinek pontosan megfeleltetett relációk egészítik ki, ezeket *marc* prefix-szel látták el, és MARCFrame-ként hivatkoznak rájuk.⁴⁹ Ilyen módon ezek a relációk nem tekintendők a bázisszótár elemeinek, a hozzájuk rendelt értékek azonban fontos részei a gráfnak, használatukkal az eredetileg MARC-rekordok formájában tárolt adattartalom konverziója rendkívül pontosan, adatvesztés nélkül végbemehet. Az alábbi példában egy specifikus adatelem rögzítése látható, amelyet két szótár elemeinek vegyítésével végeznek el. A MARC 21-rekordban a 240\$k almezőben található érték gráfban történő megjelenítéséhez a KBV-szótár nem definiál relációt, így az almező értéke csak kiegészítő szótár segítségével konvertálható RDF-re:

```
<kbv:hasTitle>
  <kbv:Title>
    <kbv:mainTitle>Works.</kbv:mainTitle>
    <marc:formSubheading>Selections</marc:formSubheading>
  </kbv:Title>
</kbv:hasTitle>
```

9.7.5. A RERO

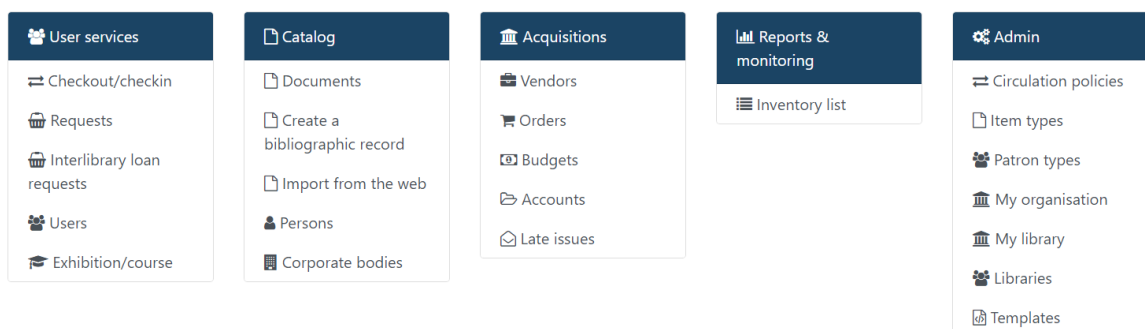
A RERO rövidítés Nyugat-Svájc könyvtári hálózatát jelöli. A csaknem harminc éve létrejött együttműködési projekt számos egyetemi könyvtárat, közkönyvtárat, műemlék-könyvtárat (heritage library), és szakkönyvtárat (összesen körülbelül 180 intézményt) tömörít Geneva, Friborg, Jura, Neuchâtel és Valais kantonokból. Explore nevű közös katalógusfelületén csaknem hétmillió bibliográfiai forrás leírása kereshető (RERO, é.n.).

⁴⁹ E speciális relációk, illetve a MARC-mezők közötti megfeleltetések ugyanis a <https://id.kb.se/marcframe/> oldalon találhatóak.

A Martigny városában működő RERO 2018-ban jelentős átalakítási projektbe fogott, amelynek keretein belül a hálózatot szolgáltatási és kompetenciaközponttá szeretnék fejleszteni. E kezdeményezés RERO21 néven vált ismertté, talán tudatos utalással a MARC 21 adatsere-formátumra. A projekt egyik fontos részfeladata egy saját integrált könyvtári platform kifejlesztése, melyet egyszerűen RERO ILS-nek neveztek el. A multitenant, SaaS szoftvereszköz az Invenio nevű, nagyméretű digitális repozitóriumok, digitális könyvtárak, illetve kutatási adatok kezelésére, menedzselésére szolgáló, open-source keretrendszer átalakításából születik, s jelen értekezés írásának idején is fejlesztés alatt van (Milhit, 2019). A rendszer használatát többnyelvű authority állomány, a Multilingual Entity File (MEF) támogatja, amelyet folyamatos szinkronizálnak a német, illetve francia nemzeti könyvtár authority állományaival.

A <https://ils.test.rero.ch/professional/> weboldalon látható a rendszer aktuális állapota. Bejelentkezés után elérhetővé válnak a könyvtárosi, adminisztrátori funkciók, köztük a katalogizálás, illetve a személy- és testületi nevekre történő keresés. A keresett személy vagy testület adatai – a MEF API-nak, illetve a külső névterek elemeivel történő adatgazdagításnak hála – nem csupán a RERO adatbázisából, hanem a GND-ből, a BNF-ből, illetve az IDREF-szolgáltatásból is származnak, illetve át lehet kattintani a szerzőhöz, közreműködőhöz kapcsolódó forrásleírásokra is.

RERO ILS administration



26. ábra – A RERO integrált könyvtári platform kezelőfelülete 2022. áprilisi állapot szerint (képernyőkép, saját szerkesztés)

A forrásleírási programmodul erősen leegyszerűsített a MARC-hoz, vagy akár az előzőekben bemutatott más linked data-alapú katalogizáló felületekhez képest. (Apróbb terminológiai zavar is jelentkezik: a funkcióhoz vezető menüpont megnevezése (create a bibliographic record) a linked data kontextusában már nem értelmezhető *rekord* fogalmat

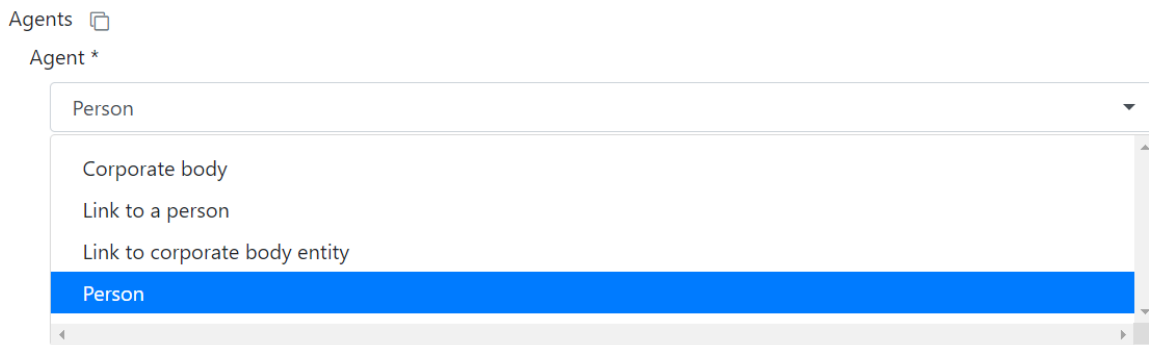
tartalmazza. A leírandó bibliográfiai forrástípustól függetlenül összesen körülbelül 20 metaadatelem-csoportot kínál a rendszer, egyesek alapértelmezetten részei az űrlapnak, míg másokat opcionálisan lehet hozzáadni. A beviteli mezők jelentős része lokális szótárból választható, vagy beírható értéket vár.

The screenshot shows a web form titled "Identifiers" with a trash icon. It is divided into three main sections:

- Identifier:** Contains a dropdown menu for "Type" with "ISBN" selected, and a text input field for "Identifier value".
- Types:** Contains a dropdown menu for "Type" with "book, text" selected, and a dropdown menu for "Subtype" with "book, text" selected.
- Mode of issuance:** Contains a dropdown menu for "Mode of issuance" with "single unit" selected, and a dropdown menu for "Subtype" with "Material unit" selected.

27. ábra – A RERO adatbeviteli űrlapjának egy részlete könyvjellegű bibliográfiai források esetén (képernyőkép, saját szerkesztés)

A metaadatelemek csupán kis részéhez, így például a személy- és testületi nevekhez, helynevekhez, műfajt jelölő fogalmakhoz társítható külső azonosító, és még ezek esetében sem kötelező; választani lehet a literálrögzítés, illetve az URI társítása között, amely személy- és testületi nevek esetében – ahogy korábban írtam – az IDREF-ből, illetve a GND-ből származhat.



28. ábra – Ágens rögzítési módjának paraméterezése (képernyőkép, saját szerkesztés)

A rendszer emellett (ugyancsak alternatívaként, nem kötelezően) beemelhetővé teszi időtartamok, sőt művek URI-jait is. Az egyes forrásleírásokhoz példányinformációk is rendelkezhetők, amelyek rögzítése egy külön beviteli űrlapon történik. Ezek az információk egyebek mellett a vonalkódot, jelzet(ek)et, illetve a kölcsönözhetőségi információkat foglalják magukba.

Negatívumként említhető, hogy a forrásleíró űrlap nem kezeli elkülönítve az egyes entitások metaadatelemeit, az aktuális entitásszintre sem az űrlap szerkezetéből, sem az elemek címkéiből nem lehet következtetni. A BIBFRAME-szótár maximális kapacitása nincsen kihasználva, a RERO rendszer nem kínálja fel mindazokat a specifikus metaadatelemeket, amelyek a BIBFRAME felhasználásával kifejezhetőek lennének (az alkalmazható megjegyzések listája például megdöbbenően alulspecifikált, csak négy elemet tartalmaz), s ugyancsak erősen válogatva teszi felhasználhatóvá az RDA értékkészleteit. A példányadatok űrlapja lokális szótáras elemeket gyakrabban, míg külső forrásból beemelhető értékekkel bíró adatmezőt egyáltalán nem tartalmaz.

Barcode
<input type="text"/>
Call number
<input type="text"/>
Unit
Example: Volume 1
Circulation category *
Select an option
Location *
Select an option
<input checked="" type="checkbox"/> New acquisition
2022-04-23
Masked
<input type="checkbox"/> Masked

*29. ábra – A példányadatok megadására szolgáló űrlap a RERO rendszerben
(képernyőkép, saját szerkesztés)*

A RERO nyilvános katalógusfelületén⁵⁰ az elkészült, RDF-logika szerint tárolt bibliográfiai leírások JSON-formátumban megtekinthetők.

⁵⁰ Hozzáférhető a <http://ils.test.rero.ch> címen [letöltés: 2022. május 19.].

```

"contribution": [
  {
    "agent": {
      "preferred_name": "Orwell, George",
      "type": "bf:Person"
    },
    "role": [
      "aut"
    ]
  },
  {
    "agent": {
      "preferred_name": "Queval, Jean",
      "type": "bf:Person"
    },
    "role": [
      "trl"
    ]
  },
  {
    "agent": {
      "preferred_name": "Decker, Jennifer",
      "type": "bf:Person"
    },
    "role": [
      "ctb"
    ]
  }
],

"identifiedBy": [
  {
    "type": "bf:Isbn",
    "value": "9782072820908"
  },
  {
    "type": "bf:Local",
    "value": "cantook-EDEN448324"
  },
  {
    "source": "cantook",
    "type": "bf:Local",
    "value": "oai:ebooks.rero.ch:cantook-EDEN448324"
  }
],

```

30. ábra – A RERO-rendszer által előállított, JSON formátumban megjelenített leírás részletei (képernyőkép, saját szerkesztés)

A 30. ábrán látottak alapján megfigyelhető, hogy a metaadatsémában jelen vannak a BIBFRAME-szótár elemei, funkciójuk a bevitt konkrét metaadat-érték valamilyen osztályba sorolása. Így például az űrlapon rögzített személy a `bf:Person` osztályba tartozik, az alkalmazható azonosítók a nekik megfelelő osztályba (`bf:Isbn`, `bf:Local`, `bf:PublisherNumber` és mások), a főcím a `bf:Title` osztályba, a tárgyszó pedig a `bf:Topic` osztályba.

```
"contentMediaCarrier": [  
  {  
    "carrierType": "rdact:1049",  
    "contentType": [  
      "rdaco:1020"  
    ],  
    "mediaType": "rdamt:1007"  
  }  
],
```

31. ábra – Metaadatelemek az RDA szótárkódolási sémáiból származó értékekkel
(képernyőkép, saját szerkesztés)

A 31. ábra bemutatja, hogy a metaadatelemek némelyikéhez tartozó szótárkódolási sémákat az RDA biztosítja. Ilyen metaadatelemek például a hordozótípus (`rdact:` prefixszel), a tartalomtípus (`rdaco`), vagy a médiatípus (`rdamt`), illetve RDA-szótár van a felhasználó segítségére a megjelenés módja (`rdami`), vagy bizonyos források (mozgóképek) esetében a színtartalom jellemzőinek (`rdacc`) rögzítésekor is.

9.7.6. Összegzés

A felsorolt forrásleíró felületek a jelen értekezés szerzője szerint a legismertebbek közé tartoznak, jelentős közösségi munkával jöttek létre, funkcióikkal, lehetőségeikkel útmutatást nyújthatnak más katalogizáló eszközök, programmodulok kidolgozásához. A példákából, illetve a mellékelt képernyőképekből jól látszik, hogy a forrásleírással foglalkozó könyvtárosok, közgyűjteményi szakemberek elé kerülő űrlapok egyáltalán nem bonyolultabbak, mint a MARC esetében, kezelésük magától értetődő. Több esetben hivatkozást találunk az alkalmazandó szabályzatpontokra, amelynek segítségével a munka még gördülékenyebbé válhat. A BIBFRAME egyedi, a fogalmi modellektől eltérő entitásstruktúrája a legtöbb esetben egyértelműen nyomon követhető, a forrásleírást végző szakember könnyen eligazodik a beviteli űrlapok között, könnyedén válthat az egyes entitásszintek leírásai között, ilyen módon az sem jelent nagyobb munkaterhet, hogy egyetlen forrás leírásakor három, vagy még több entitás-előfordulás metaadatait kell rögzíteni.

Az űrlapokat kialakító, a kollégák igényei által optimalizáló szakinformatikus kollégák munkája azonban valamivel nehezebb lett, hiszen jóval összetettebb az a

munkafolyamat, amelynek során új metaadatelemet lehet adni az egyes beviteli adatlapokhoz (profilokhoz). E munkához körültekintő linked data-alkalmazási profilt (lásd a 7.1.2. fejezetet) kell készíteni. A leíráshoz elsődlegesen alkalmazott szótár pontos ismerete, az értelmezési tartományok és értékkészletek területén kialakított tájékozottság mellett hasznos, ha a szakinformatikus jól ismeri a linked data-ökoszisztéma szélesebb körét, és új metaadatelem rögzítésére irányuló igény esetén tanácsot tud adni, hogy mely szótárat, szótárakat érdemes felhasználni. E javaslatok kialakításakor természetesen a linked data jó gyakorlataira (különösen a reláció-URI-k újrafelhasználására) is tekintettel kell lenni.

10. A BIBFRAME az Országos Széchényi Könyvtár linked data-miniprojektjében

Az Országos Széchényi Könyvtár linked data-alkalmazási kísérleteiről az értekezés több pontján is esett már szó. A 7. fejezetben utaltam az intézmény által 2010-ben végrehajtott adatpublikációs projektre, amelynek során a törzsgyűjteményi olvasói katalógust, valamint a Digitális Könyvtár teljes tartalmát és az ezekhez kapcsolódó authority adatokat tették közzé RDF-logikára épülő formátumban. Ez a publikáció a világon a második ilyen jellegű kísérlet volt: a svéd közös katalógust (LIBRIS, lásd a 9.7.4. fejezetben) tudja hazánk maga előtt. Szintén a hetedik fejezetben foglalkoztam a Magyar Elektronikus Könyvtár és a Digitális Képtár által közzétett adathalmazokkal, az 5.5. fejezetben pedig bemutattam azokat a kísérleteket, amelyek az RDA katalogizálási szabályzat elterjesztését, bemutatását, illetve az entitásszemlélet könyvtárosokkal történő megismertetését célozzák.

Az RDA-kísérleteket a *Molnárathonnak* nevezett szemantikus webes miniprojekt egészítette ki, illetve tette teljessé: a KOHÁ-ban tesztelési céllal, MARC 21-ben létrehozott bibliográfiai rekordokat a BIBFRAME szótár relációival kívántuk RDF-re konvertálni, majd a Virtuoso segítségével elérhetővé és kereshetővé tenni.

A munka kezdetén nyers, bináris formátumban rendelkezésre álló rekordokat a Library of Congress által elkészített konverziós stíluslapra épülő BIBFRAME Testbed szoftverfunkció⁵¹ segítségével ültettem át BIBFRAME-re. Ez a tesztkonverzió fontos tanulságokkal szolgált a következő lépésben kidolgozandó, az OSZK adataira optimalizált konverziós útmutató elkészítéséhez (lásd a 2. mellékletet), amely bemutatta, hogy mely almezőértékeket milyen BIBFRAME-osztályok, illetve -tulajdonságok segítségével kell triplétté konvertálni, s ugyancsak rögzítette, hogy milyen esetekben kell egyedi, intézményi URI-kat létrehozni az adatszerkezetben. Az útmutatóban rögzítettek alapján a tényleges konverziós eszközt – a 7.1.3. fejezetben tárgyalt stíluslapot –, majd az adatok Virtuoso-adatbázisba töltését⁵² az OSZK külső munkatársa, Vitéz Ildikó készítette el.

A projekt legfontosabb tanulságai az alábbiak voltak:

- a konverziós eszköz előállításakor különös tekintettel kell lenni a rögzített hosszúságú mezők, illetve a rekordfej által hordozott adatok kezelésére, leginkább

⁵¹ A BIBFRAME Testbed a Terry Reese által készített MARCEdit szoftver része, amelyről részletesebben szóltam az értekezés 9.6.3. fejezetében.

⁵² Az adatbázis hozzáférhető a <http://v.mek.oszk.hu/FlintSparqlEditor/index-KOHA.html> címen. [letöltés: 2021.01.23.]

arra, hogy pontosan ki legyen jelölve azt a karakterhely, amely a BIBFRAME-reláció szempontjából releváns információt hordozza. A karakterhelyek szabványos számozása ugyanis nem egyessel, hanem nullával kezdődik e mezők esetében;

- a Library of Congress által bevezetett, kiegészítő relációk – MatchKey, MarcKey, titleSortKey – használatakor a teljes mező adattartalmának megfelelő megőrzéséhez mindenhol nagyon aprólékosan meg kellett figyelni, hogy az adott mezőnek mely almezőit definiálja a MARC 21-szabvány;
- konverziós problémát okozott a KOHA-úrlap nem megfelelő konfigurációja. A 710-es mező adatainak átalakításakor a MARC-ban tárolt testületi név azonosítására létrejövő URI a \$0 almezőben található authority rekord azonosítója alapján készül el. Ha azonban a MARC-rekord nem tartalmazza, vagy más helyen tartalmazza a szükséges adatot – ebben az esetben a \$0 helyett a \$9-ben –, az azonosító a bibliográfiai rekord ID-ját felhasználva születik meg, így annyi különféle URI-ja lesz az adott testületnek, ahány rekordban az szerepel, s ez nyilvánvalóan téves eredményekhez vezet a visszakereséskor. A hibát ebben az esetben a MARC-úrlap kiegészítésével, és megismételt RDF-konverzióval javították, így már a testületek is minden esetben azonos URI-val rendelkeztek;
- az egységesített címek hasonló problémát okoztak: az eredeti, valamint a lefordított mű ugyanazt az azonosítót kapta. Ebben az esetben a javítás a két kifejezési forma megfelelő elkülönítésére alkalmas MARC-adatelem körültekintő kiválasztásával és megismételt konverzióval történt.

A javítások végeztével a projekt sikerrel zárult, a tesztalmazon releváns SPARQL-keresések futtathatók. Bebizonyosodott, hogy a BIBFRAME sikerrel alkalmazható egy kisebb tesztalmaz konverziójára, megfelelő körültekintéssel az adott (szoftveres és intézményi környezet által meghatározott) MARC-implementáció felhasználásával készülő rekordok jó minőségben konvertálhatók BIBFRAME-gráffá.

A nemzeti könyvtár Könyvtári Szabványosítási Irodája vezetésével jelenleg (2022. második negyedévében) is zajlanak az RDA-val kapcsolatos vizsgálatok, szabványelemzések, amelyek a remények szerint az új katalogizálási szabályzat implementációjához vezetnek. Az RDA használatba vétele csupán idő kérdése, az informatikai háttérrel kapcsolatban azonban sok nyitott kérdés vár még megválaszolásra: többek között tisztázatlan a BIBFRAME szerepe, illetve az RDA-val való viszonya az Országos Széchényi Könyvtár kialakítandó linked data-ökoszisztémájában (is). A nemzeti

könyvtárban fokozatosan, modulonként üzembe állítandó Országos Könyvtári Platformon nagyon sok múlik: mindenképp kívánatos, hogy a rendszer RDA-gráfokat, illetve BIBFRAME-gráfokat is képes legyen előállítani, így az OSZK kielégítheti a legkülönbébb adatszolgáltatási igényeket, s könnyedén csatlakozhat a jövőben minden bizonnyal kialakuló aggregációs platformokhoz is.

A továbblépést az OKP linked datára épülő katalogizálási funkcióinak üzembe állítása jelentheti, amelynek segítségével tesztelhetővé és felhasználhatóvá válnak a kisebb-nagyobb, MARC-ból BIBFRAME- vagy RDA-gráffá konvertált adathalmazok, javítható a konverzióban esetlegesen sérülő adatminőség. Az új szoftverfunkciók az entitásalapú forrásleírás módszertanának megismerését, begyakorlását és elmélyítését is szolgálják, segítségükkel megállapíthatóvá válik, hogy a natív RDF-adatok megfelelően illeszkednek-e a jövőben retrospektív konverzióval előállítandó nagy, intézményi gráfba, keletkezik-e bármiféle inkonzisztencia.

11. A Petőfi Irodalmi Múzeum bibliográfiai rekordjainak konverziója

E fejezetben kísérletet teszek egy HUNMARC-ról BIBFRAME-re történő konverzió sajátosságainak és eredményeinek a bemutatására. Jóllehet, a közelmúltban több, sikeres konverzió is lezajlott a hazai közgyűjteményi szférában ebben az irányban – a Magyar Nemzeti Múzeumot a 7.1.4. fejezetben, az Országos Széchényi Könyvtárat pedig az előző, 10. fejezetben említettem –, e kísérletek nagy hiánya, hogy eredményeiket, tanulságaikat és következtetéseiket nem dokumentálták (vagy a dokumentáció nem hozzáférhető a nyilvánosság számára), így a szakmai közösség csak e projektek sikeres befejezéséről értesülhetett, illetve megkapta a SPARQL-végpontok elérhetőségét, ahol a HUNMARC-ról BIBFRAME-re konvertált adatok hozzáférhetők. Magyarázó jegyzetek, a kiinduló állományok sajátosságait, adatszerkezetét taglaló írások azonban nem jelentek meg, s a konverziós folyamatokról is csak vázlatos tudás áll rendelkezésre. Jelen értekezésben e hiányt igyekszem pótolni, amikor nem csupán adattranszformációt hajtok végre, hanem részletesen, kritikai szempontból bemutatom a bemeneti állományt (HUNMARC, illetve MARC/XML), továbbá megvizsgálom a kimeneti adatokat (BIBFRAME-gráf), és javaslatokat fogalmazok meg az adatminőség javítására vonatkozóan. A konverzió céljaira a munkahelyem – a Petőfi Irodalmi Múzeum könyvtára – adatbázisából válogattam a bibliográfiai rekordokat. Ez az adatbázis, bár számos egyediséget tartalmaz, jórészt a HUNMARC-szabványra épül, a bibliográfiai rekordok tehát alapvetően megfelelnek az adattárolási követelményeknek, számos releváns adatmezőt és almezőt tartalmaznak, így más intézményekben is jól hasznosítható következtetések vonhatók le belőlük, amelyek felhasználhatók a konverziós projektek előkészítésénél, illetve az esetleges hibák elkerülésére való felkészülésnél, valamint az új problémák elhárítási megoldásainak kidolgozásánál.

11.1. A Petőfi Irodalmi Múzeum könyvtára

A Petőfi Irodalmi Múzeum gyűjteményének egyik kiemelt fontosságú része az irodalomtörténeti értékű könyvanyag, amelynek alapjait a Petőfi Társaság könyvgyűjteménye, illetve Jókai Mór könyvhagyatéka teremtette meg.

A múzeum könyvtára sajátos, kettős helyzetben van. Egyrészt korlátozottan nyilvános, kutatói engedéllyel látogatható irodalmi-irodalomtörténeti szakkönyvtárként működik és napjainkban megkerülhetetlen az ezzel a területtel foglalkozó kutatók számára. Gyűjtőkörébe tartoznak a kortárs szépirodalom első kiadásai, az irodalmi kánon első két

vonalaiba tartozó művek, az emigráns irodalom, illetve az írói hagyatékok. A könyvállomány jó része föld alatti raktárhelyiségben található (egy része kiemelten őrzött, kulcsra zárható gördíthető polcon kapott helyet), míg a használók kényelmét és az általános tájékozódást egy körülbelül 3000 kötetet számláló segédkönyvtár szolgálja, amely a Károlyi-palota első emeletén található olvasóteremben található. E kötetek, illetve a kutatók, dolgozók által a raktárból igényelt más bibliográfiai források kizárólag helyben használhatók.

Másrésről a PIM könyvtára egytől-egyig múzeumi műtárgynak minősülő forrásokból áll, amivel egyedülálló a magyarországi múzeumi könyvtárak között. A könyvállomány egyes darabjai fontos szerephez jutnak az intézmény állandó és időszakos kiállításaiiban, akár csak a képzőművészeti alkotások, írói relikviák vagy kéziratok. Az állománygyarapítás igen erős válogatással történik, kiemelt figyelmet fordítva a dedikált példányokra, s az intézmény a hagyatékok beszerzésénél is törekszik a duplumok keletkezésének elkerülésére.

A könyvgyűjtemény feldolgozása 2008 óta a Huntéka szoftverrel történik. Az eredetileg könyvtári funkciókat szolgáló kliensprogram a Petőfi Irodalmi Múzeum munkatársai, illetve a fejlesztők hosszú éveken keresztül tartó párbeszéde, szakmai együttműködése nyomán folyamatosan vált alkalmassá múzeumi gyűjteményezési-katalogizálási funkciók elvégzésére is, s a ma már *Huntéka-M* néven ismert szoftveralkalmazás jelentős piaci részesedéssel rendelkezik a múzeumok munkájának automatizálása területén. A gyűjtemény feldolgozása HUNMARC-on alapuló bibliográfiai (múzeumi környezetben *tárgyleíró*) rekordokba történik, amelyet hagyományos authority rekordok egészítenek ki a könyvtári szabályzatoknak megfelelően.⁵³

11.2. Célkitűzések és a kísérleti adathalmaz

Kísérleti konverziós projekt végrehajtásához a múzeum könyvtárának állományából válogattam bibliográfiai rekordokat. Céljaim a következők voltak:

⁵³ Mindenképp érdemes azonban megemlíteni, hogy a Petőfi Irodalmi Múzeum által épített személynév-authority file messze túlmutat a könyvtári feladatokon, sőt építése jórészt független a katalogizáló-tárgyleíró munkától. Az intézmény munkatársai nyomtatott adatforrások, életrajzi lexikonok, ki kicsodák stb. felhasználásával hoznak létre személynévrekordokat, amelyek a hagyományos authority rekordokhoz képest sokkal adatgazdagabbak, a kutatói tájékoztatást szolgálják. Bibliográfiai (avagy tárgyleíró) rekordok ugyanakkor nem feltétlenül kapcsolódnak hozzájuk. A PIM személynév-authority állománya (gyakran a PIM névtereként emlegetik) önállóan kereshető a <https://opac-nevter.pim.hu> címen [letöltés: 2022. május 19.], amellyel kiemelkedik a hazai közgyűjteményi szférából. Az egyes névrekordok saját URI-val rendelkeznek, így felhasználhatók a szemantikus adatháló építésében.

- tapasztalatszerzés a BIBFRAME relációinak működéséről, illetve a konverziós segédeszközről;
- olyan következtetések levonása a konvertált adathalmazból, melyek felhasználhatók a kiinduló adatállomány tisztításához, szabványosításához, illetve későbbi, kiterjesztett adatkonverziók előkészítéséhez;
- a múzeum jövőbeni adatvizualizációs törekvéseinek támogatása adathalmazokkal;
- a múzeumban rögzített egyedi adatok megjelenítésének előkészítése a szemantikus világhálón.

A kiinduló adathalmazt – az elmúlt öt évben létrehozott, az irodalomtörténeti értékű könyvanyag gyűjteményi tételekhez kapcsolódó tárgyleíró rekordjait – a Huntéka szoftver segítségével MARC/XML-be exportáltam, így egy 21.826 rekordot tartalmazó, 54 megabyte méretű állomány keletkezett, amelyet stílussal már könnyedén lehet BIBFRAME-relációkkal felépülő RDF/XML-lé konvertálni.

A MARC-rekordok adatmező- illetve almező-összetétele fontos tanulságokkal szolgál az RDF-re átalakított adathalmaz vizsgálatakor: segítségével fel lehet deríteni azokat a pontokat, amelyek specifikációs probléma, vagy egyéb adathiba miatt nem jelennek meg a konvertált RDF/XML-ben; ezeken a helyeken utólagosan kell beavatkozni, majd a javításokat az állomány megismételt konverziójával ellenőrizni. A statisztika elkészítését a MARCEdit szoftver MarcEditor moduljának FieldCount funkciójával végeztem el; a szoftver az épp megnyitott rekordhalmaz valamennyi adat- illetve almezőjét összeszámlálja, és a jelentést több formátumban letölthetővé teszi. A 28. ábrán az összeszámlálási jelentés egy részlete látható, amelyből az egyes almezők ismételhetőségére vonatkozó következtetések is levonhatók: a 090-es mező \$a almezőjének előfordulási száma például azért magasabb, mint magáé a mezőé, mert az almező több rekordban ismétlődően van jelen. A teljes jelentést a disszertáció 3. melléklete tartalmazza.

Field	Subfield	In Records	Total
090		21826	21826
	\$a		22253
100		17394	17394
	\$b		5
	\$c		503
	\$a		17394
	\$g		1603
	\$d		16401
	\$j		17249
	\$m		74
110		2	2
	\$a		2
240		585	585
	\$b		46
	\$a		585
245		21826	21826
	\$b		12796
	\$c		16854
	\$7		9
	\$a		21826

32. ábra - Az összeszámlálási jelentés egy részlete (képernyőkép, saját szerkesztés)

A konverziós folyamat megtervezésénél felhasználtam az OSZK kísérleti szemantikus konverziós miniprojektjében szerzett tapasztalataimat, és a munkát egy előzetes átalakítás lefuttatásával kezdtem, hogy kiderüljön, mely pontokon kell finomítani, korrigálni a MARCEdit szoftverbe beépített konverziós stíluslapokat. Az eltérések fakadhatnak a HUNMARC, illetve MARC 21 adatformátumok közötti különbségekből, illetve a Petőfi Irodalmi Múzeum katalogizálási munkájának bizonyos, házi feldolgozási szabályzatban rögzített sajátosságaiból.

A szoftver beállításai az alábbiak voltak:

- profil: BibFrame2
- data serialization: RDFXML
- Id Field: 001

így az eredmény egy 172 megabyte-os RDF/XML file lett.

11.3. A konverzió elemzése

A következőkben a tesztalomban előforduló MARC-hívójelek mentén haladva vetem össze egymással a bemeneti MARC/XML-állomány, illetve a kimeneti RDF/XML-állomány releváns részleteit. Az egyes adatmezők konverziójának minősége eltér egymástól: ennek indokait minden esetben megvizsgálom, és javaslatot teszek a konverzióval összefüggő, adatminőség-romlási problémák elkerülésére.

000 – Rekordfej

Az adminisztrációs metaadatok szakaszában az alábbi kódrészlet keletkezik a rekordfej **5. karakterpozíciója** alapján (rekordstátusz). A karakterhely értékének átörökítése mellett címke is keletkezik, amely a felhasználó számára történő megjelenítést szolgálja.

MARC/XML:

```
<marc:leader>00775nam a2200217 i 4500</marc:leader>
```

RDF/XML:

```
<bf:status>  
  <bf:Status>  
    <rdfs:label>new</rdfs:label>  
    <bf:code>n</bf:code>  
  </bf:Status>  
</bf:status>
```

A rekordfej **6. karakterpozíciója** határozza meg a leírt bibliográfiai forrás jellegét, avagy a rekord típusát. Az itt található „a” érték alapján minden esetben az alábbi metaadatközlés jelenik meg a gráfban a Work entitásszintjén:

MARC/XML:

```
<marc:leader>00775nam a2200217 i 4500</marc:leader>
```

RDF/XML:

```
<rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Text" />
```

A rekordfej **7. karakterpozíciója** a bibliográfiai szintet adja meg: az itt található „m” érték alapján minden esetben az alábbi metaadatközlés jelenik meg a gráfban a Work entitásszintjén:

MARC/XML:

```
<marc:leader>00775nam a2200217 i 4500</marc:leader>
```

RDF/XML:

```
<bf:issuance>  
  <bf:Issuance rdf:about="http://id.loc.gov/vocabulary/issuance/mono" />  
</bf:issuance>
```

a Work-höz kapcsolódó Instance entitásszintjén pedig az alábbi közlés található:

```
<rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Print" />
```

A leírás jellege (**17. karakterpozíció**) az alábbi metaadatközlések formájában olvasható. Az egyes értékekhez automatikusan társul a Library of Congress szótárából a megfelelő URI, és címke is keletkezik a megjeleníthetőség érdekében. A PIM rekordjaiban a 17. karakterpozíción nem található karakter, amely a „teljes, autopsziával” szint megfelelője.

MARC/XML:

```
<marc:leader>00775nam a2200217 i 4500</marc:leader>
```

RDF/XML:

```
<bflc:encodingLevel>  
  <bflc:EncodingLevel rdf:about="http://id.loc.gov/vocabulary/menc1vl/ᄁ">  
    <rdfs:label>full</rdfs:label>  
  </bflc:EncodingLevel>  
</bflc:encodingLevel>
```

A **18. karakterpozíción** a bibliográfiai leírás szabályának meghatározása történik. Minden esetben „i” érték szerepel itt, amely az alábbi módon kódolódik, szótárból származó URI-val, illetve megjelenítési címkével:

MARC/XML:

```
<marc:leader>00775nam a2200217 i 4500</marc:leader>
```

RDF/XML:

```
<bf:descriptionConventions>  
  <bf:DescriptionConventions  
  <rdf:about="http://id.loc.gov/vocabulary/descriptionConventions/isbd">  
    <rdfs:label>isbd</rdfs:label>  
  </bf:DescriptionConventions>  
</bf:descriptionConventions>
```

001 - Rekordazonosító

A rekordazonosító helyi, egyedi azonosítónak számít, megjelenítése a gráfban az alábbiak szerint történik. Az azonosítót kibocsátó ágens URI-ja a Library of Congress-re utal, ezért azt a konverzió intézményi tervezésekor módosítani kell a *ConvSpec-001-007.xml* stíluslapban, ahol ennek az URI-nak a kezelése összefügg a 003-as mezőben tárolt értékkel (vagy a mező hiányával). 003-as mezőt a HUNMARC nem definiál.

MARC/XML:

```
<marc:controlfield tag="001">PIM2255316</marc:controlfield>
```

RDF/XML:

```
<bf:identifiedBy>  
  <bf:Local>  
    <rdf:value>PIM2255316</rdf:value>  
  <bf:assigner>
```

```
<bf:Agent rdf:about="http://id.loc.gov/vocabulary/organizations/dlc" />
  </bf:assigner>
  </bf:Local>
</bf:identifiedBy>
```

002 – A leíró űrlap típusa

Egyedileg definiált adatmező (a kísérleti adathalmaz rekordjai esetében értéke mindig *book*), amely a kimeneti RDF/XML-ben nem jelenik meg. Redundáns mivolta miatt konverzióját nem tartom szükségesnek.

MARC/XML:

```
<marc:controlfield tag="002">book</marc:controlfield>
```

005 – A rekorddal végzett utolsó művelet dátuma és időpontja

Az adatmezőben található érték konverziója helyesen megtörténik.

MARC/XML:

```
<marc:controlfield tag="005">20190306143016.0</marc:controlfield>
```

RDF/XML:

```
<bf:changeDate
  rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2019-03-06T14:30:16
</bf:changeDate>
```

008 – Meghatározott jellemzők és információs adatok

A **0-5. karakterpozíció értékeiből** (az adatbázisba kerülés dátuma) az alábbi kódrészlet generálódik a Work entitásszintjén:

MARC/XML:

```
<marc:controlfield tag="008">190221s2003#### 1</marc:controlfield>
```

RDF/XML:

```
<bf:creationDate
  rdf:datatype="http://www.w3.org/2001/XMLSchema#date">2019-02-21
</bf:creationDate>
```

A **6. karakterpozíció „s”** – HUNMARC: megjelenési év (tényleges vagy feltételezett); MARC 21: single known/probable date – értéke esetén az alábbi kódrészlet jön létre a Work-höz kapcsolódó Instance leírásában:

MARC/XML:

```
<marc:controlfield tag="008">190221s2003#### 1</marc:controlfield>
```

RDF/XML:

```
<bf:provisionActivity>
  <bf:ProvisionActivity>
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Publication" />
    <bf:date rdf:datatype="http://id.loc.gov/datatypes/edtf">2003</bf:date>
    <bf:place>
      <bf:Place rdf:about="http://id.loc.gov/vocabulary/countries/" />
    </bf:place>
```

```
</bf:ProvisionActivity>
</bf:provisionActivity>
```

A félkövérrel jelölt sorban található érték a 07-10. karakterpozíciókról származik. A megjelenési hely adata, amennyiben ki van töltve, a 008-as adatmező 15-17. pozíciójáról olvasható le.

A dokumentum nyelvét a 008-as adatmező **35-37. karakterpozíciója** rögzíti, s ha ezek a pozíciók ki vannak töltve, annak megjelenítése az RDF/XML-ben a Work entitásszintjén történik, rögtön az adminisztrációs metaadatok, illetve a bibliográfiai forrás jellegének rögzítése után:

MARC/XML:

```
<marc:controlfield tag="008">181108s2016#### gre 1</marc:controlfield>
```

RDF/XML:

```
<bf:language>
  <bf:Language rdf:about="http://id.loc.gov/vocabulary/languages/gre" />
</bf:language>
```

Többynelvű kiadványok esetén az alábbi kódrészlet generálódik:

```
<bf:language>
  <bf:Language rdf:about="http://id.loc.gov/vocabulary/languages/mul" />
</bf:language>
```

A **39. karakterpozíció** értékét, amely a HUNMARC alapján a rekord eredetéről tájékoztat, a MARC 21-re épülő konverziós specifikációk nem dolgozzák fel. Amennyiben mégis megtennék, a kódrészletet mindenképp módosítani kellene, hiszen a MARC 21-ben nem fordulhat elő a 39. pozíción az egyes érték, amely a tesztalmazban található.

011 – Egyedileg definiált

Egyedi definiálású mező a PIM adatbázisában, amely egy betöltési művelet hibaüzenetét tárolja. A tesztalmazban elhanyagolható mennyiségben, három alkalommal fordult elő, így átalakításával nem foglalkozom.

MARC/XML:

```
<marc:datafield tag="011" ind1=" " ind2=" ">
<marc:subfield code="a">batch upload error: validation error</marc:subfield>
<marc:subfield code="c">2015. 10. 01. 16:15</marc:subfield>
</marc:datafield>
```

020 - ISBN

Az adatmező értéke az Instance entitásszintjén kódolódik a `bf:identifiedBy` tulajdonság, illetve az azzal használható osztályok egyike, a `bf:Isbn` segítségével. A konkrét értéket a `bf:Isbn` egy tulajdonsága tartalmazza.

MARC/XML:

```
<marc:datafield tag="020" ind1=" " ind2=" ">
<marc:subfield code="a">978-963-12-8397-6</marc:subfield>
</marc:datafield>
```

RDF/XML:

```
<bf:identifiedBy>
  <bf:Isbn>
    <rdf:value>978-963-12-8397-6</rdf:value>
  </bf:Isbn>
</bf:identifiedBy>
```

022 – ISSN

Az adatmező értéke – az ISBN-nel ellentétben – a Work entitásszintjén kódolódik a `bf:identifiedBy` tulajdonság, illetve a `bf:Issn` osztály segítségével. A MARC-ból származó értéket a `bf:Issn` egy tulajdonsága tartalmazza. Látható, hogy az `identifiedBy` tulajdonság egyaránt használható a `bf:Isbn`, illetve a `bf:Issn` osztály tagjaival (és még számos másikkal, ez a különböző típusú azonosítók transzformációjának általános sémája).

MARC/XML:

```
<marc:datafield tag="022" ind1=" " ind2=" ">
<marc:subfield code="a">1416-5368</marc:subfield>
</marc:datafield>
```

RDF/XML:

```
<bf:identifiedBy>
  <bf:Issn>
    <rdf:value>1416-5368</rdf:value>
  </bf:Issn>
</bf:identifiedBy>
```

040 – A leírás forrása

Az eredeti leíró intézmény (\$a), illetve a leírás nyelve (\$b) a Work entitásszintjén, az adminisztrációs metaadatok között kódolódik, az alábbi módon:

MARC/XML:

```
<marc:datafield tag="040" ind1=" " ind2=" ">
<marc:subfield code="a">PIM</marc:subfield>
<marc:subfield code="b">hun</marc:subfield>
</marc:datafield>
```

RDF/XML:

```
<bf:source>
  <bf:Source>
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Agent" />
    <rdfs:label>PIM</rdfs:label>
  </bf:Source>
```

```

</bf:source>

<bf:descriptionLanguage>
  <bf:Language rdf:about="http://id.loc.gov/vocabulary/languages/hun">
    <bf:code>hun</bf:code>
  </bf:Language>
</bf:descriptionLanguage>

```

amelyből kiolvasható, hogy a `bf:source` tulajdonság megköveteli, hogy azt a `bf:Source` osztály tagjaival használják; jelen konkrét előforduláshoz két metaadatközlés csatlakozik, az egyik egy tipizálás, a másik egy címke. Hasonlóan képződik le a másik almező értéke is a gráfban: a tulajdonság használatához a `Language` osztály valamely tagjára van szükség, s ezt egy kötött szótárkódolási sémából vett URI azonosítja, a nyelvkód pedig literál formájában öröklődik az RDF-struktúrába.

A `$c` almező (géppel olvasható formába átiró intézmény) kezelése nem tér el a `$a` almező átalakításától, a `bf:source` tulajdonság segítségével történik. Eltérő relációt kap azonban a `$d` almezőben tárolt módosító intézmény:

MARC/XML:

```

<marc:datafield tag="040" ind1=" " ind2=" ">
<marc:subfield code="a">PIM</marc:subfield>
<marc:subfield code="b">hun</marc:subfield>
<marc:subfield code="c">OSZK</marc:subfield>
<marc:subfield code="d">PIM</marc:subfield>
</marc:datafield>

```

RDF/XML:

```

<bf:source>
  <bf:Source>
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Agent" />
    <rdfs:label>PIM</rdfs:label>
  </bf:Source>
</bf:source>
<bf:source>
  <bf:Source>
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Agent" />
    <rdfs:label>OSZK</rdfs:label>
  </bf:Source>
</bf:source>
<bf:descriptionLanguage>
  <bf:Language rdf:about="http://id.loc.gov/vocabulary/languages/hun">
    <bf:code>hun</bf:code>
  </bf:Language>
</bf:descriptionLanguage>
<bf:descriptionModifier>
  <bf:Agent>
    <rdfs:label>PIM</rdfs:label>
  </bf:Agent>
</bf:descriptionModifier>

```

041 – Nyelvkód

Értékének konvertált formája közvetlenül a 008-as adatmező karakterpozíciójának kódolása után foglal helyet:

MARC/XML:

```

<marc:datafield tag="041" ind1="1" ind2=" ">

```

```
<marc:subfield code="a">hungerrum</marc:subfield>
</marc:datafield>
```

RDF/XML:

```
<bf:language>
  <bf:Language rdf:about="http://id.loc.gov/vocabulary/languages/hun" />
</bf:language>

<bf:language>
  <bf:Language>
    <rdf:value rdf:resource="http://id.loc.gov/vocabulary/languages/hun" />
  </bf:Language>
</bf:language>

<bf:language>
  <bf:Language>
    <rdf:value rdf:resource="http://id.loc.gov/vocabulary/languages/ger" />
  </bf:Language>
</bf:language>

<bf:language>
  <bf:Language>
    <rdf:value rdf:resource="http://id.loc.gov/vocabulary/languages/rum" />
  </bf:Language>
</bf:language>
```

A magyar nyelvről szóló metaadatközlés azért ismétlődik, mert az egyiknek a 008 karakterpozíciói, a másikkal pedig a 041-es adatmező a forrása. Az első indikátor 0 értéke esetén további információ nem jelenik meg; egyes indikátorérték esetében azonban egy további metaadatközlés generálódik:

RDF/XML:

```
<bf:note>
  <bf:Note>
    <rdfs:label>Includes translation</rdfs:label>
  </bf:Note>
</bf:note>
```

A címke magyarázható a *ConvSpec-010-048.xml* alábbi szakaszának átírásával:

```
<xsl:if test="@ind1 = '1'">
  <bf:note>
    <bf:Note>
      <rdfs:label>Fordítást tartalmaz</rdfs:label>
    </bf:Note>
  </bf:note>
</xsl:if>
```

amely szakasz tehát megvizsgálja a 041-es adatmező összes indikátor-előfordulását, s ahol egyest talál, létrehozza a megjegyzésre vonatkozó metaadatközlést.

044 – Kiadó/előállító ország kódja

A mező konverziója nem történik meg. (A PIM adathalmazában előfordulása minimális, 3 rekordot érint).

MARC/XML:

```
<marc:datafield tag="044" ind1=" " ind2=" ">
<marc:subfield code="a">hu</marc:subfield>
</marc:datafield>
```

080 – Egyetemes Tizedes Osztályozás jelzete

A mező konverziója nem történik meg. Ez annak fényében különös, hogy a BIBFRAME tartalmazza az ehhez szükséges `bf:ClassificationUdc` alosztályt, amely a `bf:Classification` alárendeltje. Ezek értékeivel vehető használatba a `bf:classification` tulajdonság. A keretrendszer e tulajdonságainak, illetve osztályainak felhasználásával új szakasz írható a stíluslapba (`ConvSpec-050-088.xsl`), amelynek mintájául a 082-es mező kezelésével foglalkozó kódrészlet használható. (A PIM-ben a mező használata nem szokás, mindössze két rekordban sikerült megtalálni.)

MARC/XML:

```
<marc:datafield tag="080" ind1=" " ind2=" ">
<marc:subfield code="a">800.87:809.451.1(038)</marc:subfield>
</marc:datafield>
```

090 – Egvedileg definiált

A HUNMARC (sőt a MARC 21) által elavultnak jelölt adatmező, mely azonban egyedi értelmezésben érvényben maradt a PIM rekordjaiban: egy hárombetűs kód kerül ide, mely azt a múzeumi gyűjteményt vagy gyűjteményrészt, esetleg bibliográfiai rész-adatbázist azonosítja, amelyben a leírt műtárgy megtalálható. A könyvtári rekordok esetében ez a kód a **KKV**. Konverziójával értelemszerűen nem számol az eredeti specifikáció, így megjelenítéséhez egy új szakaszt kell definiálni a stíluslapban, amelyben a kívánt érték a `bf:Local` osztály tagjaként jelenik meg, amely a `bf:Identifier` alosztálya. E két osztály tagjai adják az értékét a `bf:identifiedBy` tulajdonságnak, amelynek értelmezési tartománya azonban nincsen szorosabban megköthet⁵⁴: éppen ezért a stíluslapot úgy kell elkészíteni, hogy a MARC-rekord 090-es adatmezőjének értékével az Item entitásszinten elhelyezkedő metaadatközléseket hozzon létre.

MARC/XML:

```
<marc:datafield tag="090" ind1=" " ind2=" ">
<marc:subfield code="a">KKV</marc:subfield>
<marc:subfield code="a">LEK</marc:subfield></marc:datafield>
```

RDF/XML (javasolt)

```
<bf:identifiedBy>
  <bf:Local>
    <rdf:value>KKV</rdf:value>
  </bf:Local>
</bf:identifiedBy>
<bf:identifiedBy>
  <bf:Local>
    <rdf:value>LEK</rdf:value>
  </bf:Local>
```

⁵⁴ Lásd a tulajdonság adatainak összefoglalóját: https://id.loc.gov/ontologies/bibframe.html#p_identifiedBy

</bf:identifiedBy>

100 – Személynév típusú kiemelt hozzáférési pont

A kimeneti RDF/XML-file-t vizsgálva szinte rögtön kiugrik egy nagyon jelentős probléma, amely a hozzáférési pontok kezelésében, illetve az automatikus URI-generálásban rejlik.

MARC/XML:

```
<marc:datafield tag="100" ind1="1" ind2="0">
<marc:subfield code="a">Gyémánt</marc:subfield>
<marc:subfield code="j">László</marc:subfield>
<marc:subfield code="d">1935</marc:subfield>
</marc:datafield>
```

RDF/XML

```
<bf:contribution>
  <bf:Contribution>
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bflc/PrimaryContribution" />
    <bf:agent>
      <bf:Agent rdf:about="C:\Users\Hubay
Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-2020.xmlPIM2255316#Agent100-8">
        <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Person" />
        <bflc:name00MatchKey>Gyémánt László 1935</bflc:name00MatchKey>
        <bflc:primaryContributorName00MatchKey>Gyémánt László
1935</bflc:primaryContributorName00MatchKey>
        <bflc:name00MarcKey>10010$aGyémánt$jLászló$d1935</bflc:name00MarcKey>
        <rdfs:label>Gyémánt László 1935</rdfs:label>
      </bf:Agent>
    </bf:agent>
    <bf:role>
      <bf:Role rdf:about="http://id.loc.gov/vocabulary/relators/ctb" />
    </bf:role>
  </bf:Contribution>
</bf:contribution>
```

Látszik, hogy a kiindulási adatforrásként felhasznált 100-as adatmező alapján a konverter helyesen állapította meg, hogy kiemelt hozzáférési pontról van szó, helyesen sorolta a közreműködést (`bf:Contribution`) az elsődleges közreműködések osztályába (`bf:PrimaryContribution`), illetve ugyancsak helyesen állapította meg az ágens típusát (`bf:Person`). A fő problémát az jelenti, hogy a HUNMARC-ban nincsen definiálva olyan almező, amely a 100-as adatmezőbe beemelt személy URI-ját tárolhatná: sem \$0, sem \$1 nem áll rendelkezésre. Ebből pedig az OSZK-projektnél már ismertetett probléma adódik: ugyanazon személynek minden esetben eltérő URI-ja fog képződni, így a visszakereséskor téves eredmények keletkeznek.

A Library of Congress által definiált relációk helyesen adják vissza a 100-as adatmező almezőinek értékeit, és egyéb információ hiányában a közreműködői funkcióhoz is az általános, közreműködői viszonyt azonosító <http://id.loc.gov/vocabulary/relators/ctb> URI rendelődik.

A konverziós eljárás ellenőrzése céljából kézzel illesztettem be egy \$0 almezőt az érintett rekord 100-as adatmezőjébe a kiinduló MARC/XML-állományban – ez az URI a Petőfi Irodalmi Múzeum személynévteréből származott –, majd ismét lefuttattam a konverziót. Az eredmény úgy módosult, hogy a bibliográfiai rekord azonosítójával mechanikusan generált ágens-URI-t felváltotta a \$0 almezőben található érték:

```
<bf:Agent rdf:about="https://resolver.pim.hu/auth/PIM94986">
```

ami tehát igazolta, hogy a bibliográfiai rekordokban előforduló hozzáféréspont-adatmezők \$0 almezőjének MARC-ba rögzítésével lehetségessé válik az adott entitás-előfordulás valamennyi megjelenésekor azonos URI-t képezni.

A megoldás optimalizálható a konverziós eszköz módosításával, mégpedig a *ConvSpec-ControlSubfields.xsl* stíluslap átírásával, amely a MARCEdit telepítésekor kerül a felhasználó számítógépére. Ez a stíluslap gondoskodik a művek, illetve az ágensek URI-jainak generálásáról, és itt paraméterezhető, hogy a konverzió során az automatikus folyamat mely almezőben keresse az entitás-előfordulások URI-jait.

Az ágensekkel kapcsolatban feltűnhet továbbá, hogy a \$m almező (rokonsági kiegészítő) értéke nem jelenik meg a konvertált adathalmaz több relációjában: ezt az almezőt a MARC 21 nem definiálja. Ugyancsak feltűnő a \$g almező értékének (állandósult melléknév az autorizált névalak egyéb elemeként) hiánya a kimeneti adatszerkezetben. A \$c almező értéke azonban megtalálható:

MARC/XML

```
<marc:datafield tag="100" ind1="1" ind2="0">  
<marc:subfield code="a">Mitrovics</marc:subfield>  
<marc:subfield code="j">Gyula</marc:subfield>  
<marc:subfield code="m">ifj</marc:subfield>  
<marc:subfield code="d">1871-1965</marc:subfield>  
</marc:datafield>
```

RDF/XML:

```
<bflc:name00MatchKey>Mitrovics Gyula 1871-1965</bflc:name00MatchKey>  
<bflc:primaryContributorName00MatchKey>Mitrovics Gyula 1871-  
1965</bflc:primaryContributorName00MatchKey>  
<bflc:name00MarcKey>10010$aMitrovics$jGyula$mifj$d1871-1965</bflc:name00MarcKey>  
<rdfs:label>Mitrovics Gyula 1871-1965</rdfs:label>
```

MARC/XML 2:

```
<marc:datafield tag="100" ind1="1" ind2="0">  
<marc:subfield code="a">Krúdy</marc:subfield>  
<marc:subfield code="j">Gyula</marc:subfield>  
<marc:subfield code="d">1878-1933</marc:subfield>  
<marc:subfield code="c">nemes</marc:subfield>  
<marc:subfield code="g">széchénykovácsi</marc:subfield>  
</marc:datafield>
```

RDF/XML 2:

```
<bflc:name00MatchKey>Krúdy Gyula 1878-1933 nemes</bflc:name00MatchKey>
```

```

<bflc:primaryContributorName00MatchKey>Krúdy Gyula 1878-1933
nemes</bflc:primaryContributorName00MatchKey>
<bflc:name00MarcKey>10010$aKrúdy$JGyula$d1878-
1933$cnemes$gszéchenykovácsi</bflc:name00MarcKey>
<rdfs:label>Krúdy Gyula 1878-1933 nemes</rdfs:label>

```

A stíluslap vonatkozó része tehát úgy módosítandó, hogy a szükséges almezők (\$m, illetve \$g) értéke megjelenjen a vonatkozó relációkban, legfőképpen pedig a felhasználó számára is látható `rdfs:label` értékében.

110 – Testületi név típusú kiemelt hozzáférési pont

Az URI generálásával kapcsolatos következtetések megegyeznek a 100-as mezőnél leírtakkal. Az ágens típusa a 110-es mező miatt: Organization, a közreműködés típusa PrimaryContribution.

MARC/XML:

```

<marc:datafield tag="110" ind1="2" ind2="0">
<marc:subfield code="a">Auswärtiges Amt</marc:subfield>
</marc:datafield>

```

RDF/XML:

```

<bf:contribution>
  <bf:Contribution>
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bflc/PrimaryContribution" />
    <bf:agent>
      <bf:Agent rdf:about="C:\Users\Hubay
Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-2020.xmlPIM2045204#Agent110-8">
      <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Organization" />
      <bflc:name10MatchKey>Auswärtiges Amt</bflc:name10MatchKey>
      <bflc:name10MarcKey>11020$aAuswärtiges Amt</bflc:name10MarcKey>
      <bflc:primaryContributorName10MatchKey>Auswärtiges
Amt</bflc:primaryContributorName10MatchKey>
      <rdfs:label>Auswärtiges Amt</rdfs:label>
    </bf:Agent>
    </bf:agent>
    <bf:role>
      <bf:Role rdf:about="http://id.loc.gov/vocabulary/relators/ctb" />
    </bf:role>
    </bf:Contribution>
  </bf:contribution>

```

240 – Preferált cím típusú további hozzáférési pont

Értéke a Work entitásszintjén megjelenő címet adja: a MARC-ban történő helyes és körültekintő kitöltés segíthet a használatnak a művek elkülönítésében. A 240\$b almezőből nem képződik külön metaadatközlés `bf:subtitle` relációval, bár értéke a Library of Congress által bevezetett `bflc:title40MarcKey`-ben megjelenik.

MARC/XML:

```

<marc:datafield tag="240" ind1="1" ind2="0">
<marc:subfield code="a">Men against death</marc:subfield>
</marc:datafield>

```

MARC/XML 2:

```

<marc:datafield tag="240" ind1="1" ind2="0">
<marc:subfield code="a">Ha álom az élet</marc:subfield>

```

```
<marc:subfield code="b">válogatott versek</marc:subfield>
</marc:datafield>
```

RDF/XML:

```
<rdfs:label>Men against death</rdfs:label>
  <bf:title>
    <bf:Title>
      <bflc:title40MatchKey>Men against death</bflc:title40MatchKey>
      <bflc:title40MarcKey>24010$aMen against death</bflc:title40MarcKey>
      <rdfs:label>Men against death</rdfs:label>
      <bflc:titleSortKey>Men against death</bflc:titleSortKey>
      <bf:mainTitle>Men against death</bf:mainTitle>
    </bf:Title>
  </bf:title>
```

RDF/XML 2:

```
<rdfs:label>Ha álom az élet</rdfs:label>
  <bf:title>
    <bf:Title>
      <bflc:title40MatchKey>Ha álom az élet</bflc:title40MatchKey>
      <bflc:title40MarcKey>24010$aHa álom az élet$bválogatott
versek</bflc:title40MarcKey>
      <rdfs:label>Ha álom az élet</rdfs:label>
      <bflc:titleSortKey>Ha álom az élet</bflc:titleSortKey>
      <bf:mainTitle>Ha álom az élet</bf:mainTitle>
    </bf:Title>
  </bf:title>
```

245 – Cím és felelősségi közlés

Amennyiben nem áll rendelkezésre egységesített cím sem a 130-as, sem a 240-es adatmezőből, a 245-ös adatmező \$a almezőjének értéke szolgál a Work entitásszinten történő azonosításra is:

MARC/XML:

```
<marc:datafield tag="245" ind1="1" ind2="0">
<marc:subfield code="a">Ágas-bogas fa</marc:subfield>
<marc:subfield code="b">néprajzi ismeretek</marc:subfield>
<marc:subfield code="c">Liszka József</marc:subfield>
<marc:subfield code="e">Fotó Liszka József, Deutsch Zoltán</marc:subfield>
<marc:subfield code="e">ill. Thain János</marc:subfield></marc:datafield>
```

Work entitásszinten csak a 245\$a értéke látható:

RDF/XML – Work entitásszint

```
rdfs:label>Ágas-bogas fa</rdfs:label>
  <bf:title>
    <bf:Title>
      <rdfs:label>Ágas-bogas fa</rdfs:label>
      <bflc:titleSortKey>Ágas-bogas fa</bflc:titleSortKey>
      <bf:mainTitle>Ágas-bogas fa</bf:mainTitle>
    </bf:Title>
  </bf:title>
```

Instance entitásszinten azonban megjelenik a \$b, illetve \$c almezők értéke is:

RDF/XML – Instance entitásszint

```
<rdfs:label>Ágas-bogas fa</rdfs:label>
  <bf:title>
    <bf:Title>
      <rdfs:label>Ágas-bogas fa</rdfs:label>
      <bflc:titleSortKey>Ágas-bogas fa</bflc:titleSortKey>
      <bf:mainTitle>Ágas-bogas fa</bf:mainTitle>
      <bf:subtitle>néprajzi ismeretek</bf:subtitle>
    </bf:Title>
```

```
</bf:title>
<bf:responsibilityStatement>Liszka József</bf:responsibilityStatement>
```

Amennyiben kihagyandó karaktereket jelöl a második indikátor értéke, a kimenet is ennek megfelelően módosul:

MARC/XML:

```
<marc:datafield tag="245" ind1="1" ind2="4">
<marc:subfield code="a">Der ägyptische Schreiber</marc:subfield>
<marc:subfield code="b">Erzählungen</marc:subfield>
<marc:subfield code="c">Lajos Nagy</marc:subfield>
<marc:subfield code="e">[Übert., Nachbemerck. Vera Thies]</marc:subfield>
</marc:datafield>
```

RDF/XML:

```
<rdfs:label>Der ägyptische Schreiber</rdfs:label>
  <bf:title>
    <bf:Title>
      <rdfs:label>Der ägyptische Schreiber</rdfs:label>
      <bflc:titleSortKey>ägyptische Schreiber</bflc:titleSortKey>
      <bf:mainTitle>Der ägyptische Schreiber</bf:mainTitle>
      <bf:subtitle>Erzählungen</bf:subtitle>
    </bf:Title>
  </bf:title>
<bf:responsibilityStatement>Lajos Nagy</bf:responsibilityStatement>
```

A \$e értékét (második és minden további felelősségi közlés) hiába keresnénk a kimenetben, a MARC 21 nem definiál ilyen almezőt; minden felelősségi funkciót a \$c-ben tesz felsorolhatóvá. A stíluslap ebben az esetben úgy módosítandó, hogy a bf:responsibilityStatement értékei ne csak a \$c, hanem a \$e almezőből is származhassanak, így felelősségi közlésenként egy metaadat-közlés keletkezik.

A konverziós folyamat nem tesz különbséget az első indikátor értékei között: ugyanaz az eljárás fut le akkor is, ha a főcím kitüntetett, illetve akkor is, ha további hozzáférési pont.

250 – Kiadás

Értéke az alábbi formában kódolódik, literálként, Instance entitásszinten:

MARC/XML:

```
<marc:datafield tag="250" ind1=" " ind2=" ">
<marc:subfield code="a">3. kiad</marc:subfield>
</marc:datafield>
```

RDF/XML:

```
<bf:editionStatement>3. kiad</bf:editionStatement>
```

259 – Egyedileg definiált

Adatbázis-műveletek végrehajtásakor keletkezett, egyedi tartalmú adatmező, a megjelenés évét tartalmazza. Új információt nem tartalmaz, ezért konverziója nem szükséges.

MARC/XML:

```
<marc:datafield tag="259" ind1=" " ind2=" " >
<marc:subfield code="q">2014</marc:subfield>
<marc:subfield code="Q">2014</marc:subfield>
</marc:datafield>
```

260 – Megjelenés

Értéke Instance-szinten jelenik meg. Különálló osztályok és tulajdonságok segítségével adatelemenként is átöröklődnek az adatok (bf:place, bf:agent, bf:date), de a teljes adatsort is megőrzi literálként a bf:provisionActivityStatement reláció. A közreműködői funkciók adatai (kiadó, előállító) minősítőkkal vannak ellátva (Publication, illetve Manufacture, mindkettő a Library of Congress kötött szótárából származó URI-ként jelenik meg).

MARC/XML:

```
<marc:datafield tag="260" ind1=" " ind2=" " >
<marc:subfield code="a">[Budapest]</marc:subfield>
<marc:subfield code="b">Magyar Helikon</marc:subfield>
<marc:subfield code="c">1963</marc:subfield>
<marc:subfield code="e">[Gyoma]</marc:subfield>
<marc:subfield code="f">Gyomai ny</marc:subfield></marc:datafield>
```

RDF/XML:

```
<bf:provisionActivity>
  <bf:ProvisionActivity>
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Publication" />
    <bf:place>
      <bf:Place>
        <rdfs:label>Budapest</rdfs:label>
      </bf:Place>
    </bf:place>
    <bf:agent>
      <bf:Agent>
        <rdfs:label>Magyar Helikon</rdfs:label>
      </bf:Agent>
    </bf:agent>
    <bf:date>1963</bf:date>
  </bf:ProvisionActivity>
</bf:provisionActivity>
<bf:provisionActivityStatement>[Budapest]; Magyar Helikon;
1963</bf:provisionActivityStatement>
<bf:provisionActivity>
  <bf:ProvisionActivity>
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Manufacture" />
    <bf:place>
      <bf:Place>
        <rdfs:label>[Gyoma]</rdfs:label>
      </bf:Place>
    </bf:place>
    <bf:agent>
      <bf:Agent>
        <rdfs:label>Gyomai ny</rdfs:label>
      </bf:Agent>
    </bf:agent>
  </bf:ProvisionActivity>
</bf:provisionActivity>
```

300 – Terjedelem/Fizikai jellemzők

A Terjedelem adatcsoport három alapvető elemét tartalmazó kódrészlet konverziója az alábbiak szerint történik:

MARC/XML:

```
<marc:datafield tag="300" ind1=" " ind2=" ">
<marc:subfield code="a">212 p</marc:subfield>
<marc:subfield code="b">i11</marc:subfield>
<marc:subfield code="c">23 cm</marc:subfield>
</marc:datafield>
```

RDF/XML:

```
<bf:extent>
  <bf:Extent>
    <rdfs:label>212 p</rdfs:label>
  </bf:Extent>
</bf:extent>
<bf:note>
  <bf:Note>
    <bf:noteType>Physical details</bf:noteType>
    <rdfs:label>i11</rdfs:label>
  </bf:Note>
</bf:note>
<bf:dimensions>23 cm</bf:dimensions>
```

A melléklet \$e almezője külön megjegyzéstípust kap:

```
<bf:note>
  <bf:Note>
    <bf:noteType>Accompanying materials</bf:noteType>
    <rdfs:label>CD</rdfs:label>
  </bf:Note>
</bf:note>
```

A címkék magyaríthatók a stíluslap (*ConvSpec-3XX.xsl*) megfelelő szakaszának átírásával:

```
<xsl:for-each select="marc:subfield[@code='b' or @code='e']">
  <bf:note>
    <bf:Note>
      <bf:noteType>
        <xsl:choose>
          <xsl:when test="@code='b'">Fizikai jellemzők</xsl:when>
          <xsl:when test="@code='e'">Melléklet típusa</xsl:when>
        </xsl:choose>
      </bf:noteType>
    </bf:Note>
  </bf:note>
```

A HUNMARC-ban meglévő \$h, \$i, illetve \$k almezők értékei nem szerepelnek az RDF-kimenetben. Ezek az almezők nincsenek definiálva a MARC 21-ben, értékeiket a \$e almező közléseihez fűzve adják meg. A stíluslapot érdemes ennek alapján úgy módosítani, hogy a \$h, \$i, illetve \$k almezők értékei a \$e-hez fűzve jelenjenek meg a kimeneti adatszerkezetben.

MARC/XML:

```
<marc:datafield tag="300" ind1=" " ind2=" ">
<marc:subfield code="a">380 p</marc:subfield>
<marc:subfield code="b">i11</marc:subfield>
<marc:subfield code="c">15x23 cm</marc:subfield>
<marc:subfield code="e">foglalkoztató füzet</marc:subfield>
<marc:subfield code="h">[20] p</marc:subfield>
<marc:subfield code="k">15x23 cm</marc:subfield>
</marc:datafield>
```

RDF/XML:

```
<bf:extent>
  <bf:Extent>
    <rdfs:label>380 p</rdfs:label>
  </bf:Extent>
</bf:extent>
<bf:note>
  <bf:Note>
    <bf:noteType>Physical details</bf:noteType>
    <rdfs:label>ill</rdfs:label>
  </bf:Note>
</bf:note>
<bf:note>
  <bf:Note>
    <bf:noteType>Accompanying materials</bf:noteType>
    <rdfs:label>foglalkoztató füzet</rdfs:label>
  </bf:Note>
</bf:note>
<bf:dimensions>15x23 cm</bf:dimensions>
```

362 – Számozás

A PIM-adathalmazban – mivel itt könyvjellegű bibliográfiai forrásokról van szó – ez az adat mindössze néhány esetben, évkönyvek rekordjaiban fordul elő. Konverziója ugyancsak megjegyzés segítségével történik, e megjegyzés típusának magyarítása az előzőekben ismertettek mintájára történhet.

MARC/XML:

```
<marc:datafield tag="362" ind1="1" ind2="" >
<marc:subfield code="a">26</marc:subfield>
</marc:datafield>
```

RDF/XML:

```
<bf:note>
  <bf:Note>
    <bf:noteType>Numbering</bf:noteType>
    <rdfs:label>26</rdfs:label>
  </bf:Note>
</bf:note>
```

440 – Sorozatcím további hozzáférési pontként

A 440-es adatmező adatelemeinek segítségével egy önálló BIBFRAME Work metaadatközlés-halmaz generálódik a sorozat (mint mű) adataiból. Ennek kialakításában a \$a almezőnek, illetve a \$x almezőnek van szerepe – utóbbi azonosítóként jelenik meg –, a teljes metaadatközlés-halmazt pedig a bf:hasSeries reláció köti a sorozat tagjának leírásához. A \$b, illetve \$u almezőt a HUNMARC nem definiálja, azokat mégis csaknem 900, illetve 2600 rekordban alkalmazzák. Utóbbi a Petőfi Irodalmi Múzeum rekordjaiban a sorozatot közreadó kiadó nevét tartalmazza, a BIBFRAME bf:agent tulajdonsága pedig megkötés hiányában bármilyen entitáshoz társítható, így nem lenne helytelen, ha a stíluslap megfelelő módosításával ágenst kapcsolnánk a sorozati adatok felhasználásával megszülető Work szintű metaadat-közléshalmazhoz.

MARC/XML:

```
<marc:datafield tag="440" ind1=" " ind2="0">
<marc:subfield code="a">Elvek és utak</marc:subfield>
<marc:subfield code="u">Magvető (Budapest)</marc:subfield>
<marc:subfield code="x">0324-4318</marc:subfield>
<marc:subfield code="z">Lackó Mihály</marc:subfield>
</marc:datafield>
```

MARC/XML 2:

```
<marc:datafield tag="440" ind1=" " ind2="0">
<marc:subfield code="a">Milliók könyve</marc:subfield>
<marc:subfield code="b">1-490</marc:subfield>
<marc:subfield code="u">Singer-Wolfner</marc:subfield>
<marc:subfield code="v">50</marc:subfield>
<marc:subfield code="z">Barta Antal</marc:subfield></marc:datafield>
```

RDF/XML – Work entitásszint

```
<bf:hasSeries>
  <bf:Work rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-
2020.xmlPIM2040948#Work440-12">
    <rdfs:label>Elvek és utak Magvető (Budapest Lackó Mihály)</rdfs:label>
    <bf:title>
      <bf:Title>
        <bflc:title40MarcKey>440 0$aElvek és utak$uMagvető (Budapest)$x0324-4318$zLackó
Mihály</bflc:title40MarcKey>
        <bf:mainTitle>Elvek és utak</bf:mainTitle>
      </bf:Title>
    </bf:title>
    <bf:identifiedBy>
      <bf:Issn>
        <rdf:value>0324-4318</rdf:value>
      </bf:Issn>
    </bf:identifiedBy>
  </bf:Work>
</bf:hasSeries>
```

RDF/XML 2. – Work entitásszint

```
<bf:hasSeries>
  <bf:Work rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-
2020.xmlPIM2130692#Work440-11">
    <rdfs:label>Milliók könyve 1-490 Singer-Wolfner Barta Antal</rdfs:label>
    <bf:title>
      <bf:Title>
        <bflc:title40MarcKey>440 0$aMilliók könyve$b1-490$uSinger-Wolfner$v66$zBarta
Antal</bflc:title40MarcKey>
        <bf:mainTitle>Milliók könyve</bf:mainTitle>
      </bf:Title>
    </bf:title>
  </bf:Work>
</bf:hasSeries>
```

Ha a bemeneti halmazban \$n (alárendelt címmegjelölés/ágazati megjelölés) vagy \$p (alárendelt cím/ágazati cím) almező van jelen, a kód az alábbiak szerint alakul:

MARC/XML

```
<marc:datafield tag="440" ind1=" " ind2="0">
<marc:subfield code="a">Nemzetnevelők könyvtár</marc:subfield>
<marc:subfield code="n">II</marc:subfield>
<marc:subfield code="p">Nemzetszolgálat</marc:subfield>
<marc:subfield code="v">2</marc:subfield></marc:datafield>
<bf:partNumber>II</bf:partNumber>
```

RDF/XML – Work entitásszint

```
<bf:hasSeries>
  <bf:Work rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-
2020.xmlPIM2100097#Work440-11">
    <rdfs:label>Nemzetnevelők könyvtár II Nemzetszolgálat</rdfs:label>
    <bf:title>
      <bf:Title>
        <bflc:title40MarcKey>440 0$aNemzetnevelők
könyvtár$nII$pNemzetszolgálat$v2</bflc:title40MarcKey>
      </bf:Title>
    </bf:title>
  </bf:Work>
</bf:hasSeries>
```

```

    <bf:mainTitle>Nemzetnevelők könyvtár</bf:mainTitle>
    <bf:partNumber>II</bf:partNumber>
    <bf:partName>Nemzetszolgálat</bf:partName>
  </bf:Title>
</bf:title>
</bf:Work>
</bf:hasSeries>

```

Instance entitásszinten két metaadatelem tartozik a sorozat-leíró elemek közé: a `bf:seriesStatement` literálként közli a 440-es adatmező almezőinek (kivéve a \$x és \$v) konkatenált tartalmát:

```
<bf:seriesStatement>Elvek és utak Magvető (Budapest Lackó Mihály)</bf:seriesStatement>
```

a sorozati számot pedig a `bf:seriesEnumeration`-nel fejezik ki – ez az adatelem a \$v almezőből származik:

```
<bf:seriesEnumeration>50</bf:seriesEnumeration>
```

Mivel a sorozatcímekhez nem tartozik saját URI, ezért a lokálisan generált azonosítók – akárcsak az ágensek esetében – a bibliográfiai rekord azonosítójából generálódnak, és így sorozattagonként eltérő URI-t fog kapni a sorozat. A megoldás, hasonlóan a 100-as adatmezőnél leírtakhoz, egy azonosító rögzítése lehet egy előre meghatározott almezőbe.

480 – Egvedileg definiált

A PIM gyűjteménykezelő rendszerében *Gyűjteménycím* címkével található adatmező. Konceptiója, illetve funkciója igen homályos, sok esetben sorozati adatokat tartalmaz, más esetekben többkötetes könyvek kötetcímét, illetve kötetjelzését. Konverzióját nem tartom relevánsnak.

MARC/XML

```

<marc:datafield tag="245" ind1="1" ind2="0">
<marc:subfield code="a">Arany László magyar népmese gyűjteménye</marc:subfield>
<marc:subfield code="e">[közread.], (előszó Gyulai Pál)</marc:subfield>
</marc:datafield>

```

```

<marc:datafield tag="480" ind1=" " ind2=" ">
<marc:subfield code="a">Arany László összes művei. 4. köt</marc:subfield>
</marc:datafield>

```

MARC/XML 2

```

<marc:datafield tag="245" ind1="1" ind2="2">
<marc:subfield code="a">A sárszentlőrinci orvos</marc:subfield>
<marc:subfield code="b">Balassa János</marc:subfield>
<marc:subfield code="c">Nádorné Gellért Anikó</marc:subfield>
<marc:subfield code="e">ill. Barkódyné Ötvös Marika</marc:subfield>
</marc:datafield>

```

```

<marc:datafield tag="480" ind1=" " ind2=" ">
<marc:subfield code="a">Történelmi kisregények</marc:subfield>
</marc:datafield>

```

490 – Sorozatcím a kiadványon szereplő formában

Ha egy rekordban 440-es, illetve 490-es mező is található, a Work entitásszintű adatok a 440-es alapján jönnek létre. A 490-es adatmező értéke Instance entitásszinten, literálként jelenik meg.

MARC/XML:

```
<marc:datafield tag="440" ind1=" " ind2="0">
<marc:subfield code="a">Vay Sándor munkái</marc:subfield>
<marc:subfield code="v">1</marc:subfield>
</marc:datafield>
<marc:datafield tag="490" ind1="0" ind2=" ">
<marc:subfield code="a">Gróf Vay Sándor munkái</marc:subfield>
</marc:datafield>
```

RDF/XML – Work entitásszint

```
<bf:hasSeries>
  <bf:Work rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-
2020.xmlPIM2051008#Work440-12">
    <rdfs:label>Vay Sándor munkái</rdfs:label>
    <bf:title>
      <bf:Title>
        <bflc:title40MarcKey>440 0$aVay Sándor munkái$vl</bflc:title40MarcKey>
        <bf:mainTitle>Vay Sándor munkái</bf:mainTitle>
      </bf:Title>
    </bf:title>
  </bf:Work>
</bf:hasSeries>
```

RDF/XML – Instance entitásszint

```
<bf:seriesStatement>Vay Sándor munkái</bf:seriesStatement>
<bf:seriesEnumeration>1</bf:seriesEnumeration>
<bf:seriesStatement>Gróf Vay Sándor munkái</bf:seriesStatement>
```

500 – Általános megjegyzések

Generálásuk a bf:note tulajdonság, illetve a bf:Note osztály segítségével történik.

MARC/XML:

```
<marc:datafield tag="500" ind1=" " ind2=" ">
<marc:subfield code="a">Reissmann Károly Miksa portréjával</marc:subfield>
</marc:datafield>
```

RDF/XML:

```
<bf:note>
  <bf:Note>
    <rdfs:label>Reissmann Károly Miksa portréjával</rdfs:label>
  </bf:Note>
</bf:note>
```

503 - Megjegyzés a mű/kiadvány bibliográfiai történetéről

A tesztalomban mindössze két alkalommal fordul elő „PIM dokumentum” értékkel. Relevanciája és ritkasága miatt konvertálásától eltekintettem.

MARC/XML:

```
<marc:datafield tag="503" ind1=" " ind2=" ">
<marc:subfield code="a">PIM dokumentum</marc:subfield>
```

</marc:datafield>

504 – Megjegyzés a rejtett bibliográfiáról

Értéke a `bf:supplementaryContent` reláció, illetve `bf:SupplementaryContent` osztály segítségével rögzíthető. A HUNMARC és a MARC 21 meződefiníciójához (*Bibliography, etc. Note*) képest a BIBFRAME-elem *kiegészítő tartalom* megnevezése kevésbé specifikus.

MARC/XML:

```
<marc:datafield tag="504" ind1=" " ind2=" ">
<marc:subfield code="a">Bibliogr.: 193-205. p</marc:subfield>
</marc:datafield>
```

RDF/XML:

```
<bf:supplementaryContent>
  <bf:SupplementaryContent>
    <rdfs:label>Bibliogr.: 193-205. p</rdfs:label>
  </bf:SupplementaryContent>
</bf:supplementaryContent>
```

505 - Megjegyzés a tartalomról

Értéke a `bf:tableOfContents` reláció, illetve az ezzel használható `bf:TableOfContents` osztály segítségével rögzíthető. Kisebb értelmezési zavart okozhat, hogy a BIBFRAME-elem itt a szűkebb *tartalomjegyzék* címkét viseli, szemben a két MARC-formátum *Megjegyzés a tartalomról*, illetve *Formatted Contents Note* mezőnevével. A lenti példában sem tartalomjegyzékre, hanem egyszerűen a tartalomra vonatkozó megjegyzés olvasható.

MARC/XML:

```
<marc:datafield tag="505" ind1=" " ind2=" ">
<marc:subfield code="a">Csengery Adrienn számára írta Kurtág György</marc:subfield>
</marc:datafield>
```

RDF/XML:

```
<bf:tableOfContents>
  <bf:TableOfContents>
    <rdfs:label>Csengery Adrienn számára írta Kurtág György</rdfs:label>
  </bf:TableOfContents>
</bf:tableOfContents>
```

562 – Megjegyzés a példány/verzió egyediségéről

Az 562-es megjegyzésmezőben a Petőfi Irodalmi Múzeum bibliográfiai rekordjai a példány/verzió egyediségéről szóló közléseket tárolnak, például a könyvekben olvasható dedikációkat. A Library of Congress által összeállított hivatalos konverziós specifikációban az 562-es MARC 21 adatmezőnél (*Copy and Version Identification Note*) a „nac” megjegyzés szerepel, amelynek jelentése „no attempt to convert”, azaz a mező konverzióját

nem végzik el. A PIM adathalmazának konverziójából azonban nem hiányozhat e speciális egyediségeket tartalmazó adatmező, ezért egy szakasszal ki kell bővíteni a *ConvSpec-5XX.xsl* állományt. A feladat: kiolvasni az 562\$a almező értékét, és azt az Item entitásszintjének leírásába helyezni, hiszen természetesen példányegediségekről van szó.

MARC/XML:

```
<marc:datafield tag="562" ind1=" " ind2=" " >
<marc:subfield code="a">Dedikáció: Csaplár Ferencnek, akit ahogy megismertem, nemsokkal
azután kezdtem dolgozni e kötet címadó tanulmányán. Barátsággal: Tasi József</marc:subfield>
</marc:datafield>
```

RDF/XML (javasolt):

```
<bf:note>
  <bf:Note>
    <rdfs:label>Dedikáció: Csaplár Ferencnek, akit ahogy megismertem, nemsokkal
kezdtem dolgozni e kötet címadó tanulmányán. Barátsággal: Tasi József</rdfs:label>
    <bf:noteType>dedikáció</bf:noteType>
  </bf:Note>
</bf:note>
```

588 – Megjegyzés a terjedelemről/fizikai jellemzőkről

Az 588-as adatmező kezelése a HUNMARC és a MARC 21 eltérő meződefiníciója miatt helytelenül megy végbe. A HUNMARC ezt az adatmezőt „*Megjegyzés a terjedelemről/fizikai jellemzőkről*” címkével látja el, míg MARC 21-ben „*Source of Description, Etc. Note*” megnevezés található. Emiatt a tesztalomból nyert RDF/XML-ben inkonzisztencia jelenik meg a megjelenített adat és annak típusa között.

MARC/XML:

```
<marc:datafield tag="588" ind1=" " ind2=" " >
<marc:subfield code="a">Kötés: eredeti, kiadói fvk. Fedélen: Venyecianov Zaharka c.
festménye</marc:subfield>
</marc:datafield>
```

RDF/XML:

```
<bf:note>
  <bf:Note>
    <rdfs:label>Kötés: eredeti, kiadói fvk. Fedélen: Venyecianov Zaharka c.
festménye</rdfs:label>
    <bf:noteType>description source</bf:noteType>
  </bf:Note>
</bf:note>
```

Szükséges tehát úgy módosítani a megjegyzés típusát leíró reláció létrehozási folyamatát, hogy az a megfelelő – kötéstípusra vonatkozó – megjegyzésekre utaljon. A megfelelő sor a *ConvSpec-5XX.xsl* stíluslapban található, ahol az 588-as mező előfordulása esetén a konverziós folyamat a „description source” értéket tünteti fel a megjegyzés típusaként. Ezt „kötés”-re módosítva máris a helyes `noteType` jelenik meg az RDF-állományban. A stíluslapon természetesen magyar nyelvű érték is rögzíthető.

```
<xsl:when test="$vTag='588'">kötés</xsl:when>
```

592 – Megjegyzés a különlenyomatról

A MARC 21 nem ismeri az 592-es mezőt. Ha tartalmát szükséges konvertálni, a stíluslapot úgy kell kiegészíteni, hogy az alább közölt struktúra keletkezzen a kimeneti állományban. A `bf:noteType` értéke szabadon megválasztható, értékeit nem szabályozza kötött szótár.

MARC/XML:

```
<marc:datafield tag="592" ind1=" " ind2=" " >
<marc:subfield code="a">Klny.: Irodalomtörténet, 1961. 2. sz</marc:subfield>
</marc:datafield>
```

RDF/XML (javasolt):

```
<bf:note>
  <bf:Note>
    <rdfs:label>Klny.: Irodalomtörténet, 1961. 2. sz</rdfs:label>
    <bf:noteType>különlenyomat</bf:noteType>
  </bf:Note>
</bf:note>
```

600 – Személynév típusú tárgyi hozzáférési pont

Szerkezete a felelősségi közléseknél látottakhoz hasonló, ám itt az ágensre vonatkozó kódrészletet a `bf:subject` reláció foglalja keretbe. Az URI-k automatikus generálásával, illetve egyes almezők értékének elvesztésével (\$g, \$m) összefüggő problémák itt is megjelennek, akárcsak a személynév típusú kiemelt hozzáférési pont esetében. A megoldás ebben az esetben is a következő: a HUNMARC-ot alapul vevő authority rekordokban egy új, egyedileg definiált almezőt kell felvenni, s ebbe rögzíteni az adott rekord rekordazonosítója segítségével képzett URI-t, amely minden kétséget kizáróan azonosítja az adott entitás-előfordulást, s gondoskodni kell az almező értékének beemeléséről a bibliográfiai rekordba. A választott almezőt – mely a MARC 21-et követve lehet \$0 vagy \$1, de lehet más is, ugyanakkor minden authority adatmező esetében azonosnak kell lennie – specifikálni kell a stíluslapban.

MARC/XML - \$m almezőt tartalmaz

```
<marc:datafield tag="600" ind1="1" ind2="4">
<marc:subfield code="a">Nagy</marc:subfield>
<marc:subfield code="j">Géza</marc:subfield>
<marc:subfield code="m">ifj</marc:subfield>
<marc:subfield code="d">1914-1981</marc:subfield>
</marc:datafield>
```

RDF/XML:

```
<bf:subject>
  <bf:Agent rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-2020.xmlPIM2101018#Agent600-13">
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Person" />
    <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#PersonalName" />
    <madsrdf:authoritativeLabel>Nagy Géza 1914-1981</madsrdf:authoritativeLabel>
    <bflc:name00MatchKey>Nagy Géza 1914-1981</bflc:name00MatchKey>
    <bflc:name00MarcKey>60014$aNagy$jGéza$mi$fj$d1914-1981</bflc:name00MarcKey>
    <rdfs:label>Nagy Géza 1914-1981</rdfs:label>
```

```
</bf:Agent>
</bf:subject>
```

MARC/XML 2 - \$c és \$g almezőt tartalmaz:

```
<marc:datafield tag="600" ind1="1" ind2="4">
<marc:subfield code="a">Batthyány</marc:subfield>
<marc:subfield code="j">Lajos</marc:subfield>
<marc:subfield code="d">1807-1849</marc:subfield>
<marc:subfield code="c">gr</marc:subfield>
<marc:subfield code="g">németújvári</marc:subfield>
</marc:datafield>
```

RDF/XML 2:

```
<bf:subject>
  <bf:Agent rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-
2020.xmlPIM2123021#Agent600-12">
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Person" />
    <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#PersonalName" />
    <madsrdf:authoritativeLabel>Batthyány Lajos 1807-1849
gr</madsrdf:authoritativeLabel>
    <bflc:name00MatchKey>Batthyány Lajos 1807-1849 gr</bflc:name00MatchKey>
    <bflc:name00MarcKey>60014$aBatthyány$jLajos$d1807-
1849$gr$gnémetújvári</bflc:name00MarcKey>
    <rdfs:label>Batthyány Lajos 1807-1849 gr</rdfs:label>
  </bf:Agent>
</bf:subject>
```

MARC/XML 3 - \$b almezőt tartalmaz

```
<marc:datafield tag="600" ind1="1" ind2="4">
<marc:subfield code="a">Lajos</marc:subfield>
<marc:subfield code="d">1638-1715</marc:subfield>
<marc:subfield code="b">XIV</marc:subfield></marc:datafield>
```

RDF/XML 3:

```
<bf:subject>
  <bf:Agent rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-
2020.xmlPIM2138555#Agent600-13">
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Person" />
    <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#PersonalName" />
    <madsrdf:authoritativeLabel>Lajos 1638-1715 XIV</madsrdf:authoritativeLabel>
    <bflc:name00MatchKey>Lajos 1638-1715 XIV</bflc:name00MatchKey>
    <bflc:name00MarcKey>60014$aLajos$d1638-1715$bXIV</bflc:name00MarcKey>
    <rdfs:label>Lajos 1638-1715 XIV</rdfs:label>
  </bf:Agent>
</bf:subject>
```

Címreláció – azaz \$t almező – rögzítése esetén a helyzet bonyolultabb, itt ugyanis az adatmező konvertált alakja három szakaszból tevődik össze: egy Work entitásszintű szakasz jön létre, amelyet az ágens metaadatai követnek, végül a címadatok következnek:

MARC/XML 2:

```
<marc:datafield tag="600" ind1="1" ind2="4">
<marc:subfield code="a">Dante</marc:subfield>
<marc:subfield code="j">Alighieri</marc:subfield>
<marc:subfield code="d">1265-1321</marc:subfield>
<marc:subfield code="t">Isteni színjáték</marc:subfield>
</marc:datafield>
```

RDF/XML:

```
<bf:subject>
  <bf:Work rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-
2020.xmlPIM2104766#Work600-12">
    <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#NameTitle" />
    <madsrdf:authoritativeLabel>Dante Alighieri 1265-1321 Isteni
színjáték</madsrdf:authoritativeLabel>
    <bf:contribution>
      <bf:Contribution>
        <bf:agent>
```

```

    <bf:Agent rdf:about="C:\Users\Hubay
Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-2020.xmlPIM2104766#Agent600-12">
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Person" />
    <bflc:name00MatchKey>Dante Alighieri 1265-1321</bflc:name00MatchKey>
    <bflc:name00MarcKey>60014$aDante$jAlighieri$d1265-1321$tIsteni
színjáték</bflc:name00MarcKey>
    <rdfs:label>Dante Alighieri 1265-1321</rdfs:label>
    </bf:Agent>
  </bf:agent>
  <bf:role>
    <bf:Role rdf:about="http://id.loc.gov/vocabulary/relators/ctb" />
  </bf:role>
  </bf:Contribution>
</bf:contribution>
<rdfs:label>Isteni színjáték</rdfs:label>
<bf:title>
  <bf>Title>
    <bflc:title00MatchKey>Isteni színjáték</bflc:title00MatchKey>
    <bflc:title00MarcKey>60014$aDante$jAlighieri$d1265-1321$tIsteni
színjáték</bflc:title00MarcKey>
    <rdfs:label>Isteni színjáték</rdfs:label>
    <bflc:titleSortKey>Isteni színjáték</bflc:titleSortKey>
    <bf:mainTitle>Isteni színjáték</bf:mainTitle>
  </bf>Title>
</bf:title>
</bf:Work>
</bf:subject>

```

610 – Testületi név típusú tárgyi hozzáférési pont

Képzése a személynév típusú tárgyi hozzáférési ponthoz hasonlóan történik. Az alkalmazott ágenstípus az Organization, amelyet a MADS szótár megfelelő elemével is megfeleltetnek. Érdemes lenne megfontolni azoknál az authority adatmezőknél (s nem csupán a tárgyi, hanem a kiemelt és további hozzáférési pontoknál is, a 110-710 mező esetében), melyek földrajzi típusú kiegészítő elemet tartalmazó almezőt definiálnak, hogy transzformációjuk a bf:place tulajdonság, illetve az ezzel használható bf:Place osztály segítségével menjen végbe.

MARC/XML:

```

<marc:datafield tag="610" ind1="2" ind2="4">
<marc:subfield code="a">Hatvany Lajos Múzeum</marc:subfield>
<marc:subfield code="c">Hatvan</marc:subfield>
</marc:datafield>

```

MARC/XML 2:

```

<marc:datafield tag="610" ind1="2" ind2="4">
<marc:subfield code="a">Kriterion Könyvkiadó</marc:subfield>
<marc:subfield code="c">Bukarest</marc:subfield>
<marc:subfield code="d">1970</marc:subfield>
</marc:datafield>

```

RDF/XML:

```

<bf:subject>
  <bf:Agent rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-
2020.xmlPIM2039608#Agent610-14">
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Organization" />
    <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#CorporateName" />
    <madsrdf:authoritativeLabel>Hatvany Lajos Múzeum
Hatvan</madsrdf:authoritativeLabel>
    <bflc:name10MatchKey>Hatvany Lajos Múzeum Hatvan</bflc:name10MatchKey>
    <bflc:name10MarcKey>61024$aHatvany Lajos Múzeum$cHatvan</bflc:name10MarcKey>
    <rdfs:label>Hatvany Lajos Múzeum Hatvan</rdfs:label>
  </bf:Agent>
</bf:subject>

```


RDF/XML 2:

```
<bf:Agent rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-2020.xmlPIM2092102#Agent610-14">
  <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Organization" />
  <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#CorporateName" />
  <madsrdf:authoritativeLabel>Kriterion Könyvkiadó Bukarest
1970</madsrdf:authoritativeLabel>
  <bflc:name10MatchKey>Kriterion Könyvkiadó Bukarest 1970</bflc:name10MatchKey>
  <bflc:name10MarcKey>61024$Kriterion
Könyvkiadó$Bukarest$d1970</bflc:name10MarcKey>
  <rdfs:label>Kriterion Könyvkiadó Bukarest 1970</rdfs:label>
</bf:Agent>
</bf:subject>
```

611 – Rendezvénynév típusú tárgyi hozzáférési pont

Képzése az előbb ismertetett tárgyi hozzáférési pontokhoz hasonlóan történik. Az alkalmazott ágenstípus a Meeting, amelyet a MADS szótár megfelelő elemével is megfeleltetnek, hasonlóan a 610-es mezőnek megfelelő részgráf generálásához.

MARC/XML:

```
<marc:datafield tag="611" ind1=" " ind2=" ">
<marc:subfield code="a">A bárányok nem hallgatnak</marc:subfield>
<marc:subfield code="b">a juhtartás és a pásztorélet európai öröksége</marc:subfield>
<marc:subfield code="d">2014. április 11-október 31</marc:subfield></marc:datafield>
```

RDF/XML:

```
<bf:subject>
  <bf:Agent rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-2020.xmlPIM2070784#Agent611-14">
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Meeting" />
    <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#ConferenceName" />
    <madsrdf:authoritativeLabel>A bárányok nem hallgatnak 2014. április 11-október
31</madsrdf:authoritativeLabel>
    <bflc:name11MatchKey>A bárányok nem hallgatnak 2014. április 11-október
31</bflc:name11MatchKey>
    <bflc:name11MarcKey>611 $aA bárányok nem hallgatnak$ba juhtartás és a pásztorélet
európai öröksége$d2014. április 11-október 31</bflc:name11MarcKey>
    <rdfs:label>A bárányok nem hallgatnak 2014. április 11-október 31</rdfs:label>
  </bf:Agent>
</bf:subject>
```

630 – Preferált cím típusú tárgyi hozzáférési pont

Az adatmező értékének felhasználásával új, Work entitásszintű metaadatközlés-halmaz áll elő. A korábbi esetekhez hasonlóan itt is problémát jelent, hogy a művet egy olyan URI azonosítja, mely a kiinduló bibliográfiai rekord azonosítójának felhasználásával generálódik, így az elméletben azonos művet jelölő URI-k is eltérők lesznek. A probléma a 6XX mezők URI-t tartalmazó almezővel történő kiegészítésével és újrakonvertálással javítható.

MARC/XML:

```
<marc:datafield tag="630" ind1="0" ind2="4">
<marc:subfield code="a">Anabasis Alexandru</marc:subfield>
</marc:datafield>
```

RDF/XML:

```
<bf:subject>
```

```

<bf:Work rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-
2020.xmlPIM2075917#Work630-16">
  <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#" />
  <madsrdf:authoritativeLabel>Anabasis Alexandru</madsrdf:authoritativeLabel>
  <rdfs:label>Anabasis Alexandru</rdfs:label>
  <bf:title>
    <bf:Title>
      <bflc:title30MatchKey>Anabasis Alexandru</bflc:title30MatchKey>
      <bflc:title30MarcKey>63004$aAnabasis Alexandru</bflc:title30MarcKey>
      <rdfs:label>Anabasis Alexandru</rdfs:label>
      <bflc:titleSortKey>Anabasis Alexandru</bflc:titleSortKey>
      <bf:mainTitle>Anabasis Alexandru</bf:mainTitle>
    </bf:Title>
  </bf:title>
</bf:Work>
</bf:subject>

```

652 – Egyedileg definiált adatmező

A Petőfi Irodalmi Múzeum rekordjaiban a Monguz Kft. által definiált 652-es adatmező a műtárgy elkészítésének technikáját leíró tárgyszavakat tartalmazza. Amennyiben megőrzésük feltétlenül szükséges, formai tárgyszóként rögzíthetők, s érdemes valamely külső névtérből (például a Getty AAT-ból) származó URI-val kiegészíteni az adatokat.

MARC/XML:

```

<marc:datafield tag="652" ind1=" " ind2=" ">
  <marc:subfield code="a">nyomtatvány</marc:subfield>
</marc:datafield>

```

RDF/XML:

```

<bf:genreForm>
  <bf:GenreForm rdf:about="külső URI" >
    <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#Topic" />
    <rdfs:label>nyomtatvány</rdfs:label>
    <madsrdf:authoritativeLabel >nyomtatvány</madsrdf:authoritativeLabel>
  </bf:GenreForm>
</bf:genreForm>

```

700 – Személynév típusú további hozzáférési pont

Az ágensre vonatkozó metaadatközlések a 100-as adatmezőnél leírtakhoz hasonlóan jönnek létre. Fontos különbség, hogy a közreműködői funkció értéke pontosan leképződik a \$4 alapján a bf:role tulajdonság, és osztálya, a bf:Role segítségével. Miután azonban nem elsődleges felelősségi funkciókról van szó, természetesen elmarad a bf:PrimaryContribution osztályba sorolás, helyette bf:Contribution szerepel. Az almezők kezelésével kapcsolatos problémákért lásd a 600-as mező leírását.

MARC/XML:

```

<marc:datafield tag="700" ind1="1" ind2="1">
  <marc:subfield code="a">Györffy</marc:subfield>
  <marc:subfield code="j">György</marc:subfield>
  <marc:subfield code="d">1917-2000</marc:subfield>
  <marc:subfield code="g">szigethi és nádudvari</marc:subfield>
  <marc:subfield code="4">bev., jegyz. kieg., térkép terv</marc:subfield>
</marc:datafield>

```

RDF/XML:

```
<bf:contribution>
  <bf:Contribution>
    <bf:agent>
      <bf:Agent rdf:about="C:\Users\Hubay
Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-2020.xmlPIM2126620#Agent700-14">
        <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Person" />
        <bflc:name00MatchKey>Györffy György 1917-2000</bflc:name00MatchKey>
        <bflc:name00MarcKey>70011$aGyörffy$JGyörgy$d1917-2000$gszigethi és
nádudvari$4bev., jegyz. kieg., térkép terv</bflc:name00MarcKey>
        <rdfs:label>Györffy György 1917-2000</rdfs:label>
      </bf:Agent>
    </bf:agent>
    <bf:role>
      <bf:Role>
        <bf:code>bev., jegyz. kieg., térkép terv</bf:code>
      </bf:Role>
    </bf:role>
  </bf:Contribution>
</bf:contribution>
```

710 – Testületi név típusú további hozzáférési pont

Az URI generálásával kapcsolatos problémák és megoldásaik megegyeznek az előzőekben tárgyalt hozzáférési pontoknál leírtakkal. Az ágens típusa e mező konverziója esetében is Organization, a közreműködés típusa Contribution.

MARC/XML:

```
<marc:datafield tag="710" ind1="2" ind2="1">
<marc:subfield code="a">Móricz Zsigmond Színház</marc:subfield>
<marc:subfield code="c">Nyíregyháza</marc:subfield>
<marc:subfield code="4">közread</marc:subfield>
</marc:datafield>
```

RDF/XML:

```
<bf:contribution>
  <bf:Contribution>
    <bf:agent>
      <bf:Agent rdf:about="C:\Users\Hubay
Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-2020.xmlPIM2255344#Agent710-14">
        <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Organization" />
        <bflc:name10MatchKey>Móricz Zsigmond Színház Nyíregyháza</bflc:name10MatchKey>
        <bflc:name10MarcKey>71021$aMóricz Zsigmond
Színház$cNyíregyháza$4közread</bflc:name10MarcKey>
        <rdfs:label>Móricz Zsigmond Színház Nyíregyháza</rdfs:label>
      </bf:Agent>
    </bf:agent>
    <bf:role>
      <bf:Role>
        <bf:code>közread</bf:code>
      </bf:Role>
    </bf:role>
  </bf:Contribution>
</bf:contribution>
```

MARC/XML 2 - \$b almezőt tartalmaz:

```
<marc:datafield tag="710" ind1="2" ind2="1">
<marc:subfield code="a">Magyar Tudományos Akadémia</marc:subfield>
<marc:subfield code="b">Irodalomtörténeti Bizottság</marc:subfield>
<marc:subfield code="c">Budapest</marc:subfield>
<marc:subfield code="4">közreadó</marc:subfield>
</marc:datafield>
```

RDF/XML 2:

```
<bf:contribution>
  <bf:Contribution>
    <bf:agent>
```

```

    <bf:Agent rdf:about="C:\Users\Hubay
Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-2020.xmlPIM2082087#Agent710-15">
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Organization" />
    <bflc:name10MatchKey>Magyar Tudományos Akadémia Irodalomtörténeti Bizottság
Budapest</bflc:name10MatchKey>
    <bflc:name10MarcKey>71021$aMagyar Tudományos Akadémia$bIrodalomtörténeti
Bizottság$cBudapest$4közreadó</bflc:name10MarcKey>
    <rdfs:label>Magyar Tudományos Akadémia Irodalomtörténeti Bizottság
Budapest</rdfs:label>
  </bf:Agent>
</bf:agent>
<bf:role>
  <bf:Role>
    <bf:code>közreadó</bf:code>
  </bf:Role>
</bf:role>
</bf:Contribution>
</bf:contribution>

```

MARC/XML - \$d almezőt tartalmaz:

```

<marc:datafield tag="710" ind1="2" ind2="1">
<marc:subfield code="a">Magyarország szerkesztősége</marc:subfield>
<marc:subfield code="c">Budapest</marc:subfield>
<marc:subfield code="d">1893-1944</marc:subfield>
<marc:subfield code="4">címzett</marc:subfield>
</marc:datafield>

```

RDF/XML:

```

<bf:contribution>
  <bf:Contribution>
    <bf:agent>
      <bf:Agent rdf:about="C:\Users\Hubay
Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-2020.xmlPIM2055074#Agent710-16">
      <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Organization" />
      <bflc:name10MatchKey>Magyarország szerkesztősége Budapest 1893-
1944</bflc:name10MatchKey>
      <bflc:name10MarcKey>71021$aMagyarország szerkesztősége$cBudapest$d1893-
1944$4címzett</bflc:name10MarcKey>
      <rdfs:label>Magyarország szerkesztősége Budapest 1893-1944</rdfs:label>
    </bf:Agent>
  </bf:agent>
<bf:role>
  <bf:Role>
    <bf:code>címzett</bf:code>
  </bf:Role>
</bf:role>
</bf:Contribution>
</bf:contribution>

```

711 – Rendezvénynév típusú további hozzáférési pont

Az adatmező konverziója rendben végbemegy, azonban ebben az esetben is szükséges egy fix URI rögzítése a bemeneti adathalmazba, hogy elkerülhető legyen a mechanikus, előfordulásonként eltérő ágens-azonosító generálódása.

MARC/XML:

```

<marc:datafield tag="711" ind1="2" ind2="1">
<marc:subfield code="a">Fiatalok Konferenciája</marc:subfield>
</marc:datafield>

```

RDF/XML:

```

<bf:contribution>
  <bf:Contribution>
    <bf:agent>
      <bf:Agent rdf:about="C:\Users\Hubay
Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-2020.xmlPIM2268213#Agent711-17">
      <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Meeting" />
      <bflc:name11MatchKey>Fiatalok Konferenciája</bflc:name11MatchKey>

```

```

        <bflc:name11MarcKey>71121$aFiatalok Konferenciája</bflc:name11MarcKey>
        <rdfs:label>Fiatalok Konferenciája</rdfs:label>
    </bf:Agent>
</bf:agent>
<bf:role>
    <bf:Role rdf:about="http://id.loc.gov/vocabulary/relators/ctb" />
</bf:role>
</bf:Contribution>
</bf:contribution>

```

730 – Preferált cím típusú további hozzáférési pont

Az adatmező tartalmából Work entitásszintű metaadat-közléshalmazok jönnek létre, mindegyik mechanikusan generált URI-val. Ezek a metaadat-közléshalmazok a szemantikai szempontból igen alulspecifikált `bf:relatedTo` relációval kapcsolódnak az eredeti mű metaadat-közléshalmazához, nem közölve semmit a MARC-rekordból kiolvasható hierarchiaszintbeli eltérésekről (például a közös főcímű szerzői gyűjtemény, illetve a benne olvasható, önálló művek).

MARC/XML:

```

<marc:datafield tag="730" ind1="2" ind2="1">
<marc:subfield code="a">A zendülők</marc:subfield>
</marc:datafield>
<marc:datafield tag="730" ind1="0" ind2="1">
<marc:subfield code="a">Vendégjáték Bolzanoban</marc:subfield>
</marc:datafield>
<marc:datafield tag="730" ind1="2" ind2="1">
<marc:subfield code="a">A gyertyák csonkig égnek</marc:subfield>
</marc:datafield>
<marc:datafield tag="730" ind1="3" ind2="1">
<marc:subfield code="a">Az igazi ; Judit ... és az utóhang</marc:subfield>
</marc:datafield>

```

RDF/XML:

```

<bf:relatedTo>
    <bf:Work rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-2020.xmlPIM2153599#Work730-27">
        <rdfs:label>A zendülők</rdfs:label>
        <bf:title>
            <bf:Title>
                <bflc:title30MatchKey>A zendülők</bflc:title30MatchKey>
                <bflc:title30MarcKey>73021$aA zendülők</bflc:title30MarcKey>
                <rdfs:label>A zendülők</rdfs:label>
                <bflc:titleSortKey>zendülők</bflc:titleSortKey>
                <bf:mainTitle>A zendülők</bf:mainTitle>
            </bf:Title>
        </bf:title>
    </bf:Work>
</bf:relatedTo>
<bf:relatedTo>
    <bf:Work rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-2020.xmlPIM2153599#Work730-28">
        <rdfs:label>Vendégjáték Bolzanoban</rdfs:label>
        <bf:title>
            <bf:Title>
                <bflc:title30MatchKey>Vendégjáték Bolzanoban</bflc:title30MatchKey>
                <bflc:title30MarcKey>73001$aVendégjáték Bolzanoban</bflc:title30MarcKey>
                <rdfs:label>Vendégjáték Bolzanoban</rdfs:label>
                <bflc:titleSortKey>Vendégjáték Bolzanoban</bflc:titleSortKey>
                <bf:mainTitle>Vendégjáték Bolzanoban</bf:mainTitle>
            </bf:Title>
        </bf:title>
    </bf:Work>
</bf:relatedTo>
</bf:relatedTo>

```

```

    <bf:Work rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-
2020.xmlPIM2153599#Work730-29">
      <rdfs:label>A gyertyák csonkig égnék</rdfs:label>
      <bf:title>
        <bf:Title>
          <bflc:title30MatchKey>A gyertyák csonkig égnék</bflc:title30MatchKey>
          <bflc:title30MarcKey>73021$A gyertyák csonkig égnék</bflc:title30MarcKey>
          <rdfs:label>A gyertyák csonkig égnék</rdfs:label>
          <bflc:titleSortKey>gyertyák csonkig égnék</bflc:titleSortKey>
          <bf:mainTitle>A gyertyák csonkig égnék</bf:mainTitle>
        </bf:Title>
      </bf:title>
    </bf:Work>
  </bf:relatedTo>
  <bf:relatedTo>
    <bf:Work rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-
2020.xmlPIM2153599#Work730-30">
      <rdfs:label>Az igazi ; Judit ... és az utóhang</rdfs:label>
      <bf:title>
        <bf:Title>
          <bflc:title30MatchKey>Az igazi ; Judit ... és az utóhang</bflc:title30MatchKey>
          <bflc:title30MarcKey>73031$aAz igazi ; Judit ... és az
utóhang</bflc:title30MarcKey>
          <rdfs:label>Az igazi ; Judit ... és az utóhang</rdfs:label>
          <bflc:titleSortKey>igazi ; Judit ... és az utóhang</bflc:titleSortKey>
          <bf:mainTitle>Az igazi ; Judit ... és az utóhang</bf:mainTitle>
        </bf:Title>
      </bf:title>
    </bf:Work>
  </bf:relatedTo>

```

740 – További cím és felelősségi közlés

Az adatmező tartalmából, hasonlóan a 730-asnál leírtakhoz, Work entitásszintű metaadat-közléshalmazok jönnek létre. Az eredeti mű metaadat-közléshalmazához itt is a bf:relatedTo reláció hozza létre a kapcsolatokat, mely ennek a mezőnek az esetében sem árul el semmi bővebbet a kapcsolat természetéről.

*MARC/XML*⁵⁵:

```

<marc:datafield tag="740" ind1="3" ind2="1">
<marc:subfield code="a">Le Premier amour</marc:subfield>
<marc:subfield code="e">trad. Catherine Fay</marc:subfield>
</marc:datafield>
<marc:datafield tag="740" ind1="4" ind2="1">
<marc:subfield code="a">Les Révoltés</marc:subfield>
<marc:subfield code="e">trad. Ladislás Gara, Marcel Largeaud</marc:subfield>
</marc:datafield>
<marc:datafield tag="740" ind1="3" ind2="1">
<marc:subfield code="a">La Conversation de Bolzano</marc:subfield>
<marc:subfield code="e">Natalia Zarembo-Huzsvai, Charles Zarembo</marc:subfield>
</marc:datafield>
<marc:datafield tag="740" ind1="4" ind2="1">
<marc:subfield code="a">Les Braises</marc:subfield>
<marc:subfield code="e">trad. Marcelle Rénier, Georges Régnier</marc:subfield>
</marc:datafield>
<marc:datafield tag="740" ind1="0" ind2="1">
<marc:subfield code="a">Métamorphoses d'un mariage</marc:subfield>
<marc:subfield code="e">trad. Georges Kassai, Zéno Bianu</marc:subfield>
</marc:datafield>

```

RDF/XML:

```

<bf:relatedTo>
  <bf:Work rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-
2020.xmlPIM2153599#Work740-31">

```

⁵⁵ A PIM eredeti rekordjaiban téves a második indikátor egyes értéke – gyűjtemények esetében a 2 alkalmazandó.

```

<rdfs:label>Le Premier amour</rdfs:label>
<bf:title>
  <bf:Title>
    <rdfs:label>Le Premier amour</rdfs:label>
    <bflc:titleSortKey>Premier amour</bflc:titleSortKey>
    <bf:mainTitle>Le Premier amour</bf:mainTitle>
  </bf:Title>
</bf:title>
</bf:Work>
</bf:relatedTo>
<bf:relatedTo>
  <bf:Work rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-2020.xmlPIM2153599#Work740-32">
    <rdfs:label>Les Révoltés</rdfs:label>
    <bf:title>
      <bf:Title>
        <rdfs:label>Les Révoltés</rdfs:label>
        <bflc:titleSortKey>Révoltés</bflc:titleSortKey>
        <bf:mainTitle>Les Révoltés</bf:mainTitle>
      </bf:Title>
    </bf:title>
  </bf:Work>
</bf:relatedTo>
<bf:relatedTo>
  <bf:Work rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-2020.xmlPIM2153599#Work740-33">
    <rdfs:label>La Conversation de Bolzano</rdfs:label>
    <bf:title>
      <bf:Title>
        <rdfs:label>La Conversation de Bolzano</rdfs:label>
        <bflc:titleSortKey>Conversation de Bolzano</bflc:titleSortKey>
        <bf:mainTitle>La Conversation de Bolzano</bf:mainTitle>
      </bf:Title>
    </bf:title>
  </bf:Work>
</bf:relatedTo>
<bf:relatedTo>
  <bf:Work rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-2020.xmlPIM2153599#Work740-34">
    <rdfs:label>Les Braises</rdfs:label>
    <bf:title>
      <bf:Title>
        <rdfs:label>Les Braises</rdfs:label>
        <bflc:titleSortKey>Braises</bflc:titleSortKey>
        <bf:mainTitle>Les Braises</bf:mainTitle>
      </bf:Title>
    </bf:title>
  </bf:Work>
</bf:relatedTo>
<bf:relatedTo>
  <bf:Work rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-2020.xmlPIM2153599#Work740-35">
    <rdfs:label>Métamorphoses d'un mariage</rdfs:label>
    <bf:title>
      <bf:Title>
        <rdfs:label>Métamorphoses d'un mariage</rdfs:label>
        <bflc:titleSortKey>Métamorphoses d'un mariage</bflc:titleSortKey>
        <bf:mainTitle>Métamorphoses d'un mariage</bf:mainTitle>
      </bf:Title>
    </bf:title>
  </bf:Work>
</bf:relatedTo>

```

741 – Párhuzamos cím

A HUNMARC és a MARC 21 eltérően kezeli a párhuzamos címeket. Előbbi ugyan definiál almezőket a 245-ben erre a célra, később azonban megjegyzi, hogy „*A párhuzamos adatok leírása csak akkor indokolt a 245-ös mezőben, ha a szabványos bibliográfiai leírás megjelenítése nem lehetséges a párhuzamos cím 741 Párhuzamos cím mezőben történő*

leírásával” (OSZK, 2002, p. 95.). Így a párhuzamos címek elsődlegesen a 741-es adatmezőben található. A MARC 21 ellenben két helyen is előírja az ilyen adatok rögzítését: egyszer a 245\$b almezőben, egyszer pedig a 246\$a almezőben; ez utóbbi esetben a második indikátor egyes értéke jelzi, hogy párhuzamos adatról van szó. E különbségek miatt nem csoda, hogy az RDF/XML kimenetben nem találhatók párhuzamos címek, noha maga a BIBFRAME rendelkezik az ilyen adatok kezelésére megfelelő osztállyal: ez a `bf:Title` osztály `bf:VariantTitle` alosztályának `bf:ParallelTitle` alosztálya. Érdekes ugyanakkor tekintetbe venni a PMO szótárát kidolgozó szakemberek véleményét, akik szerint ennek az adatelemnek kizárólag a bibliográfiai rekordban van relevanciája, így tehát *„a `bf:ParallelTitle` kifejezésnek nincs helye a linked data-keretrendszerben, mivel az egyik nyelv kitüntettségét sugallja a többivel szemben. Az ilyen címpárokat/címcsoportokat egyenlően kellene kezelni, nyelvi minősítővel ellátva.”* (*Titles in BIBFRAME*, é.n.)

743 – Borítócím

A borítócímet a MARC 21 – hasonlóan a párhuzamos címekhez – a 246-os mező segítségével kezeli, ekkor a második indikátorpozíción négyes érték található. A HUNMARC-tól való eltérés okán a 743-as mezőben tárolt adatok sem jelennek meg az RDF/XML-ben. A specifikáció alapján Instance entitásszintű metaadat-közlések jönnek létre, a `bf:VariantTitle` osztály, illetve a `bf:variantType` tulajdonság segítségével, amelynek értéke a borítócímek esetében „cover”.

MARC/XML

```
<marc:datafield tag="743" ind1="0" ind2="1">
<marc:subfield code="a">In memoriam Szoboszlai Aladár I</marc:subfield>
</marc:datafield>
```

856 – Elektronikus hely és hozzáférés

A kimeneti állomány vizsgálatával fény derült rá, hogy egyéb adatelemek hiányában az Item entitásszint metaadat-közléshalmaza csak akkor jött létre a gráfban, ha az adott MARC-rekord tartalmazott 856-os mezőt. Ezt a közléshalmazt a `bf:hasItem`, illetve ennek inverze, a `bf:itemOf` köti össze a hozzá kapcsolódó Instance-szintű halmazzal.

MARC/XML:

```
<marc:datafield tag="856" ind1=" " ind2=" ">
<marc:subfield code="3">honlap</marc:subfield>
<marc:subfield code="u">http://www.participatorymuseum.org</marc:subfield>
</marc:datafield>
```


RDF/XML:

```
<bf:hasItem>
  <bf:Item rdf:about="C:\Users\Hubay Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-
2020.xmlPIM2104086#Item856-14">
    <bf:electronicLocator>
      <rdfs:Resource>
        <bf:target rdf:resource="http://www.participatorymuseum.org" />
        <bf:note>
          <bf:Note>
            <rdfs:label>honlap</rdfs:label>
          </bf:Note>
        </bf:note>
      </rdfs:Resource>
    </bf:electronicLocator>
    <bf:itemOf rdf:resource="C:\Users\Hubay
Miklós\Downloads\DISSZERTÁCIÓ\21826_konyv2015-2020.xmlPIM2104086#Instance856-14" />
  </bf:Item>
</bf:hasItem>
```

910 – Egyedileg definiált

Egyedileg definiált adatmező a PIM halmazában. Csekély előfordulásszáma miatt (egyetlenegy rekordban) konverziójától eltekintettem.

12. Az eredmények összegzése, értékelése

E fejezetben összegzem az előzőekben ismertetett konverziós esettanulmány eredményeit, amely összegzés központjába azt a szempontot állítom, mekkora előzetes munkára van szükség az egyes adatmezők adattartalmának BIBFRAME-konverziójához; milyen és mekkora mértékű módosítás szükséges a MARC 21-re épülő, eredeti konverziós stíluslapokon ahhoz, hogy a HUNMARC-ban tárolt bibliográfiai rekordok transzformációja a tudatosan vállalt adatvesztéseken kívül megfelelő minőségben menjen végbe. Az összegzés után a dolgozat téziseinek vizsgálata következik, zárásként pedig a további lehetséges kutatási irányvonalakat, továbbfejlesztési lehetőségeket mutatom be, amelyek hozzájárulhatnak a linked data ökoszisztéma, továbbá a BIBFRAME keretrendszer hazai és akár nemzetközi implementációjának tökéletesítéséhez.

12.1. A konverziós esettanulmány eredményeinek táblázatos összefoglalása

Hívójel	Eredmény
000	az eredeti specifikáció alapján végrehajtható
001	kisebb módosítással végrehajtható (LoC-egyediség megszüntetése)
002	nem releváns adatmező a tesztalomban
005	az eredeti specifikáció alapján végrehajtható
008	az eredeti specifikáció alapján végrehajtható, további vizsgálat szükséges a többi rögzített hosszúságú mező adattartalmának pontos tükrözéséhez, minden használati esetben
011	nem releváns adatmező a tesztalomban
020	az eredeti specifikáció alapján végrehajtható
022	az eredeti specifikáció alapján végrehajtható
040	az eredeti specifikáció alapján végrehajtható
041	az eredeti specifikáció alapján végrehajtható, fordítás szükséges
044	nem releváns adatmező a tesztalomban
080	meglévő kódrészlet újrafelhasználásával és módosításával végrehajtható
090	egyedi adatmező a kiinduló halmazban, egyedi konverziós eljárás szükséges
100	a specifikáció módosításával végrehajtható (URI-k kezelése; kimaradó almezők pótlása, LoC kiegészítő adatelemek relevanciavizsgálata)
110	a specifikáció módosításával végrehajtható (URI-k kezelése; kimaradó almezők pótlása, LoC kiegészítő adatelemek relevanciavizsgálata)
240	az eredeti specifikáció alapján végrehajtható
245	kisebb módosításokkal végrehajtható (felelősségi közlések kezelése Őe almező jelenléte esetén)
250	az eredeti specifikáció alapján végrehajtható
259	nem releváns adatmező a tesztalomban
260	az eredeti specifikáció alapján végrehajtható
300	kisebb módosítások szükségesek a mellékletek adatelemeit tartalmazó almezők kezeléséhez, fordítás szükséges
362	az eredeti specifikáció alapján végrehajtható

440	kisebb módosításokkal végrehajtható (URI-k kezelése, ágens entitás-előfordulások társítása)
480	nem releváns adatmező a tesztalmazban
490	az eredeti specifikáció alapján végrehajtható
500	az eredeti specifikáció alapján végrehajtható
503	nem releváns adatmező a tesztalmazban
504	az eredeti specifikáció alapján végrehajtható
505	az eredeti specifikáció alapján végrehajtható
562	új kódrészlet már meglévő specifikáció kisebb módosításával
588	fordítás és módosítás szükséges a konzisztencia érdekében
592	új kódrészlet már meglévő specifikáció kisebb módosításával
600	a specifikáció módosításával végrehajtható (URI-k kezelése; kimaradó almezők pótlása, LoC kiegészítő adatelemek relevanciavizsgálata)
610	a specifikáció módosításával végrehajtható (URI-k kezelése; kimaradó almezők pótlása, LoC kiegészítő adatelemek relevanciavizsgálata, BIBFRAME-reláció a földrajzi nevek kezeléséhez)
611	a specifikáció módosításával végrehajtható (URI-k kezelése; kimaradó almezők pótlása, LoC kiegészítő adatelemek relevanciavizsgálata)
630	a specifikáció módosításával végrehajtható (URI-k kezelése; kimaradó almezők pótlása, LoC kiegészítő adatelemek relevanciavizsgálata)
652	egyedi adatmező a kiinduló halmazban, új kódrészlet már meglévő specifikáció kisebb módosításával
700	a specifikáció módosításával végrehajtható (URI-k kezelése; kimaradó almezők pótlása, LoC kiegészítő adatelemek relevanciavizsgálata)
710	a specifikáció módosításával végrehajtható (URI-k kezelése; kimaradó almezők pótlása, LoC kiegészítő adatelemek relevanciavizsgálata)
711	a specifikáció módosításával végrehajtható (URI-k kezelése; kimaradó almezők pótlása, LoC kiegészítő adatelemek relevanciavizsgálata)
730	módosítás javasolt a specifikus kapcsolatok rögzítése érdekében
740	módosítás javasolt a specifikus kapcsolatok rögzítése érdekében
741	<i>a párhuzamos címek jelenléte vitatott a linked data-ökoszisztémában</i>
743	módosítás szükséges a MARC-ok eltérő specifikációja miatt
856	az eredeti specifikáció alapján végrehajtható
910	nem releváns adatmező a tesztalmazban

12.2. Tézisvizsgálat

1. tézis: A könyvtárügy a bibliográfiai univerzum entitásalapú szemléletét legmegfelelőbben a linked data-technológia segítségével ültetheti át a katalogizálási gyakorlatba. E technológia számottevő előnyökkel bír a jelenlegi megoldásokhoz képest.

Az értekezésben bemutattam, hogy az RDA katalogizálási szabályzat több metaadateleme MARC-ban is megjeleníthető, így a hagyományos hordozóformátum bizonyos mértékig igazítható az RDA rendelkezéseivel. A tartalomra, a médiumra, illetve a hordozóra vonatkozó közlések elhelyezése, a megjelenési adatok aprólékosabb kezelése,

illetve az URI-k elhelyezésének lehetősége azonban nem oldja meg az eredeti problémát: a fogalmi modellekben, az FRBR-ben és később az LRM-ben leírt entitások számának megfelelő mennyiségű rekordtípus szétfeszítené a MARC használhatóságának kereteit. A jelenlegi adatsere-formátum alkalmazásával nem biztosítható a világháló szövetébe való integrálódás, a karakterlánc-alapú adattárolás problémáinak (például az elgépeléseknek) a megfelelő kezelése, illetve a homonimák kielégítő, egyértelmű megkülönböztetése. E problémák mindegyikére kielégítő megoldást kínál a szemantikus web technológiája.

Az értekezés 5.3. fejezetében a MARC-rekordok FRBR-izált megjelenítésének technikai megoldásait ismertettem, amelyből kiderült, hogy megfelelő adatminőség és adatgazdagság esetén az alkalmazott algoritmusok megfelelően képesek felderíteni az FRBR első entitáscsoportjának entitásait, illetve azokat csoportosítva megjeleníteni. Ezek az adatok azonban nagyon erősen őrzik a MARC-rekord eredeti, kötött struktúráját, továbbá csak nagyon minimálisan lehet kívülré mutató hivatkozásokat kialakítani az egyes metaadatelemek értékein keresztül. Emiatt, bár az FRBR-esített katalógus előrelépést jelenthet a megjelenítő eszközök vonatkozásában, valójában csak igen korlátozott előnyöket mutat az adatok hálózatba integrálhatóságának területén. A linked datáról szóló 6. fejezetből kiderül, hogy ez a technológia – mivel alapvetően a kapcsolatok kialakítására épül – sokkal megfelelőbb az adatok tárolására, mint a MARC, és immanensen tartalmazza az entitások metaadatelemeinek elkülönítési lehetőségét, bármiféle utómunka elvégzésének szükségessége nélkül.

2. tézis: A könyvtárak számára előnyös választás, ha az RDF-re épülő adatszolgáltatást elsősorban a BIBFRAME bibliográfiai keretrendszerre alapozva valósítják meg.

A BIBFRAME bibliográfiai keretrendszer részletes elemzését az értekezés 9. fejezetében végeztem el, ismertettem kidolgozásának a történetét, más intézmények, illetve projektek keretében fejlesztett bővítményeit, bemutattam a más metaadat-modellek felé történő átjárás lehetőségeit és kihívásait. Szót ejtettem több, BIBFRAME-re épülő, natív linked data-katalogizálási felületről, amelyek grafikus felhasználói interfésze a legtöbb esetben megfelelően granuláris és informatív a feldolgozó könyvtárosok számára, s mindezt anélkül, hogy a MARC-hoz képest valóban összetettebb nyers adathalmazt láthatóvá tenné a felhasználónak. A BIBFRAME, a MARC-hoz való erős kötődése miatt alkalmas arra, hogy

a MARC-alapú örökséget megfelelő minőségben és teljességgel konvertálhatóvá tegye a gráfokra épülő adatszolgáltatások számára.

Azonban nem hagyhatók figyelmen kívül a BIBFRAME-szótár problémái sem, amelyeket a szótár 1.0 és 2.0 verziója, illetve Rob Sanderson munkássága nyomán mutattam be a 9.1. és 9.2. fejezetekben. A BIBFRAME, mivel a legtöbb, más elemkészletekben már meglévő relációt újradefiniálja, kevésbé alkalmas arra, hogy a könyvtárak által átkonvertált adathalmazokat valóban a globális adathálózat részévé tegye; kifejezetten kevés a külső kapcsolódási pont, amelyek hiányát más szótárakból származó elemek ekvivalencia-relációkkal történő kódolásával orvosolni lehetne. De ugyanez igaz azon metaadatelemek értékei esetében is, amelyeket szótárkódolási sémából kell választani: szükséges más adathalmazok, például az RDA Registry által biztosított sémák, vagy a Getty felé is kialakítani a kapcsolatokat, igen jelentékeny felhasználási területet biztosítva azoknak.

A metaadatsémák átjárhatóságáról szóló fejezetből, s különösen az RDA adatmodelljével összefüggésben történő vizsgálódásból látható, hogy a két szótár (ti. a BIBFRAME és az RDA) egymáshoz igazítása, a két modell alapvető entitásstruktúrája miatt nehézkes, a megfelelő körültekintés nélkül végrehajtott konverzió adatminőség-romlással jár együtt. Nagy adóssága az RDA és a BIBFRAME köré épült szakmai közösségeknek, hogy mindaddig nem készült tételes megfeleltetés a két szótár között, amely minden modellezési területre és bibliográfiai forrástípusra vonatkozóan megmutatná a konverziós nehézségeket, az adatmodellek divergenciájából származó kihívásokat.

Fentiek alapján a 2. tézisről kijelenthető, hogy annak tartalma csak részben igazolódott. A könyvtárak számára valóban előnyös, ha a BIBFRAME-szótárra alapozott adatszolgáltatást valósítanak meg, hiszen ennek segítségével könnyűszerrel kapcsolatba léphetnek más intézményekkel – e tekintetben a BIBFRAME-gráf lényegében az OAI-protokoll közbeiktatásával közvetített MARC/XML-ek utódjaként is számításba jön –, ezen kívül hatékony alapja lehet bármilyen, gráfokra épülő megjelenítő eszköz működtetésének. Bebizonyosodott, hogy az értekezésben vizsgált, a legfontosabb metaadatelemeket tartalmazó, specifikus kiinduló adathalmaz esetében nincsen szükség kiegészítő szemantikus elemkészlet alkalmazására. A 9. fejezetben közölt információk alapján azonban javasolt egy párhuzamos, RDA-szótárra épülő gráfot is szolgáltatni, egészen addig a pillanatig, amíg a két elemkészlet közötti átjárás problémáira nem születik megnyugtató megoldás.

3. tézis: A MARC 21-BIBFRAME irányú konverziós segédeszközök (stíluslapok) kisebb módosításokkal alkalmassá tehetők a HUNMARC-ban tárolt könyvtári adatvagyon megfelelő minőségű transzformációjára is.

A 3. tézis a 11. fejezetben közölt konverziós eredmények, illetve kritikai javaslatok összegzése alapján igazolódott: a MARC 21, illetve a HUNMARC közötti különbségek az értekezésben közölt, megfelelően adatgazdag kiinduló adathalmaz esetében nem igényelnek semmiféle mélyszerkezeti, entitásstruktúrát érintő, nagyobb horderejű változtatást. (A teljes HUNMARC-ra vonatkozó megállapításokat természetesen a 12.3. fejezetben említendő, valamennyi adatmezőre kiterjedően végrehajtott konverzió elvégzése és kritikai elemzése után lehet tenni). Megkerülhetetlen feladat a konverziós stíluslapok megfelelő előkészítése a két formátum eltérő adatmező-, illetve almező-kezelése miatt, emellett fontos a hazai, magyar nyelvű szakterminológia érvényesítése például az egyes megjegyzéseket tartalmazó relációk címkéi esetében.

4. tézis: A linked data-technológia bevezetése nem kizárólag a könyvtáros szakemberek feladata. Igazán hatékonyan és eredményesen közgyűjteményi dolgozók és informatikusok együttműködésével valósulhat meg az átállási vagy adatszolgáltatási projekt.

Sokakat elbátortalanít a linked data-technológiára való átállás folyamatának azon pontja, amikor közelebbről találkoznak az RDF-szintaxisú adatállományok első pillantásra átláthatatlanul bonyolult, összetett szerkezetével. Az értekezés 7.1.4. fejezetében bemutatott workflow alapján azonban látható, hogy már az adatok konverzióját sem feltétlenül magának a könyvtárosnak kell végezni, s még kevésbé igaz ez a megjelenítő felületek programozására, vagy a konverziós segédeszközök fejlesztésére. Az értekezés 3. fejezetében rámutattam, hogy az SGML vagy az XML korszakában is speciális szaktudással rendelkező informatikusok állították elő az adattranszformációt végző stíluslapokat, az egyes metaadatelemek előzetes megfeleltetése azonban jórészt a közgyűjteményi munkatársak feladata. Nincs ez másként az RDF és a tripletek korában sem. Az adatkonverziós és adatmodellezési feladatok megoldásakor a két terület (közgyűjteményi és „általános” informatika) egymásra van utalva. Jó példája ennek a BIBFRAME fejlesztésének története: a szótár Zepheira informatikai cég, valamint a Library of Congress együttműködésében

született meg. Rámutattam, hogy a BIBFRAME-ben megvalósított entitásszerkezet a közgyűjteményi dolgozók javaslata volt, sok évvel az együttműködés kezdete előtt.

Az értekezés 7.3.1., illetve 7.3.2. fejezetében bemutatott szolgáltatások esetében az adott szoftvereszközt informatikai szaktudással rendelkező szakemberek készítették, akik a projekt során együttműködtek a közgyűjtemény munkatársaival, s a kísérlet során felhasználták az intézmény teljes (Innovative Inspire) vagy célszerűen válogatott (Qulto) adathalmazát. Disszertációm 9.7. fejezetében pedig olyan katalogizáló felületeket ismertettem, amelyeket ugyancsak informatikusok állítottak elő, a tényleges adatbevitelt (az elégséges méretű *data pool* kialakítását), az esetleges inkoherenciák, az adatbevitellel összefüggő módszertani problémák felderítését pedig a közgyűjteményi dolgozók végzik.

Az értekezésben bemutatott esettanulmányok a 4. tézist megfelelően igazolják.

5. tézis: Az RDF-logikát támogató formátumokra épülő információkereső felületek valóban újszerű felhasználói élményt jelentenek.

Disszertációm 5.3. fejezetében ismertettem a MARC-rekordok utólagos kezelésével előálló, ún. FRBR-izált katalógusokat, az értekezés 7.3. fejezete pedig már valóban az új adatmodellezési módszerre, a linked data-ra épülő felhasználói felületeket mutatott be. Az RDF informatikai előnyeit az 1. tézis igazolásánál már bemutattam, ezen a ponton kizárólag a felhasználói felület megújításának kérdésére szorítkozom. Az entifikáció – akárcsak az FRBR-izálás esetében – mindegyik bemutatott példában megjelenik, a szerendipitás, azaz a váratlan felfedezések bekövetkezési esélyének növelése legfőképpen az Innovative Inspire-ra volt jellemző. Az adatok gazdagítása, továbbá a Google által kijelölt úton haladó keresőfelület-egyszerűsítési törekvések a Stanford Searchworks-felületén jelentek meg leginkább. A Qulto szemantikus katalógusa az entifikáció területén hozott újat, amennyiben külön műként mutatja be például az egyes könyvjellegű bibliográfiai forrásokban szereplő illusztrációkat.

E fejezet tanulságait összegezve az 5. tézis is csak részben igazolódott: az új típusú adattárolás nem feltétlenül garancia arra a jelenlegi gyakorlatban, hogy a 7.3. fejezet címében szereplő három hívószóval (entifikáció, szerendipitás, adatgazdagítás) jelölt előnyök valóban megjelennek egy információkereső felületen. A Qulto esetében az RDF-re

épülő megjelenítés csak részben jelent újat a „hagyományos”, utófeldolgozáson alapuló FRBR-esítéshez képest. Ennek okát a MARC-tól való elszakadás nehézségében látom: bizonytalanul lépünk az új területre, s még a felhasználóknak is időbe telik, míg megbarátkoznak az új lehetőségekkel.

Szükséges a fejlődés az adatgazdagítás nyújtotta előnyök kihasználásának területén is. Bár elismerésre méltó, egyre inkább terjedő jó gyakorlat a Wiki-univerzum szolgáltatásainak – Wikipédia, Wikidata – beépítése a szolgáltatásokba (a 7.3. fejezetben bemutatott példák közül kettő, továbbá a Petőfi Irodalmi Múzeum is használja a Wiki-univerzumot, elsősorban keresőoptimalizálási céllal, amelyről a 8. fejezetben írtam), a Qulto esetében is csak korlátozottan, a személyek adatait hivatkozzák külső adatforrásokból, ezekre az adatokra pedig nem épül szolgáltatás. A Stanford SearchWorks esetének tanulmányozása jó példával szolgál arra, hogy milyen módon lehet bővíteni az adathalmaz kereshetőségét kívülről származó adatelemek felhasználásával (például: ki végzett a Stanfordin a talált szerzők közül?), illetve hogyan lehet akár egyszerre több adatforrásból is gazdagítani például a földrajzi nevek állományát (Who’s On First, illetve Wikidata egyszerre történő alkalmazása).

12.3. További kutatási irányok, lehetőségek

A BIBFRAME a fentiekben közöltek alapján megfelelő hordozóformátuma lehet a bibliográfiai adatoknak a szemantikus világhálón, ezáltal biztos alapul szolgál mind a katalogizálás entítasalapú szemléletének gyakorlati megvalósításához, mind pedig az információkeresés folyamatának megújításához.

Ebben az értekezésben két intézmény korlátozott méretű, válogatott bibliográfiairekord-állományának konverziójáról esett szó. Jóllehet ezek az állományok tartalmazzák a legfontosabb bibliográfiai adatelemeknek megfelelő almezőket, a kép természetesen nem teljes. Átfogó, módszeres vizsgálatra van szükség annak megállapítása érdekében, hogy a HUNMARC valamennyi használati esetében igazolható legyen, hogy a BIBFRAME-konverzió megfelelő minőségben, adatvesztés nélkül kivitelezhető. Ennek kivitelezése során a HUNMARC valamennyi, jelenleg is érvényes adat- és almezője, továbbá indikátorértéke esetében el kell végezni a gráffá alakítást, és az eredmények vizsgálatával következtetéseket kell levonni a transzformációs folyamatról, s előállítani egy, teljesen a HUNMARC-ra optimalizált BIBFRAME-konverziós stíluslapcsomagot. A

megfelelő minőségű, átfogó konverziós útmutató elengedhetetlen az információkereső felületek paraméterezésének előkészítéséhez.

Fontos – és a szemantikus webes törekvések megvalósítása előtt végrehajtható feladat – a MARC 21 szerkezetét követve a HUNMARC-ba is bevezetni az RDA-adatmezőket és -almezőket (lásd például az URI-k tárolására szolgáló \$0 és \$1 almezőt), így közelítve a hazai feldolgozó könyvtárosokat és könyvtári informatikusokat az entitásalapú szemlélethez. A szoftverrendszerek forgalmazóinak a szabványosítást követően érdemes minél előbb bevezetni ezeket a módosításokat. Ezzel a külföldi adatbázisokból átemelt bibliográfiai rekordok adattartalmának jóval nagyobb része hasznosulhat a magyarországi információkeresésben is, továbbá a hazánkban előállított rekordok szerkezete idomulni fog a nemzetközi katalógusrendszerekben előforduló más tételekhez, csökken a távolság, növekszik az egységesség.

Az RDA katalogizálási szabályzat rendelkezéseit kísérő szótárak a BIBFRAME versenytársaiként vannak jelen a közgyűjteményi informatikában. A két szótár közötti átjárást azonban, a származékos bibliográfiai kapcsolatok kivételével igen kevés alkalommal vizsgálták, így ezen a területen bőven maradtak izgalmas, választ váró kutatási kérdések, például az eltérő entításszerkezetből adódó konverziós kihívásokkal kapcsolatban.

Az információkeresés szempontjából is érdemes további vizsgálatok alanyává tenni mind a BIBFRAME-, mind az RDA-gráfokat, s felderíteni, hogy ezek az adathalmazok mennyire szolgálják a felhasználó kényelmét, mennyire könnyen építhetők rájuk grafikus használói felületek, mennyire teszik az adatmodellek valóban intuitívvá, élményszerűbbé a közgyűjteményi információkeresést. A korábban leírtak értelmében nem is feltétlenül szükséges választani a két adatmodell közül, ha szükséges, az adatszolgáltatást akár mindkét szótár segítségével ki lehet építeni.

A nemzeti bibliográfiai adatvagyonot érdemes egyetlen, igen kiterjedt gráfba konvertálni (data pool), amelyet a művek, kifejezési formák, illetve megjelenési formák központi adatbázisaként lehetne használni. Ennek azonban elsődleges feltétele, hogy a hazánkban alkalmazott integrált könyvtári rendszerek, platformok fel legyenek készülve egy ilyen data poolal való kapcsolatteremtésre. Ebből a célból linked data API-kat kell implementálni, amelyek segítségével az egyes rendszerek a központi adatbázisból származó információkat teszik beemelhetővé az egyes entítások leírásaiba.

Közép- illetve hosszú távon kísérleteket kell folytatni az egyes szoftverkomponensek, programmodulok linked data-alapokra helyezésével kapcsolatban: nem csupán a katalogizálás, hanem a kölcsönzés, gyarapítás, és más funkciók esetében is. Ezeket az áttéréseket minden alkalommal a szóba jöhető külső adatforrások, adatgazdagítási lehetőségek vizsgálata kell kiegészítse, hogy az egyes közgyűjtemények minél inkább részeivé válhassanak az adatok globális hálózatának. Különösképp a katalogizálási munkafolyamatnál fontos, hogy az egyes, kívülről beemelhető adatforrások kellően fel legyenek térképezve, úgy a művek, kifejezési formák stb., mint a tárgyszavak, földrajzi nevek és más adatelemek esetében.

Az entitásalapú szemlélet meghonosítása, a technikai lehetőségek tesztelése, feltérképezése során érdemes nagy hangsúlyt fektetni a már ma is nagy számban hozzáférhető (például GitHub-ról letölthető) open source alkalmazásokra, szoftverkomponensekre, amelyek helyi implementációiról, és az elvégzett kísérletekről a könyvtári szaksajton keresztül érdemes hírt adni.

Az egyes szótárak átjárhatóságát nem csupán a konverziós eszközök kidolgozásával, hanem magán a szemantikus világhálón is biztosítani kell, amelynek egy érdekes fejlesztési iránya lehet az `owl:sameAs` azonossági reláció, illetve a gyengébb szemantikájú SKOS-elemek (lásd a 6.6. fejezetet) alkalmazása nem az entitás-előfordulások URI-jai, hanem a kapcsolatjelölő azonosítók vonatkozásában. Egy ilyen mapping tartalmazhatná az összevetésre kerülő szótárak – például és kiemelten – az RDA- és BIBFRAME-szótár egyes metaadatelemeit, amelyek között azonosságokat lehet kimutatni. Az ilyen megfeleltetéseket nem csupán szoftvermodulokban lehet alkalmazni (például a felhasználók számára történő frontendek hatékony és gyors konfigurálására), hanem gráfok formájában a világhálón is közzétehetők.

Mellékletek

1. melléklet – A BIBFRAME 1.0, illetve a BIBFRAME 2.0 osztályainak és tulajdonságainak listája

BIBFRAME 1.0 osztályok	BIBFRAME 1.0 tulajdonságok	
Agent		classificationStatus
Annotation		classificationTable
Archival	abbreviatedTitle	classificationTableSeq
Arrangement	absorbed	classificationUdc
Audio	absorbedBy	coden
Authority	absorbedInPart	colorContent
Cartography	absorbedInPartBy	componentOf
Category	accessCondition	containedIn
Classification	accessCondition	contains
Collection	accompanies	contentAccessibility
CoverArt	agent	contentCategory
Dataset	annotates	contentsNote
DescriptionAdminInfo	annotationAssertedBy	continuedBy
Electronic	annotationBody	continuedInPartBy
Event	annotationSource	continues
Family	ansi	continuesInPart
HeldItem	arrangement	contributor
HeldMaterial	aspectRatio	copyNote
Identifier	assertionDate	copyrightDate
Instance	audience	coverArt
Integrating	audienceAssigner	coverArtFor
IntendedAudience	authorityAssigner	coverArtThumb
Jurisdiction	authoritySource	creationDate
Language	authorizedAccessPoint	creator
Manuscript	awardNote	creditsNote
Meeting	barcode	custodialHistory
MixedMaterial	carrierCategory	dataSource
Monograph	cartographicAscensionAndDeclination	derivativeOf
MovingImage	cartographicCoordinates	derivedFrom
Multimedia	cartographicEquinox	descriptionAuthentication
MultipartMonograph	cartographicExclusionGRI	descriptionConventions
NotatedMovement	cartographicOuterGRI	descriptionLanguage
NotatedMusic	cartographicProjection	descriptionModifier
Organization	cartographicScale	descriptionOf
Person	cartography	descriptionSource
Place	category	descriptionStatus
Print	categorySource	dimensions
Provider	categoryType	dissertationDegree
Related	categoryValue	dissertationIdentifier
Relator	changeDate	dissertationInstitution
Resource	circulationStatus	dissertationNote
Review	classification	dissertationYear
Serial	classificationAssigner	distribution
StillImage	classificationDdc	doi
Summary	classificationDesignation	duration
TableOfContents	classificationEdition	ean
Tactile	classificationItem	edition
Temporal	classificationLcc	editionResponsibility
Text	classificationNlm	electronicLocator
ThreeDimensionalObject	classificationNumber	enumerationAndChronology
Title	classificationNumberUri	event
Topic	classificationScheme	eventAgent
Work	classificationSpanEnd	eventDate
		eventPlace

expressionOf	languageSource	reproduction
extent	lcOverseasAcq	reproductionPolicy
findingAid	lccn	resourcePart
findingAidNote	legalDate	responsibilityStatement
fingerprint	legalDeposit	retentionPolicy
formDesignation	lendingPolicy	review
format	local	reviewOf
formatOfMusic	manufacture	role
frequency	materialArrangement	separatedFrom
frequencyNote	materialHierarchicalLevel	serialFirstIssue
generationDate	materialOrganization	serialLastIssue
generationProcess	materialPart	series
genre	matrixNumber	shelfMark
geographicCoverageNote	mediaCategory	shelfMarkDdc
graphicScaleNote	mergedToForm	shelfMarkLcc
hasAnnotation	modeOfIssuance	shelfMarkNlm
hasAuthority	musicKey	shelfMarkScheme
hasDerivative	musicMedium	shelfMarkUdc
hasDescription	musicMediumNote	sici
hasEquivalent	musicNumber	soundContent
hasExpression	musicPlate	splitInto
hasInstance	musicPublisherNumber	startOfReview
hasPart	musicVersion	startOfSummary
hdl	nban	stockNumber
heldBy	nbn	strn
holdingFor	notation	studyNumber
identifier	note	subLocation
identifierAssigner	originDate	subject
identifierQualifier	originPlace	subseries
identifierScheme	originalVersion	subseriesOf
identifierStatus	otherEdition	subtitle
identifierValue	otherPhysicalFormat	succeededBy
illustrationNote	partNumber	succeeds
immediateAcquisition	partOf	summary
index	partTitle	summaryOf
instanceOf	performerNote	supersededBy
instanceTitle	place	supersededInPartBy
intendedAudience	postalRegistration	supersedes
isDerivativeOf	precededBy	supersedesInPart
isDescriptionOf	precedes	supplement
isPartOf	preferredCitation	supplementTo
isan	production	supplementaryContentNote
isbn	provider	systemNumber
isbn10	providerDate	tableOfContents
isbn13	providerName	tableOfContentsFor
ismn	providerPlace	temporalCoverageNote
iso	providerRole	title
issn	providerStatement	titleAttribute
issnL	publication	titleQualifier
issueNumber	publisherNumber	titleSource
issuedWith	referenceAuthority	titleStatement
istc	relatedAgent	titleType
iswc	relatedInstance	titleValue
itemId	relatedResource	titleVariation
keyTitle	relatedWork	titleVariationDate
label	relationship	translation
language	relationshipUri	translationOf
languageNote	relator	treatySignator
languageOfPart	relatorRole	unionOf
languageOfPartUri	reportNumber	upc

uri
urn
videorecordingNumber
workTitle

BIBFRAME 2.0 osztályok

AbbreviatedTitle
AccessPolicy
AcquisitionSource
AdminMetadata
Agent
Ansi
AppliedMaterial
Archival
Arrangement
AspectRatio
Audio
AudioIssueNumber
AudioTake
Barcode
BaseMaterial
BookFormat
BroadcastStandard
Capture
Carrier
Cartographic
CartographicDataType
CartographicObjectType
Cartography
Chronology
Classification
ClassificationDdc
ClassificationLcc
ClassificationNlm
ClassificationUdc
Coden
Collection
CollectiveTitle
ColorContent
Content
ContentAccessibility
Contribution
CopyrightNumber
CopyrightRegistration
CoverArt
Dataset
DescriptionAuthentication
DescriptionConventions
DigitalCharacteristic
Dissertation
DissertationIdentifier
Distribution
Doi
Ean
Electronic
Emulsion
EncodedBitrate
EncodingFormat
Enumeration
EnumerationAndChronology

Event
Extent
Family
FileSize
FileType
Fingerprint
FontSize
Frequency
Generation
GenerationProcess
GenreForm
GeographicCoverage
GrooveCharacteristic
Gtin14Number
Hdl
Identifier
Illustration
ImmediateAcquisition
Instance
IntendedAudience
Isan
Isbn
Ismn
Isni
Iso
Isrc
Issn
IssnL
Issuance
Istc
Iswc
Item
Jurisdiction
KeyTitle
Language
Layout
Lcen
LcOverseasAcq
Local
Manufacture
Manuscript
MatrixNumber
Media
Meeting
MixedMaterial
Mount
MovementNotation
MovingImage
Multimedia
MusicDistributorNumber
MusicEnsemble
MusicFormat
MusicInstrument
MusicMedium
MusicNotation
MusicPlate
MusicPublisherNumber
MusicVoice
Nbn
NotatedMovement

NotatedMusic
Notation
Note
Object
ObjectCount
Organization
ParallelTitle
Person
Place
PlaybackChannels
PlaybackCharacteristic
PlayingSpeed
Polarity
PostalRegistration
PresentationFormat
Print
Production
ProductionMethod
Projection
ProjectionCharacteristic
ProjectionSpeedProvisionActi
vity
Publication
PublisherNumber
RecordingMedium
RecordingMethod
ReductionRatio
RegionalEncoding
ReportNumber
Resolution
RetentionPolicy
Review
Role
Scale
Script
ShelfMark
ShelfMarkDdc
ShelfMarkLcc
ShelfMarkNlm
ShelfMarkUdc
Sici
SoundCharacteristic
SoundContent
Source
Status
StillImage
StockNumber
Strn
StudyNumber
Sublocation
Summary
SupplementaryContent
SystemRequirement
TableOfContents
Tactile
TactileNotation
TapeConfig
Temporal
Text
Title

Topic	descriptionConventions	layout
TrackConfig	descriptionLanguage	legalDate
Unit	descriptionModifier	mainTitle
Upc	digitalCharacteristic	media
Urn	dimensions	mergedToForm
UsageAndAccessPolicy	dissertation	mergerOf
UsePolicy	duration	mount
VariantTitle	edition	musicFormat
VideoCharacteristic	editionEnumeration	musicKey
VideoFormat	editionStatement	musicMedium
VideoRecordingNumber	electronicLocator	musicOpusNumber
Work	emulsion	musicSerialNumber
BIBFRAME 2.0	ensemble	musicThematicNumber
tulajdonságok	ensembleType	natureOfContent
absorbed	enumerationAndChronology	notation
absorbedBy	equinox	note
accompaniedBy	eventContent	noteType
accompanies	eventContentOf	organization
acquisitionSource	exclusionGRing	originalVersion
acquisitionTerms	expressionOf	originalVersionOf
adminMetadata	extent	originDate
agent	findingAid	originPlace
appliedMaterial	findingAidOf	otherEdition
arrangement	firstIssue	otherPhysicalFormat
ascensionAndDeclination	fontSize	outerGRing
aspectRatio	frequency	part
assigner	generation	partName
awards	generationDate	partNumber
baseMaterial	generationProcess	partOf
bookFormat	genreForm	pattern
capture	geographicCoverage	physicalLocation
carrier	grantingInstitution	place
cartographicAttributes	hasDerivative	polarity
changeDate	hasEquivalent	precededBy
classification	hasExpression	preferredCitation
classificationPortion	hasInstance	productionMethod
code	hasItem	projection
colorContent	hasPart	projectionCharacteristic
content	hasReproduction	provisionActivity
contentAccessibility	hasSeries	provisionActivityStatement
continuedBy	hasSubseries	qualifier
continuedInPartBy	heldBy	reductionRatio
continues	hierarchicalLevel	referencedBy
continuesInPart	historyOfWork	references
contribution	identifiedBy	relatedTo
coordinates	identifies	replacedBy
copyrightDate	illustrativeContent	replacementOf
copyrightRegistration	immediateAcquisition	reproductionOf
count	index	responsibilityStatement
coverArt	indexOf	review
creationDate	instanceOf	role
credits	instrument	scale
custodialHistory	instrumentalType	schedulePart
dataSource	intendedAudience	separatedFrom
date	issuance	seriesEnumeration
degree	issuedWith	seriesOf
derivativeOf	itemOf	seriesStatement
derivedFrom	itemPortion	shelfMark
descriptionAuthentication	language	soundCharacteristic
	lastIssue	soundContent

source
spanEnd
splitInto
status
subject
sublocation
subseriesEnumeration
subseriesOf
subseriesStatement
subtitle

succeededBy
summary
supplement
supplementaryContent
supplementTo
systemRequirement
table
tableOfContents
tableSeq
temporalCoverage

title
translation
translationOf
unit
usageAndAccessPolicy
variantType
version
videoCharacteristic
voice
voiceType

Általános megjegyzések

Sárga: a MARC-rekordokból származó specifikus értékek helye

Szürke: generált entitás-URI-k

Kék: opcionális szakaszok

MŰVEK (Work)

```
<bf:Work rdf:about="XYZ work URI">
```

```
<bf:adminMetadata>
```

```
<bf:AdminMetadata>
```

<bf:generationProcess> - ezt a szekciót a LoC konvertere hozza létre, tulajdonképp csak egy megjegyzés, elhagyható

```
<bf:GenerationProcess>
```

```
<rdfs:label>DLC marc2bibframe2 v1.4.0-SNAPSHOT</rdfs:label>
```

```
</bf:GenerationProcess>
```

```
</bf:generationProcess>
```

```
<bf:status>
```

```
<bf:Status>
```

```
<bf:code>000 05. karakterhely</bf:code>
```

```
</bf:Status>
```

```
</bf:status>
```

```
<bf:encodingLevel>
```

```
<bf:EncodingLevel>
```

```
<bf:code>000 17. karakterhely</bf:code>
```

```
</bf:EncodingLevel>
```

```
</bf:encodingLevel>
```

```
<bf:descriptionConventions>
```

```
<bf:DescriptionConventions>
```

```
<bf:code>000 18. karakterhely alapján</bf:code>
```

```
</bf:DescriptionConventions>
```

```
</bf:descriptionConventions>
```

```
<bf:identifiedBy>
```

```
<bf:Local>
```

```
<rdf:value>001 értéke</rdf:value>
```

```
</bf:Local>
```

```
</bf:identifiedBy>
```



```
<bf:source>
  <bf:Source>
    <bf:code>003 értéke</bf:code>
  </bf:Source>
</bf:source>
```

```
<bf:changeDate rdf:datatype="005 értéke: éééé-hh-nnTóó:pp:mp formátumban
</bf:changeDate>
<bf:creationDate rdf:datatype="008 00-05. karakterhely éééé-hh-nn formátumra
konvertálva</bf:creationDate>
```

```
<bf:source>
  <bf:Source>
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Agent" />
    <rdfs:label>040$ a értéke</rdfs:label>
  </bf:Source>
</bf:source>
```

```
<bf:source>
  <bf:Source>
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Agent" />
    <rdfs:label>040$ c értéke</rdfs:label>
  </bf:Source>
</bf:source>
```

```
<bf:descriptionLanguage>
  <bf:Language rdf:about="http://id.loc.gov/vocabulary/languages/040 $ b értéke" />
</bf:descriptionLanguage>
```

```
<bf:descriptionConventions>
  <bf:DescriptionConventions
rdf:about="http://id.loc.gov/vocabulary/descriptionConventions/040 $ e értéke" />
  </bf:descriptionConventions>
```

```
</bf:AdminMetadata>
</bf:adminMetadata>
```

```
<rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/000 06. karakterhely alapján" />
```

```
<bf:colorContent>
  <bf:ColorContent rdf:about="http://id.loc.gov/vocabulary/mcolor/007 03. karakterhely
alapján">
    <rdfs:label>007 03. karakterhely alapján</rdfs:label>
  </bf:ColorContent>
</bf:colorContent>
```

```
<bf:language>
  <bf:Language rdf:about="http://id.loc.gov/vocabulary/languages/008 35-37. karakterhely" />
</bf:language>
```

```
<bf:genreForm>
```

```
<bf:GenreForm rdf:about="http://id.loc.gov/vocabulary/marcgt/008 33. karakterhely
alapján">
  <rdfs:label>008 33. karakterhely alapján</rdfs:label>
</bf:GenreForm>
</bf:genreForm>
```

Ha 041 első indikátora 1:

```
<bf:note>
  <bf:Note>
    <rdfs:label>Includes translation</rdfs:label>
  </bf:Note>
</bf:note>
```

```
<bf:language>
  <bf:Language>
    <bf:identifiedBy>
      <bf:Identifier>
        <rdf:value rdf:resource="http://id.loc.gov/vocabulary/languages/041 $a értéke" />
        <bf:source>
          <bf:Source rdf:about="http://id.loc.gov/vocabulary/languages" />
        </bf:source>
      </bf:Identifier>
    </bf:identifiedBy>
  </bf:Language>
</bf:language>
```

Ha van 041\$h:

```
<bf:language>
  <bf:Language>
    <bf:part>original</bf:part>
    <bf:identifiedBy>
      <bf:Identifier>
        <rdf:value rdf:resource="http://id.loc.gov/vocabulary/languages/041 $h értéke" />
        <bf:source>
          <bf:Source rdf:about="http://id.loc.gov/vocabulary/languages" />
        </bf:source>
      </bf:Identifier>
    </bf:identifiedBy>
  </bf:Language>
</bf:language>
```

```
<bf:contribution>
  <bf:Contribution>
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bflc/PrimaryContribution" />
    <bf:agent>
      <bf:Agent rdf:about="XYZ Agent URI">
        <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Person" />
        <bflc:name00MatchKey>100 $abcdjq értékei indikátor és almezőjelek
nélkül</bflc:name00MatchKey>
```

```
<bf:primaryContributorName00MatchKey>100 $abcdjq értékei indikátor és almezőjelek  
nélkül</bf:primaryContributorName00MatchKey>  
<bf:name00MarcKey>a 100-as mező tartalma, hívójellel, indikátorral, almezőjelekkel és  
értékekkel</bf:name00MarcKey>  
<rdfs:label>100 $abcdjq értékei indikátor és almezőjelek nélkül </rdfs:label>  
<bf:identifiedBy>  
<bf:Identifier>  
<rdf:value>100 $0 értéke</rdf:value>  
</bf:Identifier>  
</bf:identifiedBy>  
</bf:Agent>  
</bf:agent>  
<bf:role>  
<bf:Role>  
<rdfs:label>100 $e értéke</rdfs:label>  
<bf:relatorMatchKey>100 $e értéke</bf:relatorMatchKey>  
</bf:Role>  
</bf:role>  
<bf:role>  
<bf:Role rdf:about="http://id.loc.gov/vocabulary/relators/100 $4 értéke" />  
</bf:role>  
</bf:Contribution>  
</bf:contribution>
```

```
<rdfs:label>240 $adfgklmnoprs értékei indikátor és almezőjelek nélkül</rdfs:label>  
<bf:title>  
<bf:Title>  
<bf:title40MatchKey>240 $adfgklmnoprs értékei indikátor és almezőjelek nélkül  
</bf:title40MatchKey>  
<bf:title40MarcKey>a 240-es mező tartalma, hívójellel, indikátorral, almezőjelekkel és  
értékekkel</bf:title40MarcKey>  
<rdfs:label>240 $adfgklmnoprs értékei indikátor és almezőjelek nélkül</rdfs:label>  
<bf:titleSortKey>240 $adfgklmnoprs értékei indikátor és almezőjelek nélkül, levonva az  
indikátorban jelzett kihagyandó karaktereket</bf:titleSortKey>  
<bf:mainTitle>240 $a értéke</bf:mainTitle>  
</bf:Title>  
</bf:title>  
<bf:language>  
<bf:Language>  
<rdfs:label>240 $l értéke</rdfs:label>  
</bf:Language>  
</bf:language>  
<bf:identifiedBy>  
<bf:Identifier>  
<rdf:value>240 $9 értéke</rdf:value>  
</bf:Identifier>  
</bf:identifiedBy>  
<bf:translationOf>  
<bf:Work rdf:about="XYZ Work URI - az eredeti műé">  
<rdfs:label>240 $adfgklmnoprs értékei indikátor és almezőjelek nélkül ($l nem  
szerepel!)</rdfs:label>  
<bf:title>
```

```
<bf:Title>
  <bflc:title40MatchKey>240 $adfgkmnopr értékei indikátor és almezőjelek nélkül ($I nem szerepel!)</bflc:title40MatchKey>
  <bflc:title40MarcKey>a 240-es mező tartalma, hívójellel, indikátorral, almezőjelekkel és értékekkel</bflc:title40MarcKey>
  <rdfs:label>240 $adfgkmnopr értékei indikátor és almezőjelek nélkül ($I nem szerepel!)</rdfs:label>
  <bflc:titleSortKey>240 $adfgkmnopr értékei indikátor és almezőjelek nélkül, levonva az indikátorban jelzett kihagyandó karaktereket ($I nem szerepel!)</bflc:titleSortKey>
  <bf:mainTitle>240 $a értéke</bf:mainTitle>
</bf:Title>
</bf:title>
<bf:identifiedBy>
  <bf:Identifier>
    <rdf:value>240 $9 értéke</rdf:value>
  </bf:Identifier>
</bf:identifiedBy>
</bf:Work>
</bf:translationOf>

<bf:content>
  <bf:Content rdf:about="336 $0 értéke – id.loc.gov URI">
    <rdfs:label>336 $a értéke</rdfs:label>
    <bf:identifiedBy>
      <bf:Identifier>
        <rdf:value rdf:resource="336 $0 értéke – RDA Registry URI" />
      </bf:Identifier>
    </bf:identifiedBy>
    <bf:source>
      <bf:Source>
        <rdfs:label>336 $2 értéke</rdfs:label>
      </bf:Source>
    </bf:source>
  </bf:Content>
</bf:content>

<bf:genreForm>
  <bf:GenreForm rdf:about="XYZ GenreForm URI">
    <rdf:type rdf:resource="http://www.loc.gov/mads/rdf/v1#GenreForm" />
    <rdfs:label>655 $a értéke</rdfs:label>
    <madsrdf:authoritativeLabel>655 $a értéke</madsrdf:authoritativeLabel>
  </bf:GenreForm>
</bf:genreForm>

<bf:contribution>
  <bf:Contribution>
    <bf:agent>
      <bf:Agent rdf:about="XYZ Agent URI">
        <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Person" />
        <bflc:name00MatchKey>700 $abcdjq értékei indikátor és almezőjelek nélkül</bflc:name00MatchKey>
```

```

    <bf:Name00MarcKey>a 700-as mező tartalma, hívójellel, indikátorral, almezőjelekkel és
értékekkel</bf:Name00MarcKey>
    <rdfs:label>700 $abcdjq értékei indikátor és almezőjelek nélkül</rdfs:label>
  </bf:Agent>
</bf:agent>
  <bf:role>
    <bf:Role>
      <rdfs:label>700 $e értéke</rdfs:label>
      <bf:relatorMatchKey>700 $e értéke</bf:relatorMatchKey>
    </bf:Role>
  </bf:role>
  <bf:role>
    <bf:Role rdf:about="http://id.loc.gov/vocabulary/relators/700 $4 értéke"/>
  </bf:role>
</bf:Contribution>
</bf:contribution>

<bf:contribution>
  <bf:Contribution>
    <bf:agent>
      <bf:Agent rdf:about="XYZ Agent URI">
        <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Organization" />
        <bf:Name10MatchKey>710 $abcdng értékei indikátor és almezőjelek
nélkül</bf:Name10MatchKey>
        <bf:Name10MarcKey>a 710-as mező tartalma, hívójellel, indikátorral, almezőjelekkel és
értékekkel</bf:Name10MarcKey>
        <rdfs:label>710$abcdng értékei indikátor és almezőjelek nélkül</rdfs:label>
      </bf:Agent>
    </bf:agent>
    <bf:role>
      <bf:Role>
        <rdfs:label>710 $e értéke</rdfs:label>
        <bf:relatorMatchKey>710$e értéke</bf:relatorMatchKey>
      </bf:Role>
    </bf:role>
    <bf:role>
      <bf:Role rdf:about="http://id.loc.gov/vocabulary/relators/710 $4 értéke" />
    </bf:role>
  </bf:Contribution>
</bf:contribution>

<bf:hasInstance>
  <bf:Instance rdf:about="XYZ Instance URI">
    <bf:contribution>
      <bf:PrimaryContribution>
        <bf:agent>
          <bf:Agent>
            <rdfs:label>776 $a értéke</rdfs:label>
          </bf:Agent>
        </bf:agent>
      </bf:PrimaryContribution>
    </bf:contribution>
  </bf:hasInstance>

```

```
<bf:relationship>
  <bf:Relationship>
    <bf:relation>
      <bf:Relation>
        <rdfs:label>776 $i értéke</rdfs:label>
      </bf:Relation>
    </bf:relation>
    <bf:relatedTo rdf:resource="XYZ kapcsolódó Instance URI" />
  </bf:Relationship>
</bf:relationship>
<bf:identifiedBy>
  <bf:Identifier>
    <rdfs:value>776 $w értéke</rdfs:value>
  </bf:Identifier>
</bf:identifiedBy>
<bf:provisionActivityStatement>776 $d értéke</bf:provisionActivityStatement>
  <bf:extent>
    <bf:Extent>
      <rdfs:label>776 $h értéke</rdfs:label>
    </bf:Extent>
  </bf:extent>
  <bf:title>
    <bf>Title>
      <rdfs:label>776 $t értéke</rdfs:label>
    </bf>Title>
  </bf:title>
  <bf:instanceOf rdf:resource="XYZ Work ID" />
</bf:Instance>
</bf:hasInstance>

<bf:hasInstance rdf:resource="XYZ Instance URI #2" />

</bf:Work>
```

MEGJELENÉSI FORMÁK (Instance)

```
<bf:Instance rdf:about="XYZ Instance URI">
```

```
  <bf:issuance>
```

```
    <bf:Issuance rdf:about="http://id.loc.gov/vocabulary/issuance/000 07. karakterhely alapján"
```

```
  />
```

```
  </bf:issuance>
```

```
<rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/007 01. karakterhely alapján" />
```

```
<bf:digitalCharacteristic>
```

```
  <bf:DigitalCharacteristic>
```

```
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bflc/ImageBitDepth" />
```

```
    <rdfs:label>007 06-08. karakterhely</rdfs:label>
```

```
  </bf:DigitalCharacteristic>
```

```
</bf:digitalCharacteristic>
```

```
<bf:digitalCharacteristic>
```

```
  <bf:DigitalCharacteristic>
```

```
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/EncodingFormat" />
```

```
    <rdfs:label>09. karakterhely alapján</rdfs:label>
```

```
  </bf:DigitalCharacteristic>
```

```
</bf:digitalCharacteristic>
```

```
<bf:provisionActivity>
```

```
  <bf:ProvisionActivity>
```

```
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Publication" />
```

```
    <bf:date rdf:datatype="http://id.loc.gov/datatypes/edtf">008 07-10. karakterhely</bf:date>
```

```
    <bf:place>
```

```
      <bf:Place rdf:about="http://id.loc.gov/vocabulary/countries/008 15-17. karakterhely" />
```

```
    </bf:place>
```

```
  </bf:ProvisionActivity>
```

```
</bf:provisionActivity>
```

```
<bf:illustrativeContent>
```

```
  <bf:Illustration rdf:about="http://id.loc.gov/vocabulary/millus/008 18-21. karakterhely">
```

```
    <rdfs:label>008 18-21. karakterhely alapján</rdfs:label>
```

```
  </bf:Illustration>
```

```
</bf:illustrativeContent>
```

```
<bf:identifiedBy>
```

```
  <bf:Isbn>
```

```
    <rdf:value>020 $a értéke</rdf:value>
```

```
  </bf:Isbn>
```

```
</bf:identifiedBy>
```

```
<rdfs:label>245 $afgknps értéke</rdfs:label>
```

```
<bf:title>
```

```
  <bf:Title>
```

```
    <rdfs:label>245 $afgknps értéke</rdfs:label>
```

```
<bf:TitleSortKey>245 $afgknps értéke, levonva az indikátorban jelzett kihagyandó
karaktereket</bf:TitleSortKey>
<bf:mainTitle>245 $a értéke</bf:mainTitle>
<bf:subtitle>245 $b értéke</bf:subtitle>
</bf:Title>
</bf:title>
<bf:responsibilityStatement>245 $c értéke</bf:responsibilityStatement>

<bf:editionStatement>250 $a értéke</bf:editionStatement>
```

Ha a 264 második indikátora 1:

```
<bf:provisionActivity>
  <bf:ProvisionActivity>
    <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Publication" />
    <bf:place>
      <bf:Place>
        <rdfs:label>264 $a értéke</rdfs:label>
      </bf:Place>
    </bf:place>
    <bf:agent>
      <bf:Agent>
        <rdfs:label>264 $b értéke</rdfs:label>
      </bf:Agent>
    </bf:agent>
    <bf:date>264 $c értéke</bf:date>
  </bf:ProvisionActivity>
</bf:provisionActivity>
```

Ha a 264 második indikátora 3:

```
<bf:Instance rdf:about="XYZ">
  <bf:provisionActivity>
    <bf:ProvisionActivity>
      <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Manufacture" />
      <bf:place>
        <bf:Place>
          <rdfs:label>264 $a értéke</rdfs:label>
        </bf:Place>
      </bf:place>
      <bf:agent>
        <bf:Agent>
          <rdfs:label>264 $b értéke</rdfs:label>
        </bf:Agent>
      </bf:agent>
      <bf:date>264 $c értéke</bf:date>
    </bf:ProvisionActivity>
  </bf:provisionActivity>
```

Ha a 264 második indikátora 3:

<bf:copyrightDate>264 \$c értéke</bf:copyrightDate>

<bf:provisionActivityStatement>264 \$a \$b és \$c értéke</bf:provisionActivityStatement>

<bf:extent>
<bf:Extent>
<rdfs:label>300 \$a értéke</rdfs:label>
</bf:Extent>
</bf:extent>

<bf:note>
<bf:Note>
<bf:noteType>Physical details</bf:noteType>
<rdfs:label>300 \$b értéke</rdfs:label>
</bf:Note>
</bf:note>

<bf:dimensions>300 \$c értéke</bf:dimensions>

<bf:media> - *értékének megállapításához lásd a 007-es mező 01. karakterhelyét is*

<bf:Media rdf:about="337 \$0 értéke – id.loc.gov URI">
<rdfs:label>337 \$a értéke</rdfs:label>
<bf:identifiedBy>
<bf:Identifier>
<rdf:value rdf:resource="337 \$0 értéke – RDA Registry URI" />
</bf:Identifier>
</bf:identifiedBy>
<bf:source>
<bf:Source>
<rdfs:label>337 \$2 értéke</rdfs:label>
</bf:Source>
</bf:source>
</bf:Media>
</bf:media>

<bf:carrier> - *értékének megállapításához lásd a 007-es mező 02. karakterhelyét is, továbbá a 008-as mező 23. karakterhelyét*

<bf:Carrier rdf:about="338 \$0 értéke – id.loc.gov URI">
<rdfs:label>338 \$a értéke</rdfs:label>
<bf:identifiedBy>
<bf:Identifier>
<rdf:value rdf:resource="338 \$0 értéke – RDA Registry URI" />
</bf:Identifier>
</bf:identifiedBy>
<bf:source>
<bf:Source>
<rdfs:label>338 \$2 értéke</rdfs:label>
</bf:Source>
</bf:source>

</bf:Carrier>
</bf:carrier>

<bf:otherPhysicalFormat rdf:resource="XYZ kapcsolódó Instance URI a 776-os mező alapján"/>

<bf:generation>
 <bf:Generation>
 <rdfs:label>340 \$j értéke</rdfs:label>
 <bf:source>
 <bf:Source>
 <rdfs:label>340 \$2 értéke</rdfs:label>
 </bf:Source>
 </bf:source>
 </bf:Generation>
</bf:generation>

<bf:digitalCharacteristic>
 <bf:FileType>
 <rdfs:label>347 \$a értéke</rdfs:label>
 <bf:source>
 <bf:Source>
 <rdfs:label>347 \$2 értéke</rdfs:label>
 </bf:Source>
 </bf:source>
 </bf:FileType>
</bf:digitalCharacteristic>
<bf:digitalCharacteristic>
 <bf:EncodingFormat>
 <rdfs:label>347 \$b értéke</rdfs:label>
 </bf:EncodingFormat>
</bf:digitalCharacteristic>
<bf:digitalCharacteristic>
 <bf:FileSize>
 <rdfs:label>347 \$c értéke</rdfs:label>
 </bf:FileSize>
</bf:digitalCharacteristic>

<bf:tableOfContents>
 <bf:TableOfContents>
 <rdfs:label>505 \$g és \$t értéke</rdfs:label>
 </bf:TableOfContents>
</bf:tableOfContents>

<bf:usageAndAccessPolicy>
 <bf:UsageAndAccessPolicy>
 <rdfs:label>506 \$a és \$f értéke</rdfs:label>
 </bf:UsageAndAccessPolicy>
</bf:usageAndAccessPolicy>

<bf:note>
 <bf>Note>
 <rdfs:label>588 \$a értéke</rdfs:label>

```
<bf:noteType>description source</bf:noteType>
</bf:Note>
</bf:note>
```

```
<bf:hasSeries>
  <bf:Instance>
    <rdfs:label>490 $a értéke</rdfs:label>
    <bf:seriesStatement>490 $a és $v értéke elválasztó nélkül</bf:seriesStatement>
    <bf:seriesEnumeration>490 $v értéke</bf:seriesEnumeration>
    <bf:instanceOf>
      <bf:Work rdf:about="XYZ Work ID">
        <bf:contribution>
          <bf:Contribution>
            <bf:agent>
              <bf:Agent rdf:about="XYZ Agent ID">
                <rdf:type rdf:resource="http://id.loc.gov/ontologies/bibframe/Person" />
                <bflc:name00MatchKey>800 $abcdjq értékei indikátor és almezőjelek
nélkül</bflc:name00MatchKey>
                <bflc:name00MarcKey>a 800-as mező tartalma, hívójellel, indikátorral,
almezőjelekkel és értékekkel</bflc:name00MarcKey>
                <rdfs:label>800 $abcdjq értékei</rdfs:label>
                </bf:Agent>
              </bf:agent>
              <bf:role>
                <bf:Role rdf:about="http://id.loc.gov/vocabulary/relators/ctb" />
                </bf:role>
              </bf:Contribution>
            </bf:contribution>
            <rdfs:label>800 $tfgklmnoprs értékei elválasztó nélkül</rdfs:label>
            <bf:title>
              <bf>Title>
                <bflc:title00MatchKey>800 $tfgklmnoprs értékei elválasztó
nélkül</bflc:title00MatchKey>
                <bflc:title00MarcKey>a 800-as mező tartalma, hívójellel, indikátorral, almezőjelekkel és
értékekkel</bflc:title00MarcKey>
                <rdfs:label>800 $tfgklmnoprs értékei elválasztó nélkül</rdfs:label>
                <bflc:titleSortKey>800 $tfgklmnoprs értéke, levonva az indikátorban jelzett kihagyandó
karaktereket</bflc:titleSortKey>
                <bf:mainTitle>800 $t értéke</bf:mainTitle>
              </bf>Title>
            </bf:title>
            <bf:natureOfContent>800 $k értéke</bf:natureOfContent>
            <bf:genreForm>
              <bf:GenreForm>
                <rdfs:label>800 $k értéke</rdfs:label>
              </bf:GenreForm>
            </bf:genreForm>
          </bf:Work>
        </bf:instanceOf>
      </bf:Instance>
    </bf:hasSeries>
```

```

<bf:hasSeries>
  <bf:Instance>
    <rdfs:label>490 $a értéke</rdfs:label>
    <bf:seriesStatement>490 $a és $v értéke</bf:seriesStatement>
    <bf:seriesEnumeration>830 $v értéke</bf:seriesEnumeration>
  <bf:identifiedBy>
    <bf:Issn>
      <rdf:value>490$x értéke</rdf:value>
    </bf:Issn>
  </bf:identifiedBy>
  <bf:instanceOf>
    <bf:Work rdf:about="XYZ Work URI">
      <rdfs:label>830 $a értéke</rdfs:label>
      <bf:title>
        <bf>Title>
          <bf:title30MatchKey>830 $adfgklmnops értéke elválasztó
nélkül</bf:title30MatchKey>
          <bf:title30MarcKey>a 830-as mező tartalma, hívójellel, indikátorral, almezőjelekkel és
értékekkel</bf:title30MarcKey>
          <rdfs:label>830 $adfgklmnops értékei</rdfs:label>
          <bf:titleSortKey>830 $adfgklmnops értéke, levonva az indikátorban jelzett
kihagyandó karaktereket</bf:titleSortKey>
          <bf:mainTitle>830 $a értéke</bf:mainTitle>
        </bf>Title>
      </bf:title>
      <bf:identifiedBy>
        <bf:Issn>
          <rdf:value>490$x értéke</rdf:value>
        </bf:Issn>
      </bf:identifiedBy>
    </bf:Work>
  </bf:instanceOf>
</bf:Instance>
</bf:hasSeries>

```

```

<bf:hasItem>
<bf:Item rdf:about="XYZ Item URI">
  <bf:note>
    <bf>Note>
      <bf:noteType>action</bf:noteType>
    </bf>Note>
  </bf:note>
  <bf:itemOf rdf:resource="XYZ Instance URI" />
</bf:Item>
</bf:hasItem>

```

```

<bf:hasItem>
  <bf:Item rdf:about="XYZ Item URI">
    <bf:heldBy>
      <bf:Agent rdf:about="http://id.loc.gov/vocabulary/organizations/850 $a értéke" />
    </bf:heldBy>
    <bf:itemOf rdf:resource="XYZ Instance URI" />
  </bf:Item>

```

</bf:hasItem>

<bf:hasItem>

<bf:Item rdf:about="XYZ Item URI">

<bf:heldBy>

<bf:Agent rdf:about="http://id.loc.gov/vocabulary/organizations/852 \$a értéke" />

</bf:heldBy>

<bf:itemOf rdf:resource="XYZ Instance URI" />

</bf:Item>

</bf:hasItem>

<bf:hasItem>

<bf:Item rdf:about="XYZ Item URI">

<bf:electronicLocator rdf:resource="856 \$u értéke" />

<bf:itemOf rdf:resource="XYZ Instance URI" />

</bf:Item>

</bf:hasItem>

</bf:Instance>

3. melléklet – A PIM MARC/XML-ben található adatmezők és almezők statisztikája:
MARCEdittel generált jelentés

Field	Subfield	In Records	Total
000		21826	21826
001		21826	21826
002		21826	21826
005		21826	21826
008		21826	21826
011		3	3
	\$c		3
	\$a		3
020		1012	1025
	\$a		1025
022		43	44
	\$a		44
040		3816	3816
	\$d		3
	\$b		3621
	\$c		3
	\$a		2894
041		2818	2818
	\$a		2818
044		3	3
	\$a		3
080		2	3
	\$a		3
090		21826	21826
	\$a		22253
100		17394	17394
	\$b		5
	\$c		503
	\$a		17394
	\$g		1603
	\$d		16401
	\$j		17249
	\$m		74
110		2	2
	\$a		2
240		585	585
	\$b		46
	\$a		585
245		21826	21826
	\$b		12796
	\$c		16854

	\$7		9
	\$a		21826
	\$e		17694
	\$8		7
250		1579	1579
	\$a		1579
259		11120	11120
	\$q		11120
	\$Q		11120
260		21805	21805
	\$b		22924
	\$c		21759
	\$a		22487
	\$f		14549
	\$e		12929
300		21632	21745
	\$b		7573
	\$c		21614
	\$a		21743
	\$e		233
	\$k		60
	\$h		90
	\$i		6
306		1	1
	\$a		1
362		8	8
	\$a		8
440		6404	6489
	\$b		886
	\$a		6487
	\$v		3313
	\$u		2635
	\$9		74
	\$z		2858
	\$x		1027
	\$p		78
	\$n		50
480		204	205
	\$v		1
	\$a		205
490		25	25
	\$a		25
500		3904	4441
	\$a		4441
503		2	2

	\$a		2
504		2555	2597
	\$a		2597
505		339	382
	\$a		382
562		8258	8258
	\$a		9352
588		21041	21245
	\$a		21245
592		440	441
	\$a		441
600		3411	4696
	\$b		89
	\$c		397
	\$a		4696
	\$g		1073
	\$d		4609
	\$t		59
	\$j		4515
	\$m		75
610		527	590
	\$d		53
	\$b		19
	\$c		361
	\$a		590
611		123	125
	\$b		71
	\$c		29
	\$a		125
	\$d		110
	\$n		1
630		240	274
	\$b		96
	\$a		274
652		11	11
	\$a		11
700		18945	51194
	\$b		3
	\$a		51172
	\$4		49816
	\$g		1689
	\$d		38723
	\$j		50973
	\$c		313
	\$m		315

710		805	856
	\$b		129
	\$a		855
	\$4		829
	\$c		332
	\$d		134
711		3	3
	\$a		3
730		13	22
	\$b		2
	\$a		22
740		727	1764
	\$b		221
	\$e		274
	\$c		395
	\$a		1761
741		411	508
	\$b		264
	\$a		503
743		218	218
	\$b		91
	\$a		202
856		18	19
	\$3		18
	\$f		4
	\$u		11
910		1	5
	\$a		5

Ábrajegyzék

1. **ábra** – A XOBIS főelemei és azok alapvető kapcsolatai
2. **ábra** – A XOBIS séma strukturális felépítése
3. **ábra** – Az FRBR első entitáscsoportja és elsődleges kapcsolataik
4. **ábra** – A Barbara Tillett által meghatározott bibliográfiai kapcsolatok. Számos esetben a származék azonos műnek, más esetekben már új műnek minősül
5. **ábra** – Az LRM-ben definiált 11 entitás és kapcsolataik
6. **ábra** – Időszaki kiadványok modellezése az LRM-ben
7. **ábra** – A Primo normalizált XML FRBR-kulcsokat leíró részlete
8. **ábra** – A RIMMF szoftver kifejezésforma-leíró űrlapja
9. **ábra** – Az adatközzététel öt csillagos lépcsője
10. **ábra** – Tripletok ábrázolása gráfként
11. **ábra** – A BIBFLOW projekt szakértői által javasolt áttérési menetrend
12. **ábra** – A tripletok nem közvetlenül a felhasználónak készülnek – szükség van az informatikus szakemberre is
13. **ábra** – A keresett személyhez kapcsolódó más személyek, illetve tárgyszavak megjelenítése az Innovative Inspire „context wheel”-jén
14. **ábra** – „Műtípusok” keresése a Qulto Semantics felületén
15. **ábra** – Entitáskapcsolatok MARC-alapú megjelenítése a Qulto Semantics-ban
16. **ábra** – Tudáspanel megjelenítése a SearchWorks felületén elvégzett kereséshez kapcsolódóan. A külső adatforrás a Wikidata
17. **ábra** – A MARCEdit szoftver BIBFRAME-konfigurációt paraméterező ablaka
18. **ábra** – Mű metaadatelemeit tartalmazó űrlap a BIBFRAME Editor felületén
19. **ábra** – Megnövelt felhasználói élményt kínáló, továbbfejlesztett BIBFRAME Editor-felület
20. **ábra** – A BIBFRAME profilszerkesztő felület részlete az adminisztrációs metaadatelemek beviteli mezőivel
21. **ábra** – A BIBFRAME profilszerkesztő felület részlete az egyes metaadatelemek paraméterezésére szolgáló beviteli mezőkkel
22. **ábra** – Metaadat-beviteli mezők a Sinopia felületén
23. **ábra** – A közreműködő adatainak rögzítése a LIBRIS egy leíró űrlapján
24. **ábra** – A LIBRIS felületén elkészült leírás letöltése és megjelenítése MARC 21-ben
25. **ábra** – Egy tulajdonság legfontosabb adatainak megjelenítése a LIBRIS szótárának információs weboldalán

- 26. ábra** – A RERO integrált könyvtári platform kezelőfelülete 2022. áprilisi állapot szerint
- 27. ábra** – A RERO adatbeviteli űrlapjának egy részlete könyvjellegű bibliográfiai források esetén
- 28. ábra** – Ágens rögzítési módjának paraméterezése
- 29. ábra** – A példányadatok megadására szolgáló űrlap a RERO rendszerben
- 30. ábra** – A RERO-rendszer által előállított, JSON formátumban megjelenített leírás részletei
- 31. ábra** – Metaadatelemek az RDA szótárkódolási sémáiból származó értékekkel
- 32. ábra** - Az összeszámlálási jelentés egy részlete

Rövidítések jegyzéke

AACR2 – Anglo-American Cataloging Rules, 2nd Edition

AP – Alkalmazási profil

API – Alkalmazás-programozási interfész

BDML – Bibliographic Description Markup Language

BIBFRAME – Bibliographic Framework

BIBO – Bibliographic Ontology

CHO – Cultural Heritage Object

CIDOC-CRM – CIDOC Conceptual Reference Model

DC – Dublin Core

DNB – Deutsche Nationalbibliothek

DTD – Dokumentumtípus-definíció

EAD – Encoded Archival Description

EDM – Europeana Data Model

FaBiO – FRBR-aligned Bibliographic Ontology

FOAF – Friend of a Friend

FRAD - Functional Requirements for Authority Data

FRBR - Functional Requirements for Bibliographic Records

FRBRoo – Objektum-orientált megközelítésű FRBR

FRSAD - Functional Requirements for Subject Authority Data

GND – Gemeinsame Normdatei

HTML – Hypertext Markup Language

HTTP – Hypertext Transfer Protocol

ICP – International Cataloging Principles

IFLA - International Federation of Library Associations and Institutions

ISBD – International Standard Bibliographic Description

JSON – JavaScript Object Notation

JSON-LD – JSON for Linking Data

LCNAF – Library of Congress National Authority File

LD4P – Linked Data for Production

LIDO - Lightweight Information Describing Objects

LoC – Library of Congress

LOD – Linked Open Data

LRM – Library Reference Model

LRMoo – Objektum-orientált megközelítésű LRM

MAB - Maschinelles Austauschformat für Bibliotheken

MADS - Metadata Authority Description Schema

MARC – Machine-Readable Cataloging

MESH - Medical Subject Headings

METS – Metadata Encoding and Transmission Standard

MODS – Metadata Object Description Schema

NDMSO - Network Development and MARC Standards Office

OAI-PMH - Open Archives Initiative Protocol for Metadata Harvesting

OWL – Web Ontology Language

PCC - Program for Cooperative Cataloging

PIM-DBK – Petőfi Irodalmi Múzeum Digitális Bölcsészeti Központ

PMO – Performed Music Ontology

PNX – Primo Normalized XML

PREMIS - Preservation Metadata: Implementation Strategies

RDA – Resource Description and Access

RDF – Resource Description Framework

RDFS – RDF Schema

RERO - Réseau des bibliothèques de Suisse occidentale

RIMMF – RDA in Many Metadata Formats

SEO - Search Engine Optimization

SES – String Encoding Scheme

SGML – Standard Generalized Markup Language

SHACL - Shapes Constraint Language

Share-VDE – Share Virtual Discovery Environment

ShEx – Shape Expressions

SKOS – Simple Knowledge Organization System

SPARQL – SPARQL Protocol and RDF Query Language

SRU – Search/Retrieval via URL

TEI – Text Encoding Initiative

URI – Universal Resource Identifier

URL – Universal Resource Locator

VES – Vocabulary Encoding Scheme

VIAF – Virtual Internet Authority File

W3C – World Wide Web Consortium

XML – Extensible Markup Language

XOBIS - XML Organic Bibliographic Information Schema

Felhasznált irodalom

Az elektronikus források hozzáférhetőségének utolsó ellenőrzése: 2022. május 19.

Adams, R. J. (1995). Electronic Libraries SGML Applications: background to project ELSA. *Program*, 29(4), 397-406. <https://doi.org/10.1108/eb047208>

Alemu, G., Stevens, B., Ross, P., Chandler, J. (2012). *Linked Data for Libraries: Benefits of a Conceptual Shift from Library-Specific Record Structures to RDF-based Data Models*. <http://eprints.rclis.org/17523/1/Linked%20Data%20for%20Libraries.pdf>

Allemang, D., Hendler, J. (2011). *Semantic Web for the Working Ontologist: Effective Modeling in RDFS and OWL*. (2nd ed.) Morgan Kaufmann.

Angjeli, A., Baumgartner, M., Chambers, S., Charles, V., Clayphan, R., Deliot, C., Eriksson, J., Freire, N., Huber, A., Jahnke, A., Pedrosa, G., Phillips, V., Pollecutt, N., Robson, G., Seidler, W., Rühle, S. (2012). *D5.1 Report on the alignment of library metadata with the European Data Model (EDM): Version 2.0*

https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/EuropeanaLibraries/Deliverables/D5.1%20%20Alignment%20of%20library%20metadata.pdf

Baker, T., Coyle, K., Petiya, S. (2014). Multi-Entity Models of Resource Description in the Semantic Web: A comparison of FRBR, RDA, and BIBFRAME. *Library Hi Tech*, 32(4), 565-566.

<https://doi.org/10.1108/LHT-08-2014-0081>

Bakonyi G. (2000). *Tartalomszolgáltatás - egy folyóirat digitális feldolgozása*.

<https://nws.niif.hu/ncd2000/docs/eloadas/29/index.htm>

Barry, R. K. (é.n.). *MARC 21 and Mark-Up Languages*.

<https://lekythos.library.ucy.ac.cy/bitstream/handle/10797/11950/anaefar009.pdf?sequence=1>

Beek, W., Raad, J., Acar, E., van Harmelen, F. (2020). MetaLink: A Travel Guide to the LOD Cloud. In *The Semantic Web: 17th International Conference, ESWC 2020, Heraklion, Crete, Greece, May 31–June 4, 2020, Proceedings* (pp. 481-496). Springer. https://doi.org/10.1007/978-3-030-49461-2_28

Berners-Lee, T., Hendler, J., Lassila, O. (2001). The Semantic Web. *Scientific American*, 284(5), 28-37.

Berners-Lee, T. (2006) *Linked Data*. <https://www.w3.org/DesignIssues/LinkedData.html>

Bíró Sz. (2005). *Szövegfeldolgozás XML alapokon*. Neumann János Digitális Könyvtár és Multimédia Központ Kht.

Bizer, C., Heath, T., Berners-Lee, T. (é.n.). *Linked Data – The Story So Far*.

<http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>

Callewaert, R. (2013). FRBRizing your catalogue: the facets of FRBR. In Chambers, S. (ed.) *Catalogue 2.0* (pp. 93-116). Facet. <https://doi.org/10.29085/9781783300259.007>

Campos, F.M., Lopes, M.I., Galvao, R.M. (1995). MARC formats and their use: an overview. *Program* 29(4), 445-459. <https://doi.org/10.1108/eb047213>

- Carlyle, A. (1996). Ordering author and work records: an evaluation of collocation in online catalog displays. *Journal of the Association for Information Science and Technology* 47(7), 538-554. [https://doi.org/10.1002/\(SICI\)1097-4571\(199607\)47:7%3C538::AID-ASI6%3E3.0.CO;2-V](https://doi.org/10.1002/(SICI)1097-4571(199607)47:7%3C538::AID-ASI6%3E3.0.CO;2-V)
- Carlyle, A. (2006). Understanding FRBR As A Conceptual Model: FRBR and the Bibliographic Universe. *Library Resources & Technical Services* 50(4), 264-273. <https://doi.org/10.5860/lrts.50n4.264>
- Carr, P.L. (2015). Serendipity in the Stacks: Libraries, Information Architecture, and the Problems of Accidental Discovery. *College & Research Libraries* 76(6), 831-842. <https://doi.org/10.5860/crl.76.6.831>
- Carroll, J., Stickler, P. (2004). *Trix: RDF triples in XML*. <https://www.hpl.hp.com/techreports/2004/HPL-2004-56.pdf>
- Chandler, A., Foley, D., Hafez, A.M. (2000). Mapping and Converting Essential Federal Geographic Data Committee (FGDC) Metadata into MARC21 and Dublin Core: Towards an Alternative to the FGDC Clearinghouse. *D-Lib Magazine*, 6(1). <https://doi.org/10.1045/january2000-chandler>
- Chen, P. (1976). The entity-relationship model—toward a unified view of data. *ACM Transactions on Database Systems* 1(1), 9-36. <https://doi.org/10.1145/320434.320440>
- Cole, T. W., Kazmer, M. M. (1995). SGML as a component of the digital library. *Library Hi Tech*, 13(4), 75-90. <https://doi.org/10.1108/eb047970>
- Corthouts, J., Philips, R. (1996). SGML: A librarian's perception. *The Electronic Library* 14(2), 101-110. <https://doi.org/10.1108/eb045452>
- Coyle, K. (2007). *MarcXchange*. <http://kcoyle.blogspot.hu/2007/08/marcxchange.html>
- Coyle, K. (2015). FRBR, Twenty Years On. *Cataloging & Classification Quarterly*, 53(3-4), 265-285. <https://doi.org/10.1080/01639374.2014.943446>
- Cundiff, M.V. (2004). An Introduction to the Metadata Encoding and Transmission Standard (METS). *Library Hi Tech* 22(1), 52-64. <https://doi.org/10.1108/07378830410524495>
- Dancs Sz., Mohay A., Hubay M. (2019). Az RDA és a BIBFRAME hazai implementálása. *Tudományos és Műszaki Tájékoztatás*, 66(9), 534-539.
- Dancs Sz. (2017a). Hogyan fogunk katalogizálni a jövőben? FRBR-alapú bibliográfiai feldolgozás a gyakorlatban. *Könyv, könyvtár, könyvtáros*, 26(2), 10-16.
- Dancs Sz. (2017b). RDA-kitekintés: példák az angol-amerikai katalogizálási szabályzat európai átváltéseire. *Könyvtári Figyelő*, 63(1), 85-89.
- Dancs Sz. (2018a). „RDA reborn” - a könyvtári referenciamodell és az átalakuló RDA: 1. rész. Bevezetés az IFLA LRM-be. *Tudományos és Műszaki Tájékoztatás*, 65(3), 141-152.
- Dancs Sz. (2018b). „RDA reborn” - a könyvtári referenciamodell és az átalakuló RDA : 2. rész. A 3R projekt, avagy az RDA megújulása. *Tudományos és Műszaki Tájékoztatás*, 65(4), 228-237.
- Dancs Sz. (2020). Mit jelent RDA-kompatibilisnek lenni? *Tudományos és Műszaki Tájékoztatás*, 67(9), 526-537.
- Davis, S.P. (1996). *SGML-MARC - Incorporating Library Cataloging into the TEI Environment*. <http://xml.coverpages.org/davis-MARC.html>

- De Carvalho, J., Cordeiro, M.I. (2002). *XML and bibliographic data: the TVS (Transport, Validation and Services) model*. <https://archive.ifla.org/IV/ifla68/papers/075-095e.pdf>
- De Carvalho, J.R., Cordeiro, M.I., Lopes, A., Vieira, M. (2004). Meta-information about MARC: an XML framework for validation, explanation and help systems. *Library Hi Tech*, 22(2), 131-137. <https://doi.org/10.1108/07378830410524558>
- Decourselle, J., Duchateau, F., Lumineau, N. (2015). A Survey of FRBRization Techniques. In *TPDL 2015. Lecture Notes in Computer Science, vol 9316* (pp. 185-196). Springer. https://doi.org/10.1007/978-3-319-24592-8_14
- Dean, J.W. (2013). *Charles A. Cutter and Edward Tufte: Coming to a Library Near You, via BIBFRAME*. <http://www.inthelibrarywiththeleadpipe.org/2013/charles-a-cutter-and-edward-tufte-coming-to-a-library-near-you-via-bibframe/>
- Delsey, T. (2002). *Functional Analysis of the MARC21 Bibliographic and Holdings Formats*. <https://www.loc.gov/marc/marc-functional-analysis/source/analysis.pdf>
- Dorman, D. (1995). Technically Speaking: Untangling the Web at the Exhibits. *American Libraries*, 26(8), 806-808. <https://www.jstor.org/stable/25633715>
- Dudás A. (2012). Forrásleírás és hozzáférés: az új angol-amerikai katalogizálási szabályzat (RDA) és kritikája. *Könyvtári Figyelő*, 58(4), 727-750.
- Dudás A. (2013). Nemcsak weben lenni, hanem webből lenni: a Funkcionális követelmények (FR) metaadatmodell-család névterei és a szemantikus web. *Könyvtári Figyelő*, 59(1), 45-64.
- Dudás A. (2017). Az FRBR új, szisztematizált változata: Az IFLA Könyvtári referenciamodell (IFLA-Library Reference Model – LRM). *Könyv, Könyvtár, Könyvtáros* 26(7), 3-14.
- Dunsire, G. (2016). *RDA internationalization and application profiles: applying the global to the local*. <http://www.gordondunsire.com/pubs/pres/RDAInterAP.pptx>
- Eden, B. (2006). FRBR Implementations. *Library Technology Reports* (6), 32-41.
- Espley, J.L., Pillow, R. (2012). The VTLIS Implementation of FRBR. *Cataloging & Classification Quarterly*, 50(5-7), 369-386. <https://doi.org/10.1080/01639374.2012.679882>
- Fernández, J.D., Martínez-Prieto, M.A., Gutiérrez, C., Polleres, A., Arias, M. (2013). Binary RDF representation for publication and exchange (HDT). *Web Semantics* 19, 22-41. <https://doi.org/10.1016/j.websem.2013.01.002>
- Fiander, D.J. (2001). Applying XML to the Bibliographic Description. *Cataloging & Classification Quarterly* 33(2), 17-28. https://doi.org/10.1300/J104v33n02_02
- Freire, N., Borbinha, J., Manguinhas, H. (é.n.) *UNIMARC and XML at the BN*. http://www.bnportugal.pt/agenda/unimarc/a_u_f/borbinha.pdf
- Fülöp E. (2019). *Qulto Semantics – A Qulto Platform szemantikus katalógusa* [video] <https://kifu.videotorium.hu/hu/recordings/32480>
- Gaynor, E. (1996). *From MARC to Markup: SGML and Online Library Systems* <http://xml.coverpages.org/gaynorMARC96.html>
- Giordano, R. (é.n.) *SGML: Giordano on TEI Headers*. <http://xml.coverpages.org/giord.html>

- Gjerde, A. (2008). *Design and use of XML formats for the FRBR model*.
https://web.archive.org/web/20181101031809/https://brage.bibsys.no/xmlui/bitstream/handle/11250/250624/347599_FULLTEXT01.pdf?sequence=3&isAllowed=y
- Goldfarb, C. (1996). *The Roots of SGML – A Personal Recollection*.
<http://www.sgmlsource.com/history/roots.htm>
- Gonzales, B.M. (2014). Linking Libraries to the Web: Linked Data and the Future of the Bibliographic Record. *Information Technology and Libraries*, 33(4), 10-22.
<https://doi.org/10.6017/ital.v33i4.5631>
- Gorman, M. (é.n.) *RDA: The Coming Cataloguing Debacle*. <http://www.slc.bc.ca/rda1007.pdf>
- Guenther, R.S. (2003). MODS: The Metadata Object Description Schema. *Libraries and the Academy*, 3(1), 137-150. <https://doi.org/10.1353/pla.2003.0006>
- Haász A. (2017). A Primo használata a Magyar Tudományos Akadémia Könyvtár és Információs Központban. In Szabó P., Székelyné T.T. (szerk.) *Hagyományok és kihívások V. Múlt és jövő. Országos Könyvtárszakmai Nap, 2016* (pp. 167-178). ELTE.
https://doi.org/10.21862/HagyKihiv_2016.167
- Haller, A., Fernández, J.D., Kamdar, M.R., Polleres, A. (2020). What are Links in Linked Open Data? A Characterization and Evaluation of Links between Knowledge Graphs on the Web. *Journal of Data and Information Quality*, 12(2). <https://doi.org/10.1145/3369875>
- Heath, T., Bizer, C. (2011). *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool.
- Hegyí Á., Sándor Á. (2001). *Folyóirat indexelése Zebrával*.
<http://nws.iif.hu/ncd2001/docs/eloadas/10/index.htm>
- Hickey, T.B., O'Neill, E.T. (2005). FRBRizing OCLC's WorldCat. *Cataloging & Classification Quarterly* 39(3-4), 239-251. https://doi.org/10.1300/J104v39n03_15
- Hickey, T.B., Toves, J. (2009). *FRBR Work-Set Algorithm: version 2.0*
<https://www.oclc.org/content/dam/research/activities/frbralgorithm/2009-08.pdf>
- Hider, P. (2017). A Critique of the FRBR User Tasks and Their Modifications. *Cataloging & Classification Quarterly*, 55(2), 55-74. <https://doi.org/10.1080/01639374.2016.1254698>
- Holden, C. (2020). The Bibliographic Work: History, Theory and Practice. *Cataloging & Classification Quarterly*, 59(2-3), 77-96. <https://doi.org/10.1080/01639374.2020.1850589>
- Hopkinson, A. (1994). *Traditional Communication Formats v SGML, Metadata, Dublin Core*
<https://eprints.mdx.ac.uk/5197/1/LITHPAPE.pdf>
- Horowitz, L.R. (é.n.) *CETH Workshop on Documenting Electronic Texts*
<http://nic.funet.fi/pub/doc/library/teimarc.ps>
- Horváth Á. (2015a). Az ALIADA és az URI névkonvenció. *Tudományos és Műszaki Tájékoztatás*, 62(10), 376-387.
- Horváth Á. (2015b). Gondolatok a szemantikus webről és egyben az ALIADA szoftverről. *Könyvtári Figyelő*, 61(3), 319-326.

- Horváth Á. (2018). *Online katalógusok felhozása a felszíni webre* [video] <https://kifu.videotorium.hu/hu/recordings/24689/online-katalogusok-felhozasa-a-felszini-webre>
- Hubay M. (2020a). Adatgazdagítás, adatszolgáltatás és discovery hagyományos és szemantikus metaadat-környezetben. *Tudományos és Műszaki Tájékoztatás*, 67(5), 291-300.
- Hubay M. (2020b). *Neked is jó, nekem is jó : a Wikidata és a PIM személynév-entitásainak összekapcsolása és adatgazdagítása* [video]. <https://kifu.videotorium.hu/hu/recordings/43341>
- Ilácsa Sz. (2020). Alkalmazásprofilok készítése RDA-hoz: kísérleti projekt alkalmazásprofilok módszertanának kidolgozásához. *Tudományos és Műszaki Tájékoztatás*, 67(9), 538-555.
- The Independent Header* (é.n.). <http://www.tei-c.org/Vault/GL/teip3doc.tar.gz>
- Innovative (2019). *Academic Libraries have the Most Trusted Resources, but their Tools are Hard to Use*. <http://www.iii.com/wp-content/uploads/2019/04/Innovative-Academic-Library-Research-April-2019.pdf>
- Jacsó P. (2002). Digital librarianship: XML and digital librarians. *Computers in Libraries* 22(8), 46-49.
- Jaksič, M. (2004). Mapping of bibliographical standards into XML. *Software-Practice and Experience* 34(11), 1051-1064. <https://doi.org/10.1002/spe.603>
- Jin, Q., Kudeki, D. (2019). Identity and Access Management for Libraries. *Technical Services Quarterly* 36(1), 44-60. <https://doi.org/10.1080/07317131.2018.1532056>
- Johnson, B.C. (2001). XML and MARC: Which is "Right"? *Cataloging & Classification Quarterly* 32(1), 81-90. https://doi.org/10.1300/J104v32n01_07
- Keenan, T.M. (2014). Resource description and access: cataloging standards affect reference service. *Reference Services Review* 42(3), 446-466. <https://doi.org/10.1108/RSR-04-2014-0008>
- Keith, C. (2004). Using XSLT to manipulate MARC metadata. *Library Hi Tech*, 22(2), 122-130. <https://doi.org/10.1108/07378830410524549>
- Kipp, N., Rahman, Z., Bjorklund, M. (1998). *Electronic Thesis and Dissertation Markup Language Version 2: Document Type Definition*. <http://xml.coverpages.org/etddtdBjorklund.htm>
- Kora A. (2003). *Digitalizált szövegek XML-alapú feldolgozása a W3C legújabb ajánlásaival* <https://nws.niif.hu/ncd2003/docs/ehu/EHU-2.htm>
- Kovari J., Folsom S., Younes, R. (2017). Towards a BIBFRAME Implementation: The bibliotek-o Framework. In *Advancing Metadata Practice: Quality, Openness, Interoperability: 2017 Proceedings of the International Conference on Dublin Core and Metadata Applications* (pp. 52-61.) DCMI.
- Kruk, S.R., Woroniecki, T., Gzella, A., Dabrowski, M. (2007). *JeromeDL – a Semantic Digital Library* https://aran.library.nuigalway.ie/bitstream/handle/10379/567/81n2JzC5_1.pdf
- LeBoeuf, P. (2001). FRBR and Further. *Cataloging & Classification Quarterly* 32(4), 15-52. https://doi.org/10.1300/J104v32n04_03
- LeBoeuf, P. (2012a). *An example of library and museum cooperation: the FRBRoo conceptual model*. <https://cidoc.mini.icom.museum/wp-content/uploads/sites/6/2018/12/le-boeuf-keynote.pdf>

LeBoeuf, P. (2012b). A Strange Model Named FRBRoo. *Cataloging & Classification Quarterly* 50(5-7), 422-438. <https://doi.org/10.1080/01639374.2012.679222>

Libris katalogisering av bok [180524] - primärkatalogisera en tryckt monografi (2018) [video]. Youtube. <https://www.youtube.com/watch?v=ZQ5v-OhQLzg>

Lindström, N. (é.n.). *Editing JSON-LD at the National Library of Sweden*. <https://www.kb.se/download/18.d0e4d5b16cd18f600eae9/1569322085836/Editing%20JSON-LD%20at%20the%20National%20Library%20of%20Sweden.pdf>

Lorimer, N. (2019). *Performances and Works: the Interaction of Works and Events in Modeling Sound Recordings in BIBFRAME & PMO*. <https://www.kb.se/download/18.d0e4d5b16cd18f600eafc/1569324461132/PerformancesWorks%20Lorimer.pdf>

Lubetzky, S. (1957). A katalógus funkciója. *Könyvtári Figyelő* 3(2), 62-66.

Madison, O.M.A. (2005). The Origins of the IFLA Study on Functional Requirements for Bibliographic Records. *Cataloging & Classification Quarterly* 39(3-4), 15-37. https://doi.org/10.1300/J104v39n03_02

Malmsten, M. (é.n.). *Exposing Library Data as Linked Data* http://disi.unitn.it/~bernardi/Courses/DL/Slides_10_11/linked_data_libraries.pdf

Marcum, D. (2011). *Transforming our Bibliographic Framework: A Statement from the Library of Congress (May 13, 2011)*. <http://www.loc.gov/bibframe/news/framework-051311.html>

McCallum, S. (1996). *MARC Data in an SGML Structure - the future of communication formats*. <http://xml.coverpages.org/McCallumMARC.html>

McCallum, S. (2004). An introduction to the Metadata Object Description Schema (MODS). *Library Hi Tech*, 22(1), 82-88. <https://doi.org/10.1108/07378830410524521>

McCallum, S. (2018). *BIBFRAME Introduction*. https://www.casalini.it/EBW2018/web_content/2018/presentations/McCallum_1.pdf

McCutcheon, S. (2012). RDA and the Reference Librarian: What to Expect from the New Cataloging Standard. *The Reference Librarian* 53(2), 123-137. <https://doi.org/10.1080/02763877.2011.607409>

McDonough, J. (1998). SGML and the USMARC Standard: Applying Markup to Bibliographic Data. *Technical Services Quarterly* 15(3), 21-33. https://doi.org/10.1300/J124v15n03_02

MetaMorpho: a MorphoLogic tíz éve: 1991-2001 (2001). <https://www.morphologic.hu/downloads/newspapers/morpho-konyv.pdf>

Milhit, I. (2019). *Yet another ILS? Why and how*. <https://slides.com/ignami/rero-elag-19#/2/1>

Miller, D.R., Clarke, K.S. (2003). *Putting XML to Work in the Library: Tools for Improving Access and Management*. ALA Editions.

Miller, D.R. (2000). *XML: Libraries' Strategic Opportunity* <http://cds.cern.ch/record/450652/files/ext-2000-173.pdf>

- Miller, D.R. (2005). XOBIS - An Experimental Schema for Unifying Bibliographic and Authority Records. *Cataloging and Classification Quarterly* 39(3-4), 285-303.
https://doi.org/10.1300/J104v39n03_18
- Miller, E. (2014). *Moving from MARC: How BIBFRAME moves the Linked Data in Libraries conversation to large-scale action*. http://swib.org/swib14/slides/miller_swib14_57.pdf
- Miller, L. (2011). Resource Description and Access (RDA): An Introduction for Reference Librarians. *Reference & User Services Quarterly* 50(3), 216-222.
- Miller, M. (2020a). *Cataloger's Editor*.
<https://www.loc.gov/bibframe/news/pdf/bibframefromhome-mattmiller-bfupdate-june2020.pdf>
- Miller, M. (2020b). *2020-July-09, Library of Congress BIBFRAME Editor experience, Infrastructure track* [video]. Youtube. https://youtu.be/MuS7_KOIPGk
- Mohay A., Hubay M. (2018). *Többnyelvű névtérállományok, szemantikus adatpublikáció és keresőoptimalizálás a Petőfi Irodalmi Múzeumban* [video]
<https://kifu.videotorium.hu/hu/recordings/24572/tobbnyelv-nevterallomanyok-szemantikus-adatpublikacio-es-keresooptimalizalas-a-petofi-irodalmi-muzeumban>
- Moldován I. (1999). *A számítógépen tárolt dokumentumok időtálló formátuma: az SGML/XML*.
https://web.archive.org/web/20070713124911/http://www.bibl.u-szeged.hu/mke_eksz/rendezvenyek/SGML99/
- Morris, S.R., Wiggins, B. (2016). Implementing RDA at the Library of Congress. *JLIS.it*, 7(1), 199-228. <http://dx.doi.org/10.4403/jlis.it-11824>
- Nelson, J. (2019). *Developing Sinopia's Linked-Data Editor with React and Redux*.
<https://journal.code4lib.org/articles/14598>
- Notess, M., Dunn, J.W., Hardesty, J.L. (2011). *Scherzo: A FRBR-Based Music Discovery System*.
<https://dcpapers.dublincore.org/pubs/article/view/3642/1868>
- Noy, N.F., McGuinness, D.L. (é.n.). *Ontology Development 101: A Guide to Creating Your First Ontology*. https://protege.stanford.edu/publications/ontology_development/ontology101-noy-mcguinness.html
- O'Keefe, E., Wacker, M., L'Ecuyer-Coelho, MC. (2019). The Outcome of the ArtFrame Project: A Domain-Specific BIBFRAME Exploration. *Art Documentation*, 38(1).
<https://doi.org/10.1086/703508>
- O'Neill, E.T. (2002). FRBR: Functional Requirements for Bibliographic Records: Application of the Entity-Relationship Model to Humphry Clinker. *Library Resources & Technical Services*, 46(4), 150-159. <https://doi.org/10.5860/lrts.46n4.150>
- Ontotext (é.n.) *What is Five-Star Linked Open Data?*
<https://www.ontotext.com/knowledgehub/fundamentals/five-star-linked-open-data>
- Pajzs J. (é.n.) *A magyar irodalmi és köznyelv nagyszótárának számítógépes megvalósítása*.
<http://www.linguistics.hu/oszt/lexi/pajzs/publ/A%20magyar%20irodalmi%20.pdf>
- Panchyshyn, R.S., Park, A.L. (2015). Resource Description and Access (RDA) Database Enrichment: The Path to a Hybridized Catalog. *Cataloging & Classification Quarterly* 53(2), 214-233.
<https://doi.org/10.1080/01639374.2014.946574>

- Peroni, S., Shotton, D. (2012). FaBiO and CiTO - Ontologies for describing bibliographic resources and citations. *Journal of Web Semantics* 17, 33-43.
<https://doi.org/10.1016/j.websem.2012.08.001>
- Peters, B. (é.n.) *Tutorial*. <http://projects.freelibrary.info/freelib-marc4j/tutorial.html>
- Pettee, J. (1936). The Development of Authorship Entry and the Formulation of Authorship Rules as Found in the Anglo-American Code. *The Library Quarterly* 6(3), 270-290.
<https://doi.org/10.1086/613854>
- Pietriga, E., Bizer, C., Karger, D., Lee, R. (2006). Fresnel: A Browser-Independent Presentation Vocabulary for RDF. In *The Semantic Web : 5th International Semantic Web Conference, ISWC 2006, Athens, GA, USA, November 5-9, 2006. Proceedings* (pp. 158-171.) Springer.
<http://iswc2006.semanticweb.org/items/Bizer2006kl.pdf>
- PIM (2021). „A jövő katalógusa felé” [video]. Youtube.
<https://www.youtube.com/watch?v=hQbB8MtMxIU>
- Pitti, D. V. (1995a). Standard Generalized Markup Language and the Transformation of Cataloging. *The Serials Librarian*, 25(3-4), 243-253. https://doi.org/10.1300/J123v25n03_26
- Pitti, D.V. (1995b). The Berkeley Finding Aid Project: Standards in Navigation.
<http://archive1.village.virginia.edu/dvp4c/arlpap.htm>
- Pitti, D.V. (1997). Encoded Archival Description: The Development of an Encoding Standard for Archival Finding Aids. *American Archivist* 60(3), 268-283.
<https://doi.org/10.17723/aarc.60.3.f5102tt644q123lx>
- Riva, P. (2016). On the new conceptual model of the bibliographic universe: the FRBR Library Reference Model. *AIB Studi*, 56(2), 265-275. <http://dx.doi.org/10.2426/aibstudi-11480>
- Riva, P. (2018). *The IFLA Library Reference Model: lectio magistralis in library science*. Casalini Libri. <https://www.torrossa.com/en/resources/an/4302108?digital=true#>
- Riva, P., Le Boeuf, P., Žumer, M. (2017). *IFLA könyvtári referenciamodell: a bibliográfiai információk elméleti modellje*. https://www.ifla.org/files/assets/cataloguing/frbr-lrm/ifla-lrm-august-2017_rev201712-hu.pdf
- Riva, P., Oliver, C. (2012). Evaluation of RDA as an Implementation of FRBR and FRAD. *Cataloging & Classification Quarterly*, 50(5-7), 564-586. <https://doi.org/10.1080/01639374.2012.680848>
- Rodríguez, E.E. (2016). RDA and ISBD: history of a relationship. *JLIS.it*, 7(2), 49-82.
<http://dx.doi.org/10.4403/jlis.it-11703>
- Rubinsky, Y. (1992). *The SGML Year in Review – 1992*. <http://xml.coverpages.org/sgml1992.html>
- Salgáné M.M. (2004). Az XML: új perspektívák a könyvtár-informatikában. *Tudományos és Műszaki Tájékoztatás*, 51(2), 61-71.
- Salgáné M.M. (2005). A Neumann-galaxis: az elektronikus kor és a számítógépes adatformátumok. *Könyvtári Figyelő*, 51(2), 275-285.

- Sanderson, R. (2015). *Analysis of the BIBFRAME Ontology for Linked Data Best Practices*. <https://goo.gl/KRiuTt>⁵⁶
- Sanderson, R., (2015b). *Linked Data Best Practices and BibFrame*. <https://www.slideshare.net/azaroth42/linked-data-best-practices-and-bibframe>
- Sanderson, R. (2018). *Shout It Out: LOUD*. <https://www.slideshare.net/BaltimoreNISO/sanderson-shout-it-out-loud>
- Sanderson, R. (2020). *The Importance of Being LOUD*. <https://www.slideshare.net/azaroth42/the-importance-of-being-loud>
- Schneider, J. (2008). *FRBRizing MARC records with the FRBR Display Tool*. http://jodischneider.com/pubs/2008may_frbr.html
- Schreur, P.E. (2019). Sinopia: A New Linked-Data Editing Environment Designed for Libraries. In *Metadata and Semantics Research : 13th International Conference, MTSR 2019 Rome, Italy, October 28–31, 2019, Proceedings* (pp. 425-430). Springer. https://doi.org/10.1007/978-3-030-36599-8_39
- Seikel, M., Steele, T. (2020). Comparison of Key Entities Within Bibliographic Models and Implementations : Definitions, Evolution, and Relationships. *Library Resources & Technical Services*, 64(2), 62-71. <https://doi.org/10.5860/lrts.64n2.62>
- Sfakakis, M., Zapounidou, S., Papatheodorou, C. (2020). Mapping Derivative Relationships from BIBFRAME 2.0 to RDA. *Cataloging & Classification Quarterly*, 58(7), 603-631. <https://doi.org/10.1080/01639374.2020.1821856>
- Smiraglia, R.P. (2003). The History of “The Work” in the Modern Catalog. *Cataloging & Classification Quarterly* 35(3-4), 553-567. https://doi.org/10.1300/J104v35n03_13
- Smiraglia, R.P. (2012). Be Careful What You Wish For: FRBR, Some Lacunae, A Review. *Cataloging & Classification Quarterly* 50(5-7), 360-368. <https://doi.org/10.1080/01639374.2012.682254>
- Smiraglia, R.P., Leazer, G.H. (1999). Derivative Bibliographic Relationships: The Work Relationship in a Global Bibliographic Database. *Journal of the American Society for Information Science*, 50(6), 493-504. [https://doi.org/10.1002/\(SICI\)1097-4571\(1999\)50:6<493::AID-ASIA>3.0.CO;2-U](https://doi.org/10.1002/(SICI)1097-4571(1999)50:6<493::AID-ASIA>3.0.CO;2-U)
- Smith, M., Stahmer, C.G., Li, X., Gonzalez, G. (2017). *BIBFLOW: A Roadmap for Library Linked Data Transition*. https://www.library.ucdavis.edu/wp-content/uploads/2020/01/bibflow_roadmap_revised_3_14_2017.pdf
- Solodovnik, I. (2013). Development of a metadata schema describing Institutional Repository content objects enhanced by “LODE-BD” strategies. *JLIS.it*, 4(2), 109-144. <http://dx.doi.org/10.4403/jlis.it-8792>
- Sporny, M. (2012). *The Problem with RDF and Nuclear Power*. <http://web.archive.org/web/20140515020408/http://manu.sporny.org/2012/nuclear-rdf>
- Stanford University Libraries. (2019). *LD4P2 Knowledge Panels Work Cycle Demo* [video]. Youtube. <https://www.youtube.com/watch?v=XjBFmBE3Pck>

⁵⁶ Eredeti forráshelyén már nem elérhető, a bírálati eljárás céljára korlátozottan nyilvános megosztással hozzáférhető a <https://tinyurl.com/5n8tc3be> címen.

- Suominen, O. (2017). *From MARC to schema.org - via BIBFRAME*.
<https://wiki.dnb.de/download/attachments/125433008/Suominen-FromMARCtoSchema.pdf?version=1&modificationDate=1506077926000&api=v2>
- Suominen, O. (2018). *Converging [sic!] BIBFRAME to Schema.org*.
https://www.casalini.it/EBW2018/web_content/2018/presentations/Suominen.pdf
- Szeredi P., Lukácsy G., Benkő T. (2005). *A szemantikus világháló elmélete és gyakorlata*. Typotex.
- Tallerås, K. (2017). Quality of Linked Bibliographic Data: The Models, Vocabularies, and Links of Data Sets Published by Four National Libraries. *Journal of Library Metadata*, 17(2), 126-155.
<https://doi.org/10.1080/19386389.2017.1355166>
- Taniguchi, S. (2002). A conceptual model giving primacy to expression-level bibliographic entity in cataloging. *Journal of Documentation*, 58(4), 363-382.
<https://doi.org/10.1108/00220410210431109>
- Tennant, R. (2001). XML: The Digital Library Hammer. *Library Journal*, 126(5), 30-31.
- Tennant, R. (2002). *MARC Must Die*. <http://lj.libraryjournal.com/2002/10/ljarchives/marc-must-die/>
- Tennant, R. (2010). *MARC Exit Strategies*. <https://www.libraryjournal.com/?detailStory=marc-exit-strategies>
- Tillett, B.B. (2001). Bibliographic Relationships. In Bean, C.A., Green, R. (eds.) *Relationships in the Organization of Knowledge* (pp. 19-35.) Springer. https://doi.org/10.1007/978-94-015-9696-1_2
- Tillett, B.B. (2005). FRBR and Cataloging for the Future. *Cataloging & Classification Quarterly*, 39(3-4), 197-205. https://doi.org/10.1300/J104v39n03_12
- Tillett, B.B. (2016). RDA, or, The Long Journey of the Catalog to the Digital Age. *JLIS.it*, 7(1), 7-24.
<http://dx.doi.org/10.4403/jlis.it-11643>
- Tószegi Zs., Kora A. (2003). Az irodalmi Nobel-díj hatása a Neumann János Digitális Könyvtár életére. *Tudományos és Műszaki Tájékoztatás*, 50(3), 95-99.
- Tóvári J., Szabó B. (2011). *Meta-adat tárolási technikák*. Eszterházy Károly Főiskola.
- Trove (2019). *The perfect Trove record – Part 1*.
<https://trove.nla.gov.au/blog/2019/09/16/perfect-trove-record-part-1>
- Ungváry R. (2005). *Ontológiák és könyvtárak*. <http://nws.iif.hu/ncd2005/docs/ehu/046.pdf>
- Vander Sande, M., Verborgh, R., Hochstenbach, P., Van de Sompel, H. (2018). Toward sustainable publishing and querying of distributed Linked Data archives. *Journal of Documentation*, 74(1), 195-222. <https://doi.org/10.1108/JD-03-2017-0040>
- Vass J. (2017a). Az RDA kapcsolata a nemzetközi könyvtári modellekkel, és elemkészletekkel *Tudományos és Műszaki Tájékoztatás* – I. 64(4), 168-188. ; II. 64(10), 477-495.
- Verona, E. (1959). Literary Unit versus Bibliographical Unit. *Libri* 9(2), 79-104.
<https://doi.org/10.1515/libr.1959.9.1-4.79>

- Wacker, M., Han, M., Dartt, J. (2011). Testing Resource Description and Access (RDA) with Non-MARC Metadata Standards. *Cataloging & Classification Quarterly*, 49(7-8), 655-675.
<https://doi.org/10.1080/01639374.2011.616451>
- Wallis, R. (2018). *Three Linked Data Choices for Libraries: Beyond MARC*.
<https://www.slideshare.net/rjw/three-linked-data-choices-for-libraries>
- Williamschen, J. (2018). *The Library of Congress BIBFRAME Editor*.
http://www.casalini.it/ebw2018/web_content/2018/presentations/Williamschen_1.pdf
- Wilson, P. (1989). Interpreting the Second Objective of the Catalog. *Library Quarterly*, 59(4), 339-353.
- Wusteman, J. (1997). The evolving use of SGML in Electronic Journals. In *Electronic Publishing '97 - New Models and Opportunities: Proceedings of an ICC/IFIP Conference*. University of Kent.
- Yee, M.M. (1994). What is a Work? Part 1: The User and the Objects of the Catalog. *Cataloging & Classification Quarterly*, 19(1), 9-28. https://doi.org/10.1300/J104v19n01_03
- Yee, M.M. (1995a). What is a Work? Part 2: The Anglo-American Cataloging Codes. *Cataloging & Classification Quarterly*, 19(2), 5-22. https://doi.org/10.1300/J104v19n02_02
- Yee, M.M. (1995b). What is a Work? Part 3: The Anglo-American Cataloging Codes. *Cataloging & Classification Quarterly*, 20(1), 25-46. https://doi.org/10.1300/J104v20n01_03
- Yee, M.M. (1995c). What is a Work? Part 4: Cataloging Theorists and a Definition Abstract. *Cataloging & Classification Quarterly*, 20(2), 3-24. http://dx.doi.org/10.1300/J104v20n02_02
- Yee, M. M. (2000). Lubetzky's Work Principle. The Future of Cataloging: Insights from the Lubetzky Symposium, April 18, 1998, University of California, Los Angeles, 1, 72-104.
<https://escholarship.org/uc/item/60d2c41c>
- Yee, M.M. (2005). FRBRization: A Method for Turning Online Public Finding Lists into Online Public Catalogs. *Information Technology and Libraries*, 24(2), 77-95.
<https://doi.org/10.6017/ital.v24i2.3368>
- Zapounidou, S., Sfakakis, M., Papatheodorou, C. (2014a). Integrating library and cultural heritage data models: the BIBFRAME - EDM case. In *PCI '14: Proceedings of the 18th Panhellenic Conference on Informatics*. ACM. <https://doi.org/10.1145/2645791.2645805>
- Zapounidou, S., Sfakakis, M., Papatheodorou, C. (2014b). Library Data Integration: Towards BIBFRAME Mapping to EDM. In *Metadata and Semantics Research: 8th Research Conference, MTSR 2014 Karlsruhe, Germany, November 27–29, 2014, Proceedings* (pp. 262-273.) Springer.
https://doi.org/10.1007/978-3-319-13674-5_25
- Zapounidou, S., Sfakakis, M., Papatheodorou, C. (2019). Mapping Derivative Relationships from RDA to BIBFRAME 2. *Cataloging & Classification Quarterly*, 57(5), 278-308.
<https://doi.org/10.1080/01639374.2019.1650152>
- Zuber, P.A. (2005). *MARC vs XML*.
<http://scholarsarchive.byu.edu/cgi/viewcontent.cgi?article=1997&context=facpub>

Dokumentációk, információs oldalak

About XOBIS (*The XML Organic Bibliographic Information Schema*) (é.n.).

<http://web.archive.org/web/20051214131234/http://www.xobis.info/>

BIBFRAME 2.0 Event Model. (2017). <https://www.loc.gov/bibframe/docs/pdf/bf2-eventmodel-march2017.pdf>

BIBFRAME Authorities. (2014). <https://www.loc.gov/bibframe/docs/bibframe-authorities.html>

BIBFRAME Expectations for ILS tenders. (2018).

https://wiki.dnb.de/download/attachments/125433008/BIBFRAME_Expectations_for_ILS_Tenders.pdf

BIBFRAME Model & Vocabulary. (é.n.)

<http://web.archive.org/web/20151105022115/http://www.loc.gov/bibframe/docs/index.html>

BIBFRAME Vocabulary List View. (é.n.)

<http://web.archive.org/web/20141213174012/http://bibframe.org/vocab-list/#HeldItem>

A Bibliographic Framework for the Digital Age (2011).

<http://www.loc.gov/bibframe/news/framework-103111.html>

„BiblioML” project. (1999).

<https://web.archive.org/web/20020219154829/http://www.culture.fr/BiblioML/fr/index.html>

BP Digital (é.n.). *Schema markup útmutató lépésről lépésre*. <https://bpdigital.hu/schema-markup-utmutato>

Coyle, K., Baker, T. (2009). *Guidelines for Dublin Core™ Application Profiles*.

<https://www.dublincore.org/specifications/dublin-core/profile-guidelines/>

D’Arcus, B., Giasson, F. (2009). *Bibliographic Ontology Specification*. <http://bibliontology.com/>

DB-Engines Ranking of RDF Stores (é.n.). <https://db-engines.com/en/ranking/rdf+store>

DCMI (é.n.). *Metadata expressed in RDF Schema Language*. <https://dublincore.org/schemas/rdfs/>

DCMI Metadata Terms (é.n.). <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>

Doerr, M., Gradmann, S., LeBoeuf, P., Aalberg, T., Bailly, R., Olenky, M. (é.n.). *Final Report on EDM – FRBRoo Application Profile Task Force*.

<http://www.cidoc-crm.org/sites/default/files/TaskfoApplication%2BProfile%2BEDM-FRBRoo.pdf>

DTD-k és metaadatkezelés a MEK-ben (é.n.). <http://mek.oszk.hu/html/irattar/dtd.htm>

ExLibris (é.n.). *The FRBR Vector*.

https://knowledge.exlibrisgroup.com/Primo/Product_Documentation/Primo/Technical_Guide/04_0FRBRization/010The_FRBR_Vector

Futornick, M. (é.n.). *Discovery (WP4)*.

<https://wiki.lyrasis.org/pages/viewpage.action?pageId=101783940>

IFLA. (1998). *A bibliográfiai tételek funkcionális követelményei*.

<http://www.ifla.org/files/assets/cataloguing/frbr/frbr-hu.pdf>

- IFLA (2004). *ISBD(G): General International Standard Bibliographic Description*.
<https://archive.ifla.org/VII/s13/pubs/isbdg2004.pdf>
- IFLA (2011). *ISBD: International Standard Bibliographic Description: Consolidated Edition*.
https://www.ifla.org/wp-content/uploads/2019/05/assets/cataloguing/isbd/isbd-cons_20110321.pdf
- IFLA (2015). *Definition of FRBROO: A Conceptual Model for Bibliographic Information in Object-Oriented Formalism*. https://repository.ifla.org/bitstream/123456789/659/1/frbroo_v_2.4.pdf
- IFLA (2016). *Nyilatkozás a nemzetközi katalogizálási alapelvekről (ICP)*.
https://www.oszk.hu/sites/default/files/ICP_2016_magyarul_2019_09_27.pdf
- KB (2019). *Versionsinformation Libris 1.13.2 (2019-12-11)*.
<https://www.kb.se/download/18.d0e4d5b16cd18f600e3da/1576131347869/Libris-versionsinformation.pdf>
- Library of Congress (é.n.). *Outline of Elements and Attributes in MADS 2.1*.
<http://www.loc.gov/standards/mads/mads-outline-2-1.html>
- Library of Congress (2008). *On the Record: Report of The Library of Congress Working Group on the Future of Bibliographic Control*. <http://www.loc.gov/bibliographic-future/news/lcwg-ontherecord-jan08-final.pdf>
- Library of Congress (2012). *Bibliographic Framework as a Web of Data: Linked Data Model and Supporting Services*. <https://www.loc.gov/bibframe/pdf/marclid-report-11-21-2012.pdf>
- Library of Congress (2014). *RDA in MARC*. <https://www.loc.gov/marc/RDAinMARC.html>
- Linked Data* (é.n.). https://developers.exlibrisgroup.com/alma/integrations/linked_data/
- Logos Research Systems (2001). *MARC Record Resources*.
<http://web.archive.org/web/20010605051444/http://www.logos.com:80/marc/>
- MARC SGML Archive (2003). <https://www.loc.gov/marc/marcsgmlarchive.html>
- MARC Standards (é.n.). <https://www.loc.gov/marc/>
- A Nemzeti Névtér értelméről és hasznáról* (é.n.). <http://abcd.hu/2019/05/13/a-nemzeti-nevter-ertelmerol-es-hasznarol>
- OSZK (2002). *HUNMARC: a bibliográfiai rekordok adatcsere formátuma*.
<https://ki.oszk.hu/sites/default/files/dokumentumtar/hunmarc.pdf>
- OSZK (é.n.). *Szemantikus web: az Országos Széchényi Könyvtár a szemantikus weben*.
http://nektar.oszk.hu/wiki/Szemantikus_web
- PCC (é.n.). *PCC Task Group on URIs in MARC*.
<https://www.loc.gov/aba/pcc/bibframe/TaskGroups/URI-TaskGroup.html>
- PCC (2017). *MARC Proposal No. 2017-08*. <https://www.loc.gov/marc/mac/2017/2017-08.html>
- PCC (2020). *Task Group on Metadata Application Profiles Final Report*.
<https://www.loc.gov/aba/pcc/documents/PCC-TGMAPs-FinalReport-2020-05-01.pdf>

RDA Toolkit (é.n.). *Diachronic works*. https://access.rdatoolkit.org/Guidance/Index?externalId=en-US_ala-443c463e-b075-3e6a-9377-eb230c6b8281

RDA Carrier Type (é.n.). <https://www.rdaregistry.info/termList/RDACarrierType/>

RDA Content Type (é.n.). <https://www.rdaregistry.info/termList/RDAContentType/>

RDA FAQ (2022). <http://www.rda-rsc.org/node/641#1>

RDA Media Type (é.n.). <https://www.rdaregistry.info/termList/RDAMediaType/>

RDF Stores (é.n.). <https://db-engines.com/en/article/RDF+Stores>

Report and Recommendations of the U.S. RDA Test Coordinating Committee (2011). <https://www.loc.gov/aba/rda/source/rdatesting-finalreport-20june2011.pdf>

SGML: TEI to MARC (1996). <http://xml.coverpages.org/tei-marc.html>

TEI (2004). *The TEI Header*. <http://www.tei-c.org/Vault/P4/doc/html/HD.html#HD8>

TEI (2021) *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. <http://www.tei-c.org/release/doc/tei-p5-doc/en/Guidelines.pdf>

Titles in BIBFRAME (é.n.).

<https://wiki.lyrasis.org/download/attachments/79795208/Titles%20in%20BIBFRAME.docx?version=1&modificationDate=1492184599961&api=v2>

W3C (2004). *RDF Primer: W3C Recommendation 10 February 2004*. <https://www.w3.org/TR/rdf-primer/>

W3C (2008). *Cool URIs for the Semantic Web*. <https://www.w3.org/TR/cooluris/>

W3C (2014). *RDF 1.1 N-Quads: a line-based syntax for RDF datasets*. <https://www.w3.org/TR/n-quads/>

W3C (2014b). *RDF 1.1 TriG : RDF Dataset Language*. <https://www.w3.org/TR/trig/>

W3C (2015). *RDFa 1.1 Primer : Third Edition : Rich Structured Data Markup for Web Documents*. <https://www.w3.org/TR/rdfa-primer/>

W3C (2017). *Shapes Constraint Language (SHACL) : W3C Recommendation 20 July 2017*. <https://www.w3.org/TR/2017/REC-shacl-20170720/>

W3C (2019). *Shape Expressions (ShEx) 2.1 Primer*. <http://shex.io/shex-primer/>

W3C (2020). *JSON-LD 1.1 : a JSON-based Serialization for Linked Data*. <https://www.w3.org/TR/json-ld11/>

Wikidata (é.n.) *Wikidata: Property proposal*.

https://www.wikidata.org/wiki/Special:MyLanguage/Wikidata:Property_proposal