



Hálózati metaadat univerzum - könyvtárak válaszátnál

Az OCLC EMEA 2017. évi konferenciájának útkeresései

HORVÁTH Zoltánné

A digitális információ hálózati evolúciójának is nevezett jelenkorban az adatmenedzsment olyan eredményeket generált az utóbbi években, melyeket a régi rendszerek és platformok szintjén már nem lehet intézményi keretbe szorítani. Az adatok az új webtechnológia és a szemantikai irányzat fejlődése nyomán gráfokba rendeződnek a hálózaton, és a leírás alapegységévé a dokumentumok helyett az adatok váltak. A géppel is értelmezhető kapcsolati sémákba rendeződő adatok kiugró méretű ún. big data (adatsokaság) gráfokat alkotnak, és kezelésükhöz különleges adattárolókra, felhő alapú rendszerekre van szükség.

Az adatok hálózati kezeléséhez szükséges speciális adatformátumok könyvtári alkalmazása néhány éve megkezdődött, és az adat- és információkapcsolatok is folyamatosan bővülnek. A teremtett vagy virtuális valóságban tükröződik környezetünk, gondolkodásunk, felkészültségünk, mely a korábitól eltérő, még nem teljesen ismert módon hat a tudományok és a könyvtárak fejlődésére. Az új eredmények erős nemzetközi konzorciumokban, gyakran nyílt közösségek körében születnek az új technológiák által szükségessé váló folyamatokkal együtt. A stratégiai gondolkodás elkerülhetetlen: több évre kell tervezni. A tárolás a működőképesség feltételévé vált a digitalizálás következtében, ezért előttünk áll a könyvtári felhőrendszerekkel kapcsolatos döntések kockázatos lépése is. Még ezen sem jutottunk túl, és már itt az új technológiai korszak, az „Internet of Things” (IoT), vagyis a tárgyak internete, amely az „okos” rendszerek működést vezérlő tranzakciói révén bekapcsolja a tárgyakat is a virtuális adat- és információforgalomba. A problémák áttekintéséhez az OCLC EMEA Libraries at the Crossroads című 2017. évi berlini konferenciájának témáit és előadásait idéztem fel, amelyek közül kiemelkedik a nemzeti és globális infrastruktúrák, platformok együttes fejlesztése.*

A hálózati interakciók dinamizmusa és a könyvtári értékszemplélet

Kutatások és felmérések vizsgálatai szerint a hálózati interakciók elsődlegessége már nem csak az információkeresésben jellemző. Az egyes mikroközösségek saját hálózati útvonalakat követnek, tömeges

csoportosulások kialakulása jósolható egyes események, média hírek, kulturális jelenségek stb. hatására, gyorsan sikeressé vagy perifériakussá válva. Az érték-elvű könyvtári szemlélet elkerülhetetlenül fontos a minőségi szelekció és a hektikus változások kezelése miatt. A gyorsan módosuló, divergens publikációs jelenségek a könyvtárak értékelemző képességének

* EMEA – Europe, the Middle-East and Africa

fejlesztését igénylik a mértékadó arányosság és minőség fenntartása érdekében, amely épp a dinamikus és divergens jelenségek megfelelő kezelésében válik hangsúlyossá a különböző identitású könyvtárak napi gyakorlatában.

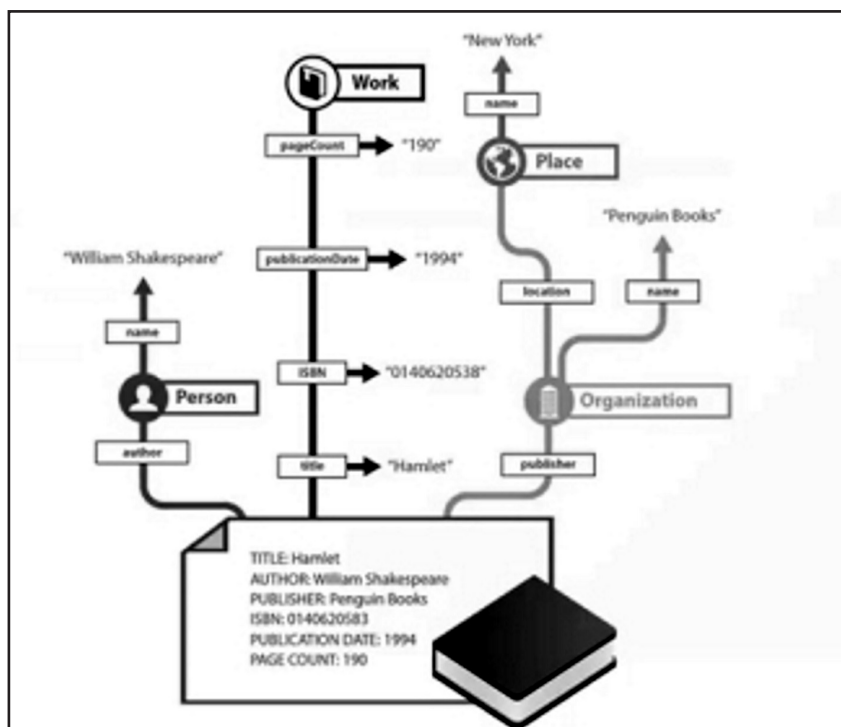
Az *információtechnológia dinamikus hálózati fejlődése* növelte a szolgáltatások komplexitását, a szolgáltatások hatékonyságát és produktivitását, az összetett vagy éppen unikális, speciális szolgáltatások iránti igényt. Az olvasótermek egyre inkább tanuló- és kutatóhelyekké válnak a sok külső, egyébként drága előfizetésű online folyóiratok és adatbázisok, valamint a célorientált, személyre vagy csoportra szabott szolgáltatások igénye miatt. *A hatás és következmény láncolatának* logikája szerint haladva nehéz rangsorolni, melyik tényező hívta életre a transzformációk sorozatának elkerülhetetlenségét a változás és fejlődés folyamatában. Az új média eszközök elterjedése a fiatal generációk kezelésében, az információforrások digitális változatainak gyors növekedése (gyakran megfizethetetlen árakkal) kiegészült a hálózati webtechnológia és a hálózati tartalmak közötti kapcsolati tranzakciók új jelenségeivel. Az információtömeg dinamikus növekedésével egyenes arányban nőtt a rendezettség iránti igény a strukturálatlan információk körében. Jelenlegi ismereteink szerint a szemantikus webtechnológia és adatkezelés jelenti a megoldást, melynek sok stációja volt *Tim Berners-Lee* 2001-es tanulmányát követően. 2006 óta a könyvtári informatikában is megjelentek az elméleti és gyakorlati elemzések és kezdeményezések. Napjainkban milliárdos tömegben láthatók a szemantikus webadat-technológia linked data* fejlődési irányának leírási sémái (FRBR, RDF, SKOS stb.) a hálózaton, és folyamatosan fejlődik a sémához tartozó szabványozási eljárásrend. Könyvtári területen a leírási sémák hálózati elterjedése a nagy könyvtári kutatóközpontok munkájának köszönhető. Szoros együttműködésben végzik az elméleti és gyakorlati fejlesztő tevékenységüket: a Library of Congress, az OCLC, a British Library, a Deutsche Nationalbibliothek, a Bibliothèque nationale de France, valamint további nemzeti könyvtárak közreműködésével. A fejlődés világrendszerek alapjaira épül (ld. az OCLC WorldCat világcatalógusát a 2,3 milliárd rekordjával, a VIAF authority (névtér) adatbázist vagy a DBPédia, az Europeana és más rendszereket).¹ Az elkészült linked data adathalmazok több

leírási sémában is hozzáférhető a nyilvános forrásokban (egy klikkeléssel például a Wikipédia szócikkek alatt, a nemzeti könyvtárak linkjeinél).

Hálózati adatkapcsolat és jelentésbővülés

Az új leírási sémák alapján a művek, alkotások azonosított, egységes adataihoz korlátlanul kapcsolódhatnak új adat- és információelemek, növelve az információs háló (tudásgráf, könyvtári tudásgráf, csomópontok és élek) méretét és összetételét. A növekedés nem egyszerűen mennyiségi, hiszen a kapcsolati háló kiterjedése új értelmezési lehetőségeket, új összefüggéseket is hozhat, amelyek átrendezik, ill. módosítják a háló szerkezetét és jelentését. Az értelmezési lehetőségek gazdagítják a tudományos fejlődést, és a következő lépés tágabb körből indítható. Minden új elem vagy új kapcsolat új szempontokat vehet fel, új kutatást vagy tudományos értekezést generálhat. A digitális hálókbán gyorsan megjelenhetnek az új adatok, a gyorsulás állandó, mértékét legplasztikusabban a nagyon fiatal generáció eszközhasználatára jelzi: játékaikban az előző generáció képtelen lenne hozzájuk hasonló reakcióidőket elérni. A játékok használatában észrevétlenül alakulnak ki új módszerek, amelyek majd megjelennek az új generációk rendszere iránti igényeiben. A hálózati kapcsolatok által egyidejűen feltárolt új összefüggések és új összetevők rendelkezésre állnak ahhoz, hogy a könyvtár falain belül lévő struktúrákat kiegészítsük a korlátlan hálózati forrásokkal, amelyek akkor is léteznek, ha azokat nem vonjuk be könyvtári folyamatainkba vagy adataink kapcsolataiba. A nagy kérdés, hogy a hálózati információforrásokon és új eszközökön nevelkedő használóinkkal hogyan tudjuk saját érték-növelt szolgáltatásainkat megismertetni, és hol vannak azok a pontok, ahol új lépésekre, új szolgáltatásokra, újonnan felvetődött kérdésekre kell megoldást találni? A könyvtár az adatok és információk kezelésében megtartotta háromezer éves gyakorlatát, miközben képessé vált az új technológiák és eljárások alkalmazására. Néhány évenként mégis ugrásszerű változásokkal kell szembenézni. Jelen pillanatban a globális hálózati és felhő alapú információkezelés bevezetése előtt állva, átélhetjük egy hatalmas lépéshez szükséges döntés kockázatát.

* Egyes angol kifejezéseket (pl. linked data, discovery stb.) az eredeti formájukban használom, mert szerintem a fordításuk (pl. kapcsolt adat) nem fejezi ki pontosan az eredeti fogalmat.



1. ábra

Az OCLC felhő-alapú linked data sémája²

A digitális információ hálózati evolúciója – miért globális az információ?

A könyvtári digitalizálás tömegessé válása kölcsönhatásban áll a tudományos eredmények digitalizált megjelenésével, gyorsabb a tájékoztatás, szélesebb a merítési kör, eleve digitálisan jelenik meg a dokumentumok nagy része. Az adatkapcsolatokkal is jellemzett hálózati forrásokban az információ értelmezése a korábbitól eltérő módon változhat. A szemantikus webtechnológia a jelentésalapú adatfeltárás fejlesztésével, az adatok és információk gyors hálózati kapcsolódásával a digitális tartalmak evolúciójaként jelenik meg a tudományokban és ezzel korrelálva a könyvtárakban. Sajnos a tömegesség és gyorsaság nincs átfedésben a minőséggel. Viszont az áttekinthető hálózati lehetősége, a humán tudományok fejlődő gyakorlata, az adatkurátorok gyakran több intézményt átfogó tevékenysége új fejlődési szakaszt teremt a könyvtárak számára is a tendenciák hálózati mintákra alapozott minőségi szempontjaival. A tanult módszerek, a szakterületi együttműködés, az online tudásbázisok és jó gyakorlatok, a nyílt e-learning képzések állandó kínálata nyomán új szerepeket kap a szakma humán összetevőinek együttese, színesítve az

utóbbi időszak technológiai alapú kultúráját (ld. a digital humanities/digitális bölcsészet elemzéseit). Az azonnali nyilvánosság a hálózaton felnagyítja az értelmezési és egyéb hibákat és fokozott figyelmet igényel. A gyors megjelenítés és hozzáférés követelése mellett a minőség megőrzése nagyobb erőket köt le. Kockázatokkal jár, melyik követelményt helyezük előtérbe, ezért a kompetencia és a képesség előfeltétele alapján átrendeződnek a korábbi kompetencia struktúrák és kirajzolódik néhány fontosabb új irány: kommunikációs és kapcsolatteremtési készség, adatelemző és értékelő képesség, technológiai és fejlesztői beállítódás.

A kapcsolati háló fejlesztése attól is függ, hogy a könyvtáros feltáró képessége, műveltsége milyen szintű struktúrák létrehozására alkalmas. Az egyre szélesebb tudástartalom indokoltá teszi az intézeti határok átlépését a külső szakértelem igénybevétele érdekében akár adatbázisokból, akár szakmai virtuális közösségekből merítünk. Még eléggé feltáratlan az egyes szakterületek információs hálózatainak hasonlósága, a struktúrák szabályos vagy szabálytalan alakulása, a növekedés dinamikus vagy eseti jellegének oka és következménye, ezért a tervezhetőség és az elemzés általános gyakorlattá válása még várat magára. Mennyire lehet általános következtetéseket levonni az egyes tudományok hálórendszerének és kapcsolatrendszerének alakulásából?³ Mennyiben tekinthető egy hálóstruktúra mintának, hogyan jelennek meg és hogyan vehetők át az azonos struktúrák, hogyan teremthetünk mintákat? Második, leképezett vagy teremtett világunk („second life”) hogyan tükrözi saját könyvtárunk, műveltségünk, érdeklődésünk jellegét, mélységét vagy felületességét? Ebben a „felülnézetben” milyen törvényszerűségek láthatók, amelyek beépíthetők saját vagy intézményi fejlődésünkbe?

Digitális választónál – az OCLC EMEA 2017. évi konferenciájának útkeresései

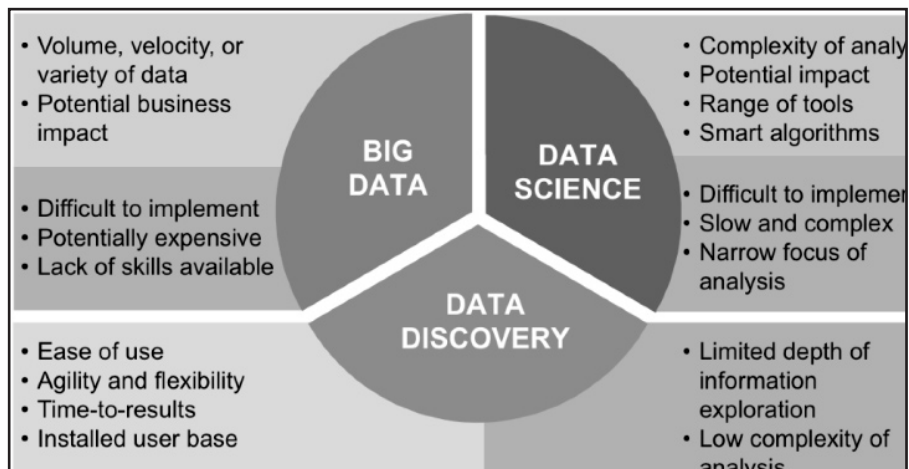
Az OCLC 2017. évi regionális konferenciája a „Könyvtárak választónál” (Libraries at the crossroads) címmel jelöli jelenkorunk döntési kérdéseit, amelyekkel szakmánk szembesül a digitalizálás és a hálózati technológia fejlődésének sodrásában, a hálózati és metaadat univerzumnak aposztrofált⁴ technológiai és fejlődési környezetben, a szemantikus web és az új felhő alapú globális vagy regionális infrastruktúrák világában. A választónál másik lényeges kérdése a könyvtár önmeghatározó képességének (*resolving identities*) vizsgálata volt. A specializálódás, az önmeghatározás és a társadalmi folyamatokba való közvetlen beépülés számomra *Mary Sauer-Games OCLC strategic update* című előadásában jelent meg a legvilágosabban az egyedi használói igények kiszolgálásának szorgalmazásával: „*válgj részévé használóid életének fejlett eszközök, formák, okos technológiák használatával; illeszd be könyvtárad a környezeted működésébe egyedi programokkal, unikális szolgáltatásokkal, tanuló programokkal; integráld és realizáld adataid beépülését a környező gazdaságba, nyújts döntés előkészítő forrásokat és módszereket (pl. statisztika, benchmark stb.)*”.

Digitális hálózati metaadat univerzum (web of data, big data)

Az OCLC konferenciáján a digitális hálózati metaadat univerzumnak⁵ nevezett technológiai környezetben

nemcsak a könyvtár, hanem a tudományok területén is megjelenik a rendezettséghez szükséges adatmenedzsment tevékenység fejlesztése, az adatok kulcsszerepének felismerése. Az adatmenedzsment reneszánszának is nevezett korszakban az adatok kiemelt, gazdasági értéként kezelése a megnevezésben is jelentkezik: adatvagyon, adatgazdagság, adatmenedzsment stb.

A MARC21 és UNIMARC formátumú adatok jelenleg – elsősorban az ún. authority adatbázisoknak köszönhetően biztos háttérrel jelentenek az új szemantikus adatleírás sémák kialakításához, az adatok azonosításához, a hálózati kapcsolatok létrehozásához. De ha megvizsgáljuk, hány állítást lehet – főleg használói szemmel – megfogalmazni a jelenleg rendelkezésre álló MARC rekord adataiból, nyilvánvaló, hogy sokkal kevesebbet, mint amennyi az internetforrások gazdag kínálatában könnyedén elérhető. Az adatok rendszerszemléletű megalkotásához egyelőre a MARC olyan mankót jelent, amelylyel a tartalmi és terjedelmi teljesség mérlegelése, az arányok és mértékek szemmel tartása, az adatok rendszerbe állítása könnyebbnek látszik, mint az adatkapcsolatok gráfokban történő tömeges kezelése és áttekintése. A technológiai fejlődés jelenlegi szintjén két alapvető lehetőség látszik az adatmenedzsment korszerűsítésében. Az egyik a „web of data” (webadat) technológiai fejlődési irány érvényesítése a könyvtári gyakorlatban. Túlhaladva a „document of web” (webdokumentum) korszakot, jelenleg a web of data technológiát bevonjuk könyvtári rendszereinkbe (pl. Wikipédia, VIAF, ISNI linkek beépítése a



2. ábra

Gartner „big data” sémája⁶

rekordokba), gazdagítjuk azokat az új technológia által kínált lehetőségekkel, elfogadjuk, hogy ebben a szakaszban a leírás alapegysége az adat és nem a dokumentum vagy a rekord. Az adatokhoz szabványos azonosítók alapján lehetséges a távoli, akár gépi hozzáférés, értelmezés és kapcsolódás. A hálózati adatkapcsolatok dinamizmusának egyik tünete a big data jelenség, mely kiugró adattömörülések kezelésére szolgáló technológiát követel.

A big data⁷ jelenségnek három lényeges vonása van: mennyiség (volume), sebesség (velocity) és változatosság (variety), melyekre a „3V”-ként is szokás utalni. Az IBM negyediknek a veracityt (érvényesség) is idesorolja, és így már a „4V”-ről beszélhetünk. A mennyiség a másodpercenként létrejövő hatalmas adatözönre vonatkozik, amelyet egyre gyorsabban és gyorsabban kell feldolgozni, és lehetőleg valós időben. Az egyik legnagyobb kihívást a változatosság jelenti, a cél a kontrollálatlan adatfolyamok formázása az értékes információk kinyeréséhez.

MARC adatgazdagítás mellett strukturálhatjuk a rekordokat, ha a linked data gráfok azonosított entitáisait beillesztjük a MARC struktúrába (pl. hely, idő, intézmény, fazettás keresés). A másik lehetőség is összefonódik a linked data fejlődéssel: az ellenőrzött MARC mezőket alkalmazhatjuk a kapcsolt adatstruktúrák csomópontjainak, éleinek megnevezésére. A megnevezésekhez e célra a nagy könyvtárak kulcsszókészlete alapján készített jelölő szótárakat alkalmazunk (pl. a Library of Congress, az OCLC és a francia és német nemzeti könyvtárak közreműködésével fejlesztett *Schema.org*-ot vagy a *Wikidata* gyűjteményt), amelyek nemcsak könyvtári, hanem kereskedelmi szolgáltatók számára is hasznos eszközök, és vannak közös fejlesztések is (pl. Google).⁸

MARC adatmanipulációk és a szemantikus hálók

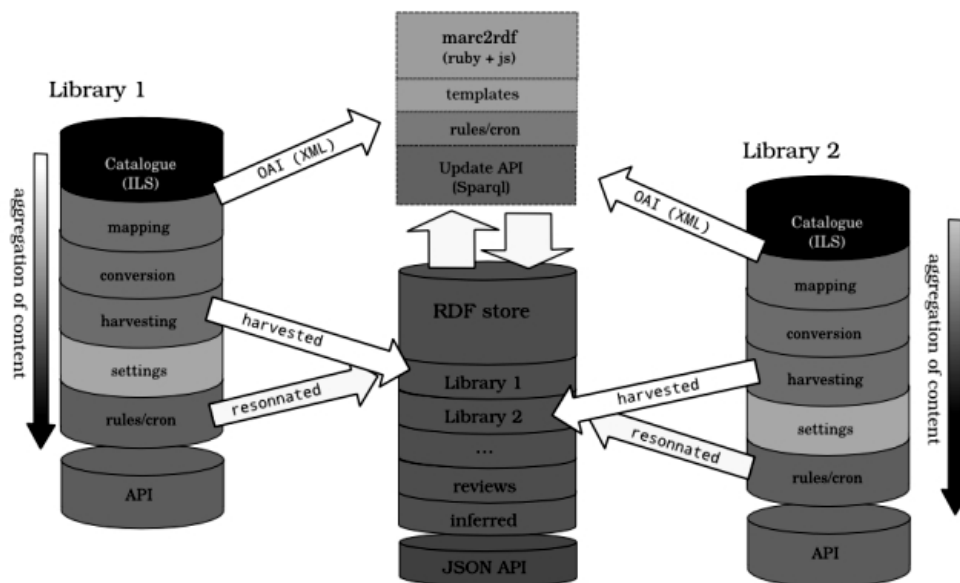
Ha feltesszük azt a kérdést, nélkülözhetők-e jelenleg a MARC formátumú leírások, akkor nyilvánvalóan azonnal azt válaszoljuk, hogy a könyvtárosok számára nélkülözhetetlenek. Továbbá: az új típusú szemantikai fejlődési irány általános alkalmazásához még nincs elegendő tömegesen elérhető, egyértelműen értelmezhető publikált, nyilvános linked data adatkészlet, bár ezek általában publikussá válnak a fejlesztéseket követően. A MARC-tól eltérő egyéb adatleíró sémák napi, rutinos alkalmazására nincs általánosan rendelkezésre álló kompetencia, a tudás transzformálása nagyobb előkészületet igényel. A kutatási projekteknek még be kell épülniük a min-

dennapi gyakorlatba, és szükség van az informatikai képzésben vagy továbbképzésben részt vett könyvtárosokra és informatikusokra. Vannak automatikus konvertáló eszközök, azonban a rendszer felállítása, felkészítése, működtetése speciálisan képzett munkatársakat igényel. Ezért van jelentősége a nagy konzorciumok által végzett kooperációs adatleírási kutatásoknak és fejlesztéseknek, melyek tapasztalatai, gyakorlati megoldásai átvehetők.

A MARC adatformátumokkal számos problémánk van: géppel olvasható, de nem értelmezhető, kevés és nem bővíthető állítás megfogalmazására alkalmas. Hektikusan bővülő adattartalmak leképezését nem oldja meg, ezért a big data jelenséget nem tudja jól kezelni. Viszont a MARC adatok jóval többre használhatók az azonosításnál és keresésnél (pl. linkelés, gépi műveletek, webaratás, gyűjteményanalízis összehasonlítás alapján, találatok rangsorolása stb⁹). A szabványos MARC adatokból jól generálhatók az RDF sémák elsősorban az egységesített authority adatfájlok alapján. Születtek együttműködést segítő kutatási eredmények, de egyelőre senki nem merné állítani, hogy a linked data megoldások teljesen kiválthatnák a MARC típusú adatfeldolgozást, viszont a két eszköz egymást erősíti.

Adatspecialisták, kurátorok, adatenedzsment és adattudomány

A szemantikus webről visszafejthető adatok kiemelik az adatokban gazdag intézmények (könyvtárak, kutatóintézetek, múzeumok, archívumok és levéltárak) adatenedzsment tevékenységének közös vonásait. Az adatspecialisták, köztük a könyvtárosok és adatkurátorok adatenedzsment tevékenysége és felkészültsége konvergens irányban fejlődik. Egy másik irány a hazánkban is létrejött adattudományi képzés, amely 2016-ban indult az ELTE Adattudományi és Adattechnológiai Tanszéken, a Telekom anyavállalata, a Deutsche Telekom berlini intézményének (Telekom Innovation Laboratories, T-Labs)¹⁰ támogatásával. Az adattudomány elsősorban az üzleti intelligencia és az adatokra építő, üzleti alkalmazások informatikai szakembereinek képzését szolgálja az adatbányászat, a mesterséges intelligencia és az adat-vizualizáció felhasználásával. Ezeknek a modelleknek a különböző platformokra és architektúrákra való programozása és megvalósítása lesz az adattudományi szakemberek feladata.¹²



3. ábra

MARC bibliográfiai rekordok RDF konvertálása automatikus eszközökkel¹¹

Az FRBR szemlélet: a szemantikus web és az RDF séma

Az metaadatok új típusú fejlődése az FRBR¹³ leírási séma és a szemantikus webtechnológia elterjedésének köszönhető. Az FRBR típusú adatmenedzsment felfogás szerint nem az egyedi dokumentum, hanem a mű (alkotás) a leírás alapegysége (mint a környezet világ egyedi névvel illelhető entitása), amelynek különböző megjelenési formái vannak. Az entitások leírására alapozott szemantikus webtechnológiában az adatok kapcsolatai jelentéstartalmat hordoznak, és gépi felismerésre, értelmezésre is alkalmasak. Az open linked data fejlődési irány ennek az adatkezelési sémának az alapjaiból építkezik és a szemantikus webadatkezelés szabványos megjelenési formáját képezi a web of data technológiai környezetében. A linked data gráf struktúrái leképezik az entitások kapcsolatát és értéket adnak az entitások tulajdonságainak. A linked data séma formanyelve az RDF (Resource Description Framework, forrásleírás és hozzáférés), amely egyszerű gráf struktúrát alkot, ahol a csomópontok és élek modellezik a leírt adatokat. A gráfban egyszerű *alany – állítmány – tárgy* hármásban (triplets) írjuk le az adatkapcsolatokat¹⁴:

Vegyünk egy azonosítókkal ellátott hármass adatkapcsolatot linked data formában:

a mű: <http://www.worldcat.org/title/winnie-the-pooh-the-collection/oclc/224839609> (Micimackó regény)

; <http://www.worldcat.org/title/winnie-the-pooh-the-collection-includes-winnie-the-pooh-the-house-at-pooh-corner-a-party-for-pooh/oclc/634939175> (Micimackó Sound Recording, stb. 15 > <http://schema.org/author> (has an author relationship to), > szerzője van: A szerző neve: <http://www.viaf.org/viaf/9850718/#> Milne, A.A. (Alan Alexander), 1882-1956; ISNI: 0000 0001 0796 9410 ;

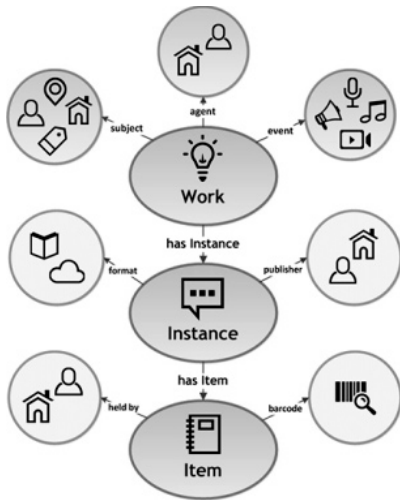
A hangzó dokumentumhoz természetesen több szerző tartozhat, pl. Composer, dalszöveg író stb.

Minden osztálynak és tulajdonságnak van saját azonosítója URI-ja (Uniform Resource Identifier, egységes erőforrás azonosító),¹⁶ amely felismerhető, kapcsolható, géppel kereshető és biztosított az újra használat és az átvétel. Az URI-azonosító olyan rövid karaktersorozatból áll, amelyet a webes erőforrás azonosítására használunk. Közismert példái a webcímek, más néven URL-ek. Az URI az erőforrást hely szerint (URL) vagy név szerint (URN) azonosítja.

A források leírására használt RDF leíró nyelv alkalmazása az egyes forrástulajdonosoknál és intézménytípusoknál igényli az általános adatmodellek előzetes tervezését adathalmaz osztályok meghatározásával.

Az adatmodell készítésének főbb lépései:

- az adatmodell osztályainak meghatározása (pl. személy, mű, hely, szervezet, egyetemenél: kurzus, és fogalmak stb.);
- az adatok azonosításához szükséges URI-tervezés (URL és URN);



Valami dollog (entitás)	Alany	A Mű, amelyről állítást alkotunk (Mici-mackó)
Tulajdonság	Állítmány	a műnek szerzője van (írója, fordítója...) A műnek kiadója van (Publisher) A műnek kiadása van (item)
Tulajdonság felruházása értékkel	Tárgy	A szerző neve (Írta A. A. Milne, Ford. Karinthy Frigyes, illusztrálta A kiadó neve:(Szépirodalmi Kvk.) 3. kiad., stb.

4. ábra

RDF triplet MARC konverzióból (Bibframe)

- a létező ontológiák (adatszótárak stb.) körének meghatározása, amelyeket az adatmodell tervezésben az osztályok és tulajdonságok megnevezésére használni fogunk,
- az adatmodellekben használt névtér prefixek meghatározása, melyek az adatmodell osztályainak URI-jeiből, ill. a használt ontológiák URI-jeiből állnak (pl.: dbr-de <http://de.dbpedia.org/resource>).

Cichon (Biblioteka Narodowa) előadását, melyben ismertetést kaptunk a MARC adatgazdálkodás tárgyában indított projekt eredményéről (ld. az 5. ábra összefoglalását).

OCLC adatmodellek, könyvtári tudásgráf és linked data konverzió

Ahhoz, hogy dolgokra azonos elnevezéseket használjunk, az emberek és gépek számára is azonosítható fogalmakat kell használnunk. Az OCLC a Schema.org egységes jelölőnyelv kialakítását a Library of Congress és az európai nemzeti könyvtárak bevonásával oldotta meg. Jelenleg is szoros együttműkö-

A MARC21 adatgazdálkodás és adatszegmentáció

Az OCLC konferenciáján a MARC21 adatbővítés és adatszegmentáció jelentőségét erősítette, hogy a folyamat összekapcsolódott az RDF entitások gráfjainak megfeleltetésével. Érdemes kiemelni *Marta*

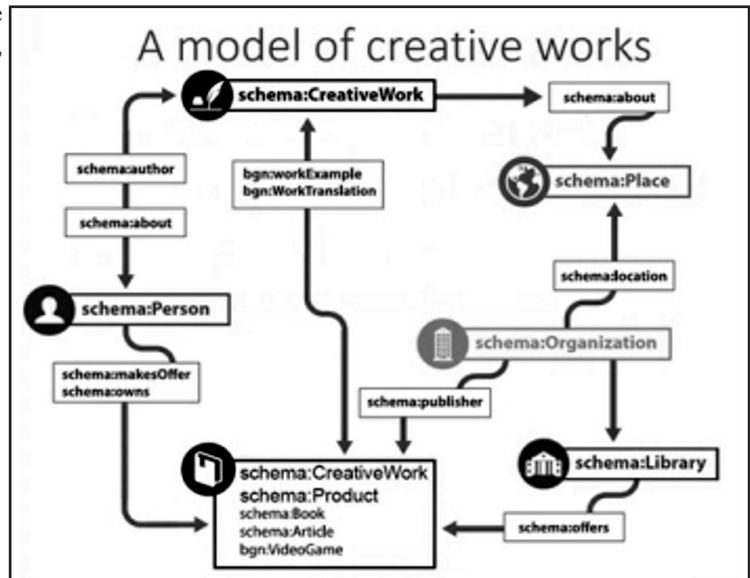
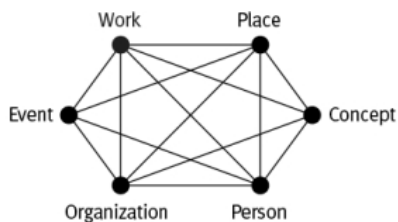
- © Bővítsd új entitásokkal szerzői és tárgyi adataidat (hely, időpont)!
- © Szegmentáld jobban az adataidat!
- © Használj új attribútumokat népszerű források alapján (pl. link import Wikipédiából)!
- © Használj a rekordtartalom gazdálkodására külső speciális adatokat, változatos formákban!
- © Kombináld az automatikus és fél-automatikus rendszereket az adatgazdálkodásnál!

5. ábra

MARC adatgazdálkodás¹⁷

désben dolgozik a W3C munkabizottságaival és más szabályozó testületekkel. A fejlesztésekre a stratégiai szemlélet jellemző, az alapja a WorldCat világcatalógus, melyhez szervesen kapcsolódnak az authority és a szótár rendszerek. A stratégiai szemlélet jellemző az OCLC nemzetközi kooperációjára: szakmai munkabizottságok együttműködésében épül a WorldCat adatbázis és a különböző szemantikai technológiákhoz tartozó egyéb rendszerek készlete (VIAF, ISNI, Wikidata, Schéma.org,

stb.). A Schema.org átfogó megnevezéseket tartalmazó, globális szinten használt szótár a könnyű adaptáció érdekében.¹⁸ A Schema.org folyamatos fejlesztése a nyílt szakmai közösség kezében van, a public-schemaorg@w3.org (<http://lists.w3.org/Archives/Public/public-schemaorg>) levelezési lista és a GitHub (<http://github.com/schemaorg/shemaorg>) használata alapján.¹⁹



6. ábra

OCLC tudásgráf ábrázolása, az entitások általános szintű megnevezése, kapcsolatok²⁰

A szemantikus web az információk rendezésével és szolgáltatásával hivatásszerűen foglalkozók számára két alapvető szempont miatt fontos: egyrészt az adatsémák gráfszerkezete alapján az adatleírások jelentéstartalmakkal bővülnek, másrészt alkalmasak gépi olvasásra és automatikus értelmezésre. Harmadik szempont, hogy az adatkapcsolatokba beépíthetjük a könyvtárhoz vezető automatikus útvonalakat is a weben található adatkapcsolati hálók csomópontjain lévő adataink azonosítói alapján.

- *Az OCLC stratégiai adatkezelése – a WorldCat globális bibliográfiai adatbázis*

A WorldCat a világ legnagyobb katalógusa (kb. 2,3 milliárd tétellel, 72 000 könyvtári kapcsolattal, több mint 50 nemzeti könyvtár rekordjaival), a WorldShare Management könyvtári rendszer és a linked data fejlesztések alaprendszere. A VIAF névtér adatbázis linkekkel kapcsolódik a rendszerhez, ezért a VIAF közvetlen adatbetöltéshez a WorldCat tagságot, és a rekordbetöltést megelőző adattisztítást és eljárási rendet javasolják. A lokális vagy regioná-

lis adathálózatról a betöltés és aratás emberi beavatkozás nélkül történik batch üzemmódban, főként MARC21 adatformátumokról (ld. a holland, dán, svájci, német és ausztrál nemzeti könyvtár és konzorciumaik gyakorlatát). Rengeteg magyar rekord szerepel a rendszerben: 2015-ig a külföldi könyvtárak anyagai alapján, aztán a hazai könyvtárak közül elsőként a Szegedi Tudományegyetem Klebelsberg Könyvtár töltötte fel adatait a rendszerbe. Jelenleg a Magyar Tudományos Akadémia Könyvtár és Levéltár készíti elő a HUNMARC konverzióit MARC21-re a 2017 nyarán induló adatfeltöltéshez. A szervezett felkészülés az OCLC e-learning felkészítő anyagai alapján történik.

- *Kooperatív kutatások nemzeti könyvtárakkal*

Az OCLC adatstratégiájának és működésének legfontosabb tényezője a kooperatív felkészülés, rendszerépítés, fejlesztés és adatmenedzsment. A tagsági viszonyt létesítő könyvtárak az irányító testületekkel együttműködve részt vehetnek a vezetőválasztásban, a fejlesztési stratégia irányainak kijelölésében, külön-

féle közös kutatásokban. Közeli példaként a Lengyel Nemzeti Könyvtár fejlesztéseit említhetnénk.²¹ Az OCLC legfontosabb partnere az új típusú szemantikus adatmenedzsment kialakításában a Library of Congress, gyűjteménye a WorldCat bibliográfiai adatbázis kulcs tényezője.

Az authority adataik fejlesztési kooperációja a Library of Congress és az OCLC együttműködésében a NACO és SACO, vagyis LC Name Authority és az LC Subject Authority fejlesztési programban valósul meg. A résztvevő könyvtárak összeállítják authority rekordjaikat, egységesített és sorozati címeiket a NACO Authority File segítségével. A felkészítésben oktatásokat, tanulmányokat vehetnek igénybe²² a rekordok direkt rendszerbe töltése érdekében. A NACO (Name authority records, <https://www.loc.gov/aba/pcc/naco/index>) és SACO (Subject authority cooperative program, <https://www.loc.gov/aba/pcc/saco/index>) számos eredménye publikálásra került és igénybe vehető.

- *Az OCLC WorldCat integrációk*

Az OCLC WorldCat integrációk linked data kapcsolatokkal és authority adatokkal

- ⊙ VIAF (Virtual International Authority File) – Egységesített Besorolási adatok Nemzetközi Virtuális Adatbázisa; VIAF Identity Network – besorolási adatok összegzése grafikusán is;
- ⊙ ISNI (International Standard [author] Name Identifier) – Nemzetközi Szabványos Névazonosító;
- ⊙ FAST (Faceted Application of Subject Terminology) – webes tárgyszó kezelés az OCLC és a Library of Congress együttműködésében az LC Subject Headings (LCSH) alapján;
- ⊙ Schema.org – egységes jelölőnyelv az OCLC, a Bing, a Google, a Yahoo és a Yandex (orosz keresőmotor) fejlesztésében az interneten szereplő információk strukturálására;
- ⊙ Wikidata – a Wikipédia, WorldCat és VIAF kapcsolat az adatminőség javítására, adatok egységesítésére, azonosítására, nemzeti könyvtárak együttműködésében.

- *VIAF – ISNI – kooperatív fejlesztések*

A VIAF építésében együttműködő nemzeti könyvtárak (British Library, Bibliothèque nationale de France, Deutsche Nationalbibliothek) fejlesztették ki az ISNI rendszert, amely az OCLC CBS (Central Bibliographical System) szerveren fut, és speciális szerzői adatokat biztosít a nemzeti könyvtárak szá-

mára, elsősorban a szerzők nemzetközi egységes számozási rendszerével és más kiegészítő adatokkal. Tagjaik (pl. a finn és norvég nemzeti könyvtárak, az új-zélandi és a lengyel nemzeti könyvtár, jelenleg összesen 25 intézmény) folyamatos fejlesztéseket végeznek. Az eredményt nemcsak a könyvtárak, hanem szövegkiadással foglalkozók, valamint zenei intézmények, tudományos kutatók és különféle szakmai társaságok is használják.

A linked data manipulációk és konverziók könyvtári lépéseinek összefoglalása

Az OCLC nagyon világosan fogalmazza meg a *linked data* alaplépéseit:

- ⊙ a weben minden dolognak és fogalomnak van neve és helye (URI);
- ⊙ az URI fontos információkat hordoz szabványos formátumokban;
- ⊙ ezek az információk explicit rokonságokkal rendelkeznek más dolgokhoz (URI-khoz).²³

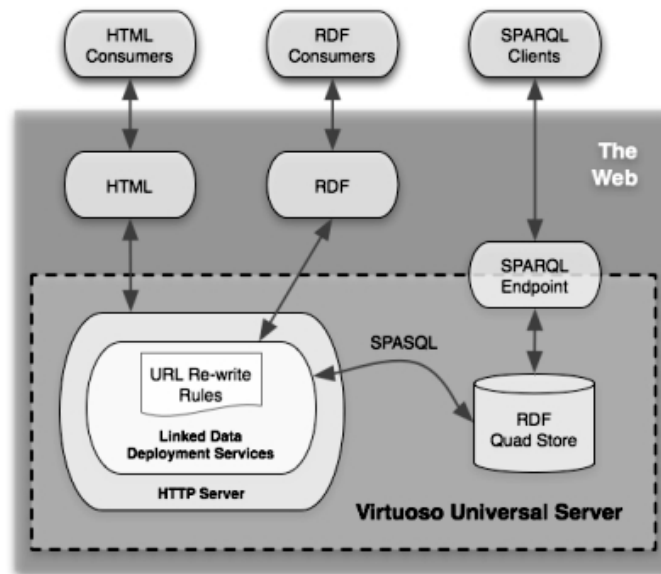
Smith-Yoshimura, az OCLC kutatója a *D-Lib Magazine* 2016. július-augusztusi számában a linked data könyvtári bevezetéséről készült vizsgálat eredményeiről számolt be.²⁴ A kutatás a linked data fejlesztések értékelésével kapcsolatban vizsgálta a következőket: adatok újrahaználata, a discovery funkciók kiterjesztése, új ismeretek előállítása a weben látható aktivitások és a linked data szótárak vizsgálata alapján, a folyamatok és a MARC leírások előnyeinek értékelése, felkészülés a szemantikus webre, a működés eredményessége a mutatók alapján, a szervezet és a munkatársak fejlődése az új folyamatok beépítésével, a MARC alapú leírások RDF modellre konvertálása a linked data alapelvek alapján, nem könyvtáros tartalomspecialisták bevonása az ellenőrzésekbe.

A kutatásban a következő jelentős és gyakran használt *linked data* forrásokat vizsgálták:

- ⊙ Europeana (www.europeana.eu), mely aggregálta a digitális múzeumi, archívumi és audiovizuális archívumok metaadatait Európában;
- ⊙ OCLC WorldCat Linked Data (www.worldcat.org), amely 370 millió bibliográfiai rekordot jelentett meg elérhetően linked data formában;
- ⊙ OCLC's Virtual International Authority File, VIAF (<http://viaf.org>) – különböző országokból és régiókból származó negyven besorolási adatfájl aggregált gyűjteménye;
- ⊙ Getty Vocabularies (<http://vocab.getty.edu>): Art and Architecture Thesaurus (AAT, Művésze-

- ti és Építészeti Tezaurusz), Getty Thesaurus of Geographic Place Names (TGN, a tezauruszhoz tartozó földrajzi helyek névtára), Union List of Artist Names (ULAN, a művészek neveinek nemzeti névtára);
- © Library of Congress' Linked Data Service (<http://ld.loc.gov>) – az LC linked data szolgáltatása több mint ötven szóttal;
- © DBpedia (<http://wiki.dbpedia.org>) – a Wikipédia szócikkeiből összeállított rendszer;
- © GeoNames (<http://www.geonames.org>) – földrajzi névtér rendszer;

- © data.bnf.fr* – a Bibliothèque nationale de France digitális könyvtárakból aggregált adatai;
- © Deutsche Nationalbibliothek Linked Data Services* (http://dnb.de/EN/Service/DigitaleDienste/LinkedData/linkedata_node.html) – a Német Nemzeti Könyvtár linked data szolgáltatásai;
- © FAST Linked Data* (<http://experimental.worldcat.org/fast/>) – Faceted Application of Subject Terminology– az OCLC webes tárgyszó sémája;
- © WorldCat.org* (<http://worldcat.org>) – könyvtári tartalmakat és szolgáltatásokat kínáló központi katalógus hálózat az OCLC működtetésében.



7. ábra

A DBPedia rendszere ²⁵

- *A könyvtárak gyakorlata alapján az alábbi projekt szolgáltatásokat jelölték meg, amelyeknek használata meghaladta a 100 000 kérést naponta:*
- Europeana • OCLC's WorldCat Linked Data • OCLC's Virtual International Authority File (VIAF) • Getty Vocabularies • Library of Congress' Linked Data Service • North Rhine-Westphalia Library Service Center's Linked Open Data szolgáltatásai (bibliográfiai források, könyvtárak és más intézmények elérhetőségével és authority adataival).
- További hat linked data adatállomány napi 10–50 000 kérés kezelésével:
- Bibliothèque nationale de France a könyvtár gyűjteményének és adat-csomópontjának elérésével • British Library's British National Bibliography •

- National Diet Library's [Japán Nemzeti Könyvtár] Authority Data • OCLC's WorldCat Works • OCLC's FAST.

A linked data technológia egységes nemzeti vagy globális alkalmazásaihoz szótárakat és ontológiákat használnak a megnevezések egységes értelmezéséhez, amelyek közül a következők voltak a legáltalánosabban használtak:

- Simple Knowledge Organization System (skos) • Friend of a Friend (foaf) • DCMI Metadata Terms (dcterms) • Dublin Core Metadata Element Set (dce) • Schema.org vocabulary (schema) • The Bibliographic Ontology (bibo) • Local Vocabulary VOCABS rda • Europeana Data Model vocabulary (edm) • ISBD elements (isbd) • WGS84 Geo Positioning (geo) • BIBFRAME Vocabulary (bf) • Expression of Core

FRBR Concepts in RDF (frbr) • The Event Ontology (event) • Metadata Authority Description Schema (mads) • CIDOC Conceptual Reference Model (crm) • BIO: A vocabulary for biographical information (bio) • MODS RDF Ontology • The OAI ORE terms vocabulary (ore) • Core Organization Ontology (org) • DPLA Metadata Application Profile (MAP) • FRBR-aligned Bibliographic Ontology (fabio) • Library extension of schema.org (lib) • Music Ontology (mo) • VIVO Core Ontology (vivo) • Archival collections ontology (arch) • ISNI (International Standard Name Identifier — ontology yet to be published) • RDA Group 2 Elements (rdag2) • British Library Terms RDF schema (blt) • Data Catalog Vocabulary (dcat) • EAC-CPF Descriptions Ontology for Linked Archival Data (eac-cpf) • Nomisma Ontology (The archaeological context of a numismatic object) • Radatana (norvég személynév authority fájlok linked data • formátumban) • Review Vocabulary (rev) (biológiai tárgyú szótár linked data formátumban).

Az általános gyakorlat szerint a szemantikus web-formátumok már létező adatállományok generálásával jönnek létre a különböző forrásokból. A kiinduló adatállomány hagyományos adatformátumokban van, szakaszos (batch módú) betöltés és transzformáció alapján készülnek a szemantikus formátumok. Pl. a *Swissbib*, a svájci nemzeti könyvtár és konzorciuma rendszer esetében is, melyben 21 millió publikáció leírása található. A publikációk központi repozitóriumba kerülnek a résztvevő könyvtárak együttműködésében, az adatokat MARCXML formátumra exportálják. A konverzióhoz egységes szótárakat és ontológiákat is használtak. Az URI azonosítók generálását MARCXML formátumok RDF-re transzformálásával és a kapcsolatok kialakításával oldották meg. Névévadataikhoz a VIAF és a DBPédia authority rendszereket használták az Library of Congress, a British Library és a DNB korábbi tapasztalatai felhasználásával. Végül az RDF reprezentációkat külön repozitóriumban tárolták linkekkel kapcsolódva az eredeti adatokhoz.²⁶

Az OCLC 2015-ig 20 milliárd RDF triplet²⁷ konverziót hajtott végre MARC rekordokból az authority fájlokban. Eddig ez a világ legnagyobb aggregált linked data gyűjteménye. Az authority adatok a WorldCat világcatalógus alapján, a VIAF authority fájlokban szerepelnek a FAST, stb. szótár megnevezéseinek igénybevételével, amelyek a tíz leggyakrabban használt linked data források között szerepelnek a Smith-Yoshimura által vezetett, korábban már ismertetett 2015. évi kutatásban.²⁸

• *Az OCLC entitás alapú adatstratégiájának lépései időrendben*

2010: VIAF, ISNI, FAST;

2012: WorldCat.org linked data kiadása – a Schema.org szótár rendszer megnevezéseivel;

2013: Adatbányászat WorldCat források alapján;

2014: WorldCat újabb linked data bővítés a Schema.org használatával; Schema.org megnevezések a VIAF RDF rendszerében; WorldCat Discovery API (Alkalmazás Programozási Interfész) megfeleltetés a Schema.org RFD formátumához (Béta verzió);

2015–2016: Alkalmazás integráció a WorldCat Discovery, a katalogizálás és az analitikák, valamint az entitások megfeleltetésével kapcsolatban.

A rendelkezésre álló *linked data* rendszerek (pl. VIAF, ISNI stb.) adaptálása még nem kezdődött meg Magyarországon, de több könyvtár tervezi vagy elindította csatlakozását a WorldCat rendszerhez.

Nemzeti könyvtárak linked data fejlesztései

Három nemzeti könyvtár linked data adatállománya kiemelkedik a Smith-Yoshimura kutatás top 12 könyvtári listájából. Mindhárman jelentős tevékenységet mutathatnak fel ebben a témában: a Library of Congress, a Bibliothèque nationale de France, valamint a Deutsche Nationalbibliothek. A British Library²⁹ elsőként tette elérhetővé nemzeti bibliográfiáját *open linked data* formájában is. Az alábbi szabvány sémákat használják: • Bibliographic Ontology – kiadványtípusok leírására szolgáló szótár a kapcsolódó forrásokkal együtt • Bio – életrajzi adatok publikálására szolgáló szótár és kiegészítése • British Library Terms, Dublin Core – alapvető metaadatok • Event Ontology – események és résztvevők megnevezésének leírása • FOAF – emberek kapcsolatainak leírása • ISBD szótár • RDA – nem strukturált tulajdonságok az elemek készletéből • Org – szervezetek linked adatainak leírására szolgáló készlet • OWL • Web Ontology Language • RDF Schema • RDF schema nyelv • SKOS – tárgyszó és taxonómia megnevezések szótára stb.

A DNB (Német Nemzeti Könyvtár) 2010 óta szerkeszti adatait az RDF szabvány alapján, export formátumban is biztosítva az újrahaználás lehetőségét. A szolgáltatások fejlesztését és optimalizálását technológiai és szervezeti szinten is a Library of Congress BIBFRAME – Bibliographic Framework Transition Initiative alapján végzik, együttműködve a Zepheira³⁰ céggel, az OCLC-vel és egyéb partnerekkel. A Német Nemzeti Könyvtár engedélyezi teljes linked data

adatkészletének nyilvános használatát, köztük az összes authority adatot – nemcsak a könyvtárak, hanem a kereskedelmi szolgáltatók számára is.

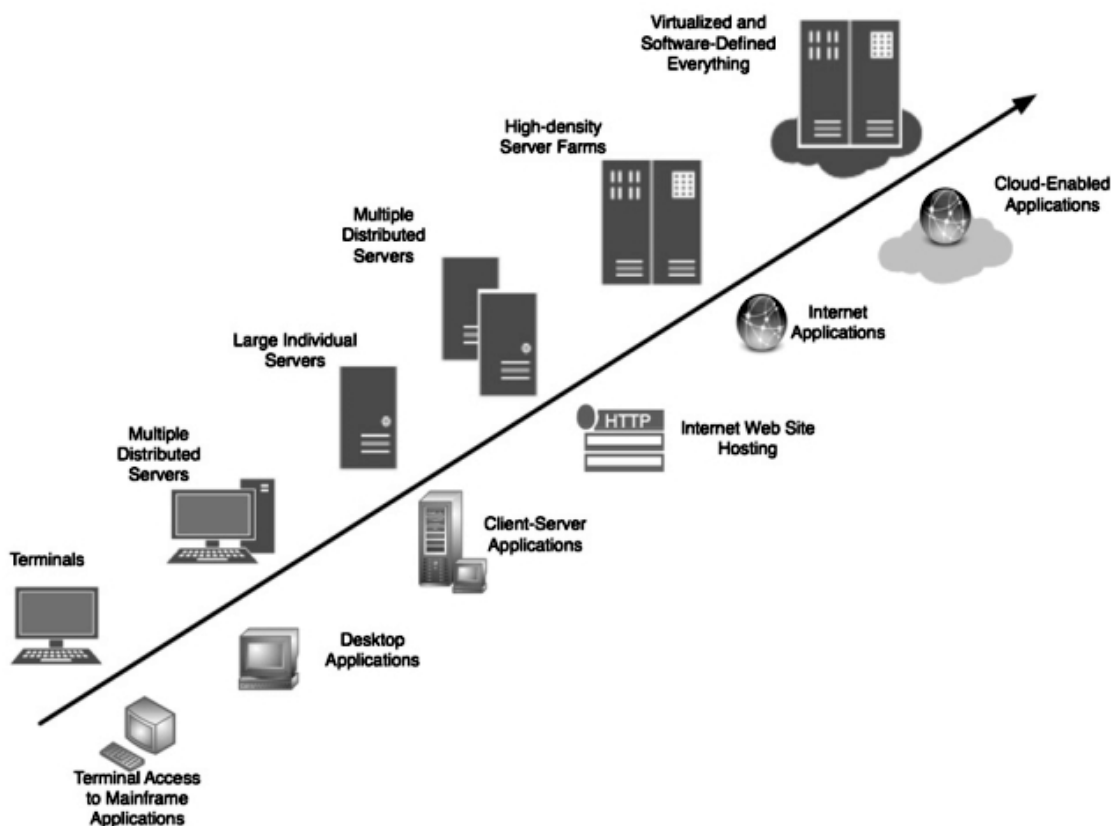
A japán National Diet Library 2015-ben 5 linked data projektről számolt be: bibliográfiai, authority adatok, nemzetközi szabványos azonosítók a könyvtárak és hasonló intézmények számára (ISIL), NIPPON decimális osztályozási rendszer és a nagy Kelet-Japán földrengés adatai. A cikkben van egy táblázat a linked data kérdéssel foglalkozó könyvtárakról, amelyben az OSZK is szerepel.³¹

Globális, regionális, nemzeti szintű platformok

A digitális rendszerek és a hálózati adatmanipulációk kinőtték az intézményi platformok infrastruktúráit, folyamatosan elégtelennek bizonyulnak az intézményi keretek között működtetett szerverek és szoftver együttesek. Az adatmenedzsment megfelelő szintű biztosítása nem maradhat egy intézmény keretén belül, regionális, nemzeti és globális együttműködésre

van szükség az átjárhatóság, a cserélhetőség, és az egységes szabályozás miatt. A kooperáción alapuló adatkezelés olyan technológiai alkalmazás és eljárás együtteseket igényel, amelyek biztosítják a rendszerek közötti kooperációt, az aggregációs folyamatokat, az adat- és információcserét, valamint a webes szolgáltatások használói rendszereit, online tranzakciókat és folyamatkövetést.

Az OCLC EMEA konferencián a globális, regionális és nemzeti szintű összetett infrastruktúra platformok kialakítása és felépítése fő szerepet kapott. Az általában felhő alapú új generációs rendszerek kooperatív, digitális platformjaival kapcsolatban jellemző törekvés a nemzeti adatvagyon és a digitális lokális data hub igény fenntartása a globális felhő rendszerek birtokba vétele mellett. Jellemzőjük a megosztott integrált alkalmazás, a rendszerintegráció, az intézményi igények alapján rendezett tudományrendszertan, a belső és külső szolgáltatások egysége, az automatikus aggregáció, az erős infrastruktúra háttér, adatok, adat-



8. ábra

Evolution of Cloud Computing-alkalmazások fejlődése a felhő alapú rendszerekig³³

bázisok, tudástartalmának szétsugárzása, megújuló használati és kommunikációs alkalmazások.³²

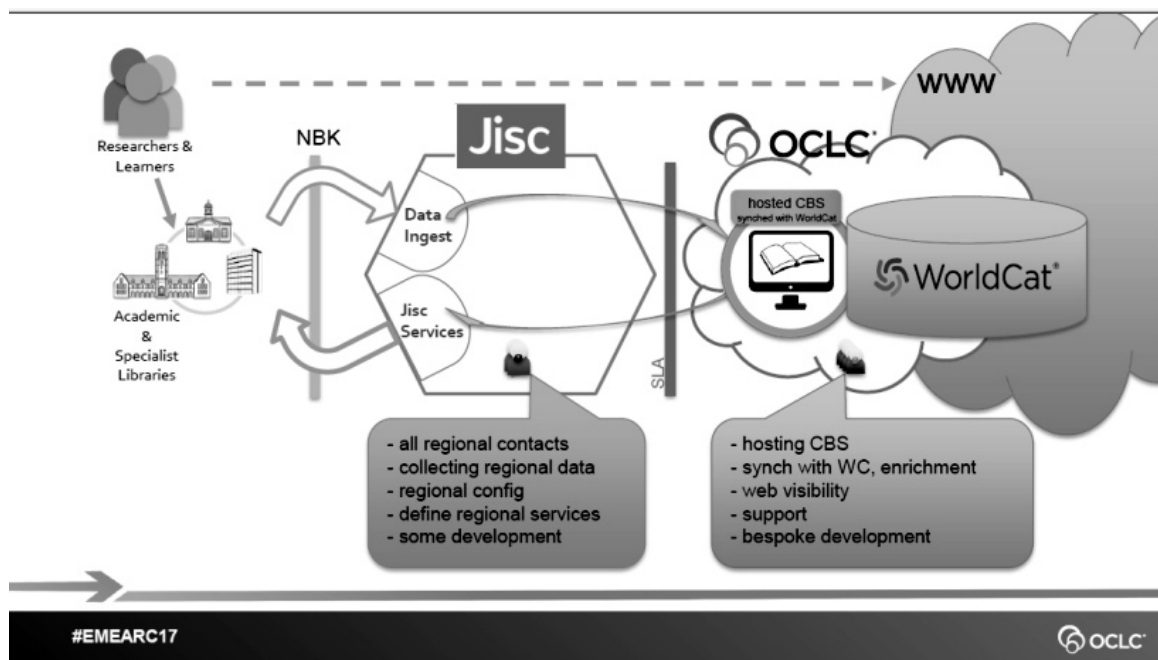
Az OCLC új generációs felhő alapú rendszereihez biztosított digitális platformok globális, osztott infrastruktúrát nyújtanak a könyvtári szolgáltatásokhoz. Két adatközpontjuk van az USA-ban, egy Nagy-Britanniában, 2012-től Európában, Ausztráliában és Kanadában is erős tárolókapacitást hoztak létre. A WorldShare Management platform képes a tagkönyvtárak és használók teljes ellátására egy globális osztott infrastruktúrán. Olyan fejlesztési és alkalmazási készletet nyújt, amely kiterjesztett technológiai háttérrel jelent globális szinten biztosított könyvtári web alapú szolgáltatási rendszerhez. A platformot a szolgáltató gondozza, frissíti, és megfelelő, biztonságos tárhelykapacitást biztosít. (lásd 8. ábra)

Az OCLC a nemzeti, regionális és globális infrastruktúra rendszerek együttműködésének érdekében új megoldásokat alkalmaz a metaadat univerzumként aposztrofált új generációs felhő alapú rendszerek és a könyvtári kapcsolt adatok fejlesztései során. Legtöbbször a nemzeti könyvtár és több egyetemi könyvtár konzorciumaiban valósulnak meg a fejlesztések, amelyek tárgya az egységes könyvtári platformok, a discovery szolgáltatások, a metaadatok minőségi fejlesztése, az adatgazdagítás, adatsere és

aggregációs folyamatok hatékonyabbá tétele, a privát (regionális, nemzeti) és a globális felhő alapú rendszerek összehangolása. A fejlesztések a WorldCat világcatalógus, az authority rendszerek (VIAF, ISNI) alapján, a WorldShare Management rendszer egységes platformjainak adaptációjával, nyílt API-kkal (application programming interface, alkalmazás programozási interfész) történnek.

A privát vagy nemzeti data hub alkalmazásokat a CBS rendszerrel oldják meg, amely elvégzi az adatok aggregálását, az adattisztítást, a különböző MARC formátumok konvertálását, és a linked data formátumok konverzióit.

- *Az egységes digitális platformok általános jellemzői:*
- ⊙ osztott és felhő alapú infrastruktúra a többféle alkalmazáshoz;
- ⊙ big data kezelésére nagy teljesítményű szerver clusterek (HBASE szerverek);
- ⊙ legfrissebb technológia, magas szintű webalkalmazások;
- ⊙ nemzeti és globális adatközpontokban extenzív és szolgáltatás központú architektúrák, regionális adatközpontok data hub alkalmazásokkal;
- ⊙ szinkronizáció, interoperabilitás, integráció az együttműködő rendszerekhez;



9. ábra

NBK – National Bibliographical Knowledge base (holland) WorldCat és CBS integrációval³⁴

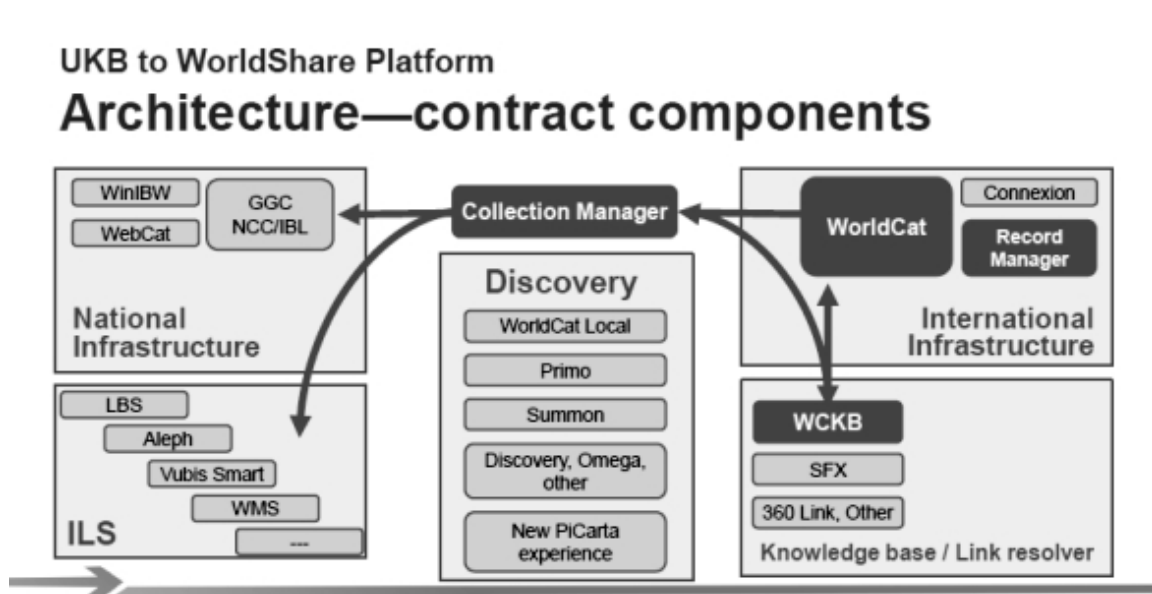
- © új megközelítésű könyvtári munkafolyamat (workflow), amely illeszkedik az új webes technológiákhoz és az új felhasználói szokásokhoz (mobil keresés, szociális média, tranzakciós folyamatok stb.);
- © WorldCat Discovery – felhasználó központú keresési környezet, amely a keresést és kutatást megtervezett útvonalakkal támogatja – szállítással és szükség esetén számlázással vagy külső szolgáltatóhoz vezető linkekkel. Nyílt kutatási és együttműködési felületet biztosítanak a könyvtárak, a könyvtári IT szolgáltatók és harmadik fél (pl. kereskedelmi szolgáltatók) számára megfelelő részvételi feltételek mellett.³⁵
- *Discovery platformok – keresés, kutatás előre épített útvonalakon*

A kutatások tárgyát képező összefüggések korábban is léteztek, de manuális módszerekkel teljes felfedezésük csak kutatómunkával volt megoldható. A discovery platformoknál a kutatás és keresés a rendszerekbe előzetesen beépített „útvonalai” miatt befolyásolható a keresési munkafolyamat. A discovery funkció az egyre növekvő tudáshalmazokban leképezi a megismerés folyamatát a lehetséges kapcsolatok kiépítésének nyomán, de egyúttal a tudományos fejlődés ösztönzőjeként is működik: a hálózati környezet az asszociációs gondolkodást és a tudomá-

nyos kutatói és megismerési módszerek követését és fejlődését is támogatja a könyvtár típusához illő adatbázisok és adatforrások szervezett gyűjteményei, az elérés útvonalai és a felelős könyvtáros rendszerismerete alapján.

- *Interoperabilitás, rendszerintegráció, middleware*

Fontos tényező a rendszerek közötti interoperabilitást biztosító *middleware*³⁷ alkalmazások beillesztése. Általában nyílt API-kat alkalmaznak annak érdekében, hogy a fejlesztők és külső közreműködők megfelelő jogosultság háttérrel részt vehessenek a rendszerek tranzakcióiban, adat- és információ cserében, testre szabásban, fejlesztésben különböző könyvtári katalógusok, webes és egyéb felületek, platformok, digitális dokumentumok és leíró adataik, külső források beillesztésében, leíró szabványok, MARC formátumok, linked data formátumok, ellenőrzött szótárak, stb. egységesítési, fejlesztési és testre szabási folyamataiban. Az összetett platformokon megtervezett és kialakított discovery funkció a különböző belső és külső rendszerek összehangolt együttműködését igényli. A felhő alapú rendszereknél a könyvtárnak nincs gondja a szerver beszerzéssel, fenntartással, a szoftverek és rendszerek frissítésével, a rendszerek együttműködéséhez szükséges tervezőmunka és a rendszerintegráció viszont a könyvtár és a fejlesztő



10. ábra

Holland nemzeti információs infrastruktúra komponensek (WorldCat, ALMA, Summon, Primo, Aleph, WorldShare Management (UKB = Consortium of Dutch University libraries + Royal Library)³⁶

cég részéről szoros együttes munkát igényel követelményjegyzék és folyamatleírás alapján.

Az OCLC WorldCat globális discovery szolgáltatás a forrásmegosztást az alábbiak szerint nyújtja:

300 millió MARC alapú katalógus rekord, 2.1 milliárd állományi adat, 40 millió digitális gyűjtemény tétel, 1,5 milliárd cikk a központi indexből.

A WorldCat Discovery API a linked data technológia alapján RDF formátumban is lehetővé teszi az adatok hozzáférését. A megnevezésekhez elsődleges linked data szótárát, a Schema.org rendszert használja. A központi index cikkeket, e-könyveket, és egyéb digitális anyagokat tartalmaz az alábbi digitális szöveges és képállományokból: ArchiveGrid, ArticleFirst, Camio, Catalog of Art museum images, Worldcat Dissertations, ERIC, GPO Monthly Catalog, MEDLINE, stb. További információk: Bibliographic Resource documentation³⁸

A legfontosabb index-alapú könyvtári discovery szolgáltatások: Primo and Primo Central from Ex Libris Group (2017-től Proquest), EBSCO Discovery Service az EBSCO Information Services alapján; Summon® ProQuest; WorldCat® Discovery Service OCLC.

A discovery felületeket szokás next-generation katalógusként is jelölni, amely a korábbi OPAC funkciókat teljesen új alapokon, bővített szolgáltatásokkal nyújtja: pl. relevancia szerinti rangsorolás, fazettás navigáció, web-alapú források bevonása.

A discovery rendszer és szolgáltatás részei:

© Végfelhasználó felület, amely egy webböngésző keresőablak szolgáltatáson felül alternatív kereső oldalakat is biztosít a „haladó” keresések opciójával a strukturált mezők tartalmában, és változatos listázási szolgáltatásokat is kínál. Együttműködik a linkszolgáltatásokkal és teljes szövegű dokumentumokhoz ad linkeket a keresésnél, gyakran az Apache SOLR™ (lucene.apache.org/solr/) vagy az ElasticSearch® (www.elasticsearch.org/) lokálisan alkalmazott kereső eszközökkel. Képes interaktív kommunikációra a könyvtári IKR-rel az elérhető tartalmak státusával kapcsolatban, és listázza a tételek elérésének forrásait, árát, vagy megrendelési lehetőségeit. A platformhoz biztosított API által távoli indexek elérhetőségét nyújtja az eredmények listázásával.³⁹

© A discovery szolgáltatásokkal szemben kihívást jelent, ha a források elérése korlátozott, csak azo-

nosított felhasználók számára lehetséges. Több adatbázis-szolgáltató esetében komoly informatikai problémát jelent a különböző szintű hozzáférések megoldása a discovery platformon. Az egyes szolgáltatók absztrakt és index szolgáltatása eltérő lehet akár a lefedett szakterület, akár más tényezők miatt. A diszciplína által meghatározott szótárak, kivonatok kezelése az egységes discovery szolgáltatástól **speciális megoldást** kíván. Ha nincs kooperáció a különböző szolgáltatók között, gondok jelentkezhetnek az általános vagy a szakmai szótárak, az eltérő metaadat formátumok összehangolása, átvétele tárgyában⁴⁰. A szolgáltatók kiválasztásánál érdemes megfontolni, hogy mennyire akar vagy mennyire képes könyvtári szempontok tárgyában kooperációra? Fontos szempont a találati lista rangsorolásának módszere és egyeztetési lehetősége.

© A discovery felületeken a szolgáltatók támogatják a könyvtár és a használók értékelemző és szelektív tevékenységét a legjobb források kiválasztásával, az ún. magfolyóiratokkal. Épp ezért fontos összetevő az oktató vagy e-learning rendszerek szerves összekapcsolása a discovery rendszerrel. Ha van lehetőség kész discovery szolgáltatás igénybevételére, érdemes mérlegelni az önálló kialakítás kockázatait.

Internet of Things (IoT) – intelligens technológiák és a megszelídített adattenger

A könyvtárakban jelenleg fejlődő szemantikus és digitális adatáramlás a teljes tudományos, gazdasági és irányítási környezetben „smart” (intelligens) technológiaként terjed el, és az Internet of Things fejlődési irányban a tárgyakra is vonatkozni fog, amelyek beépített szenzorokkal kommunikálnak majd a szerverekkel. Az internethez kapcsolódva adatokat küldenek és fogadnak, távolról vezérelhetővé válnak. A hálózatba kapcsolt eszközök erős IT háttérrel vesznek részt a virtuális adatforgalomban, mint közlekedési, turisztikai, kulturális és gazdasági információk, amelyeket már keresni sem kell, mert életünk és eszközeink részévé válnak.

Könyvtáraknál elsősorban a „smart” technológiák hatására lehet számítani (pl. intelligens város könyvtári funkciói az állampolgári tájékoztatásban, hozzáférési és azonosítási kártyarendszerek, mobil fizetés, esemény regisztrációk, „okos” könyvek, „*kiterjesztett valóság*” a tájékoztatásban, objektum-alapú, táv- és mobil oktatás, „self” e-learning programok, kompa-

tibilis eszközök és technológiák szélesítése az általános használathoz, a linked data technológia kiterjedt alkalmazása a tárgyakra vonatkozó adatok körében, interaktív kijelzők, szabadtéri wifi hálózat felhasználása tájékoztatásra, marketingre stb.

A könyvtárak jelenleg is élen járnak az IoT megvalósításának kezdeményezéseiben. A könyvek mint tárgyak digitális adataik révén jelenleg is elérhetők a hálózatról, sőt mobil eszközzel „mozgathatók” is (kölcsonzés, hosszabbítás). Az IoT legfontosabb célja a fenntartható fejlődés elősegítése, a használók informálása, tájékoztatása intelligens eszközök útján, a monoton és manuális cselekvések kiváltása. A cél tehát számunkra ismerős, a megvalósulás formáit pedig szokás szerint az életformánkká vált folyamatos tanulással kell elsajátítanunk.

A táguló horizontot nemcsak a fejlődő technológia adja, hanem az emberi kreativitás, amely egységbe foglalja a lehetőségeket és az alkotókészséget.

Jegyzetek

- OCLC linked data adatközpont: <https://www.oclc.org/developer/develop/linked-data.en.html>;
- Godby, Jean Carol et al. <https://www.youtube.com/watch?v=mj0CVMzM7dQ&feature=youtu.be>
- BARABÁSI Albert László: <http://mindentudas.videotorium.hu/hu/recordings/details/8316>,
- A konferencia több előadásában szereplő kifejezés: network metadata universe – abban az értelemben, hogy a hálózat és metaadat világrendszerek vannak születőben, a könyvtárak számára a hálózat döntő szerepet tölt be, és ezzel összefüggésben a felhő technológia alkalmazások aktualitása elemzést igényel.
- Id. még: GRUPI, Gianferanco: Beyond the pillars of Hercules. Linked data and cultural heritage = J LIS . vol. 4. 2013. no. 1.
- Közli: TIMO, Elliott: <http://timoelliott.com/blog/2015/03/what-is-big-data-discovery.html>
- Sven Löfflerttel, a T-Systems Üzleti Intelligencia és Big Data üzletfejlesztési igazgatójával készült interjúból = <https://www.it-services.hu/hirek/mi-az-a-big-data/>; <https://www-01.ibm.com/software/data/bigdata/>
- SMITH-YOSHIMURA et al. <http://www.oclc.org/content/dam/research/publications/library/2010/2010-06.pdf>;
- SMITH-YOSHIMURA: <http://www.dlib.org/dlib/july16/smith-yoshimura/07smith-yoshimura.html>;
- <https://www.elte.hu/content/uj-tanszek-indul-az-elte-n-a-deutsche-telekom-tamogatasaval.t.12297>
- ROKSETH, Benjamin: <https://github.com/digibib/marc2rdf>;
- <https://www.elte.hu/content/uj-tanszek-indul-az-elte-n-a-deutsche-telekom-tamogatasaval.t.12297>
- FRBR: Functional Requirements for Bibliographic Records: <https://www.ifla.org/publications/functional-requirements-for-bibliographic-records>
- McCALLUM, Sally : <http://www.loc.gov/bibframe/pdf/BFI-IGeLU-ppt2003.pdf>
- Bövebben ld. Linked data explorer: <http://experiment.worldcat.org/entity/work/data/2406166>
- Uniform Resource Identifier – (pl. 'issn' defined by [RFC 3043], Using The ISSN as URN within an ISSN-URN Namespace (W3C <http://www.w3.org/TR/uri-clarification/#RFC3043>)
- <https://www.oclc.org/content/dam/oclc/events/2017/EMEARC2017/EMEARC-2017-Session-J-Resolving-the-Identities-of-Entities-Marta-Chichon.pdf>
- Friend-of-a-Friend (FOAF); Simple Knowledge Organization System (SKOS); Creative Commons (CC), Schema.org
- <http://f.estontorise.hu/archives/288>
- Godby, Jean Carol-Wang, Shenghui-Mixer, Jeffrey K. <https://www.youtube.com/watch?v=mj0CVMzM7dQ&feature=youtu.be>
- Id. Marta Chichon már idézett előadását.
- Name Authority Cooperative Program <https://www.loc.gov/aba/pcc/naco/index.html>, Guidelines for reporting NACO BFM , SACO; <https://www.loc.gov/aba/pcc/naco/bfmguid.html>, SACO - Subject Authority Cooperative Program; <https://www.loc.gov/aba/pcc/conser/index.html> ;
- <https://www.youtube.com/watch?v=fWfEYcnk8Z8>; (Linked data for libraries) API Application Programming Interface
- A nemzetközi linked data implementációkról Id. az OCLC 2014-2015. évi kutatását 20 ország 90 intézményének válaszai alapján. Az elemzés 112 linked data projektet összesít. Smith-Yoshimura: <http://www.dlib.org/dlib/july16/smith-yoshimura/07smith-yoshimura.html>
- <http://wiki.dbpedia.org/about/architecture>
- <http://journal.frontiersin.org/article/10.3389/fdigh.2017.00005/full>
- Triplet: RFD hármas struktúra: alany, állítmány, tárgy, amelyből az alany az entitás megnevezése (pl. mű), az állítmány, amelyet az alanyról állítok, a tárgy az állítás attribútummal

való felruházása

28. <http://www.dlib.org/dlib/july16/smith-yoshimura/07smith-yoshimura.html>
29. http://www.bl.uk/bibliographic/pdfs/publishing_bnb_as_lod.pdf; <http://bnb.data.bl.uk/>; <http://www.bl.uk/bibliographic/natbib.html>; http://www.dnb.de/EN/Service/DigitaleDienste/LinkedData/linkedata_node.html
30. Zepheira: <https://zepheira.com/>
31. Smith-Yoshimura: <http://www.dlib.org/dlib/july16/smith-yoshimura/07smith-yoshimura.html> Institutions Responding to International Linked Data Survey for Implementers táblázat
32. Id. erről Kortekaas, Simone: UKB projekt. A kooperatív platformok
33. <https://www.safaribooksonline.com/library/view/the-enterprise-cloud/9781491907832/ch01.html>
34. KASCHTE, Axel: <http://www.oclc.org/content/dam/oclc/events/2017/EMEARC2017/EMEARC-2017-Session-A-Regional-and-National-Infrastructure>
35. <https://www.oclc.org/developer/develop/web-services/worldcat-discovery-api.en.html>
36. <http://www.oclc.org/content/dam/oclc/events/2017/EMEARC2017/EMEARC-2017-Session-A-Regional-and-National-Infrastructure-David-Whitehair-Simone-Kortekaas-and-Axel-Kaschte.pdf> /Simone Kortekaas prezentációja. További információkat ld. <https://www.ukb.nl/english/worldshare>
37. Middleware: megosztott rendszerek közötti kommunikációt és adatkezelést lehetővé tevő alkalmazás
38. <https://www.oclc.org/developer/develop/web-services/worldcat-discovery-api/bibliographic-resource.en.html>; <https://github.com/OCLC-Developer-Network/worldcat-discovery-php>; https://www.oclc.org/content/dam/oclc/services/brochures/215159-WWBE_WCDDiscovery-Content.pdf
39. NISO: The Future of Library Resource Discovery
http://www.niso.org/apps/group_public/download.php/14487/future_library_resource_discovery.pdf
40. BREEDING, Marshall: <https://americanlibrariesmagazine.org/2014/01/14/web-scale-discovery-services/>

Beérkezett: 2017. április 26.



Az MKE sikeres nemzetközi pályázata: IFLA International Advocacy Programme

Regionális érdekérvényesítési szeminárium megrendezésére nyert 12 ezer eurót a Magyar Könyvtárosok Egyesülete (MKE) az IFLA (International Advocacy Programme, <https://ifla.org/node/11149>) pályázatán. A világ minden tájáról érkezett, összesen 115 pályázatból az MKE pályázati anyaga bekerült legjobb tíz közé.

Az „Együtt a könyvtárak érdekérvényesítéséért!” (Let’s work together on the library advocacy!

Regional workshop for librarians from the Czech Republic, Poland, Romania, Slovakia, the Ukraine and Hungary”) című pályázat célja egy regionális konferencia megszervezése 2018 tavaszán, melyen cseh, lengyel, román, szlovák, ukrán és magyar könyvtárosok részvételére számítanak. A szervezésben a Lengyel Könyvtárosok Egyesülete (SBP) lesz az MKE partnere. A rendezvényt azt szeretnék demonstrálni, hogy a könyvtárak magukénak vallják az ENSZ 2030-ig tartó, a fenntartható fejlődéssel kapcsolatos programjainak céljait, és a kormányzatok partnereiként részt vállalnak a helyi, regionális és nemzetközi szintű megvalósításban, és közvetíteni szeretnék az érdekérvényesítéshez szükséges ismereteket.

IFLA-HUN, 2017. július 3. (Barátné Hajdu Ágnes tájékoztatása alapján Haraszi Katalin híre)