



KÖNYVTÁRI INTÉZET

Bibliográfiai számbavétel 2021

KÖNYV

KÖNYVTÁR

KÖNYVTÁROS



2021/10

# A digitális könyvtári szolgáltatás új lehetőségei az olvasási képességet érintő fogyatékossgal élők számára

## Az OSZK és a Könyvtári Intézet ajánlása a Marrákesi szerződés átültetésével kapcsolatos könyvtári feladatok ellátásával kapcsolatban

A szerzői jogi szabályozás egyik legfontosabb célja, hogy egysúlyt teremtsen a kreatív gazdaság alapjait jelentő *alkotói érdekek* és az információhoz és kulturális javakhoz való hozzáféréshez fűződő társadalmi érdekek között. Ebből következően a törvény *kivételeket* fogalmaz meg, illetve *korlátozásokat* vezet be a szerzőnek a mű felhasználásának engedélyezésére vonatkozó kizárólagos jogával kapcsolatban. E korlátozásokat és kivételeket nemzetközi szerződések, uniós irányelvek határozzák meg.

Az olvasási képességet érintő fogyatékossgal élő személyeknek kíván kedvezni az a 2017 szeptemberében megjelent EU-s rendelet, amely az úgynevezett Marrákesi szerződés átültetéséről szól. A rendeletet ismertető hivatalos összefoglaló szerint:

„A rendelet szabályokat határoz meg arra vonatkozóan, hogy egyes művek vagy más teljesítmények hozzáférhető formátumú példányait hogyan osztják meg az EU és a Marrákesi szerződésben részes nem EU-országok között a vakok, látáskárosultak és nyomtatott szöveget egyéb okból használni képtelen személyek érdekét szolgálva, és anélkül, hogy a szerzői jogi jogosult engedélyére szükség lenne.”

Az átültetésről gondoskodó törvénynek köszönhetően új szabad felhasználási esetek kerültek be a szerzői jogi törvényünkbe. Kifejtve:

„41. § (1) A mű nem üzletszerű felhasználása a szabad felhasználás körébe tartozik, ha az kizárólag fogyatékossgal élő személyek javára, fogyatékossgal közvetlen összefüggésben történik, és nem haladja meg a fogyatékossgal által indokolt mértéket.

(1a) Az (1) bekezdés szerinti szabad felhasználásnak minősül különösen:

a) a mű hozzáférhető formátumú példányának az olvasási képességet érintő fogyatékossgal élő kedvezményezett személy kizárólagos használatára történő előállítás, akár a kedvezményezett személy, akár valamely, az ilyen személy nevében eljáró személy vagy jogosított szervezet által, feltéve, hogy a hozzáférhető formátumú példány alapjául szolgáló műhöz a hozzáférhető formátumú példány előállítója jogszerűen jutott hozzá;

b) a jogosított szervezet által az olvasási képességet érintő fogyatékossgal élő kedvezményezett személyek vagy más jogosított szervezetek számára hozzáférhető formátumú példányok terjesztése és nyilvánossághoz közvetítése, ideértve az oly módon történő hozzáférhetővé tételt is, amikor a nyilvánosság tagjai a hozzáférés helyét és idejét egyénileg választják meg.

(1b) A Magyarországon letelepedett jogosított szervezet az (1a) bekezdés b) pontjában meghatározott tevékenységet az Európai Unió bármely tagállamában letelepedett jogosított szervezet vagy az olvasási képességet érintő fogyatékossgal élő kedvezményezett személy számára szabadon végezheti.

(1c) A Magyarországon letelepedett jogosított szervezet és az olvasási képességet érintő fogyatékossgal élő kedvezményezett személy a hozzáférhető formátumú példányt az Európai Unió más tagállamában letelepedett jogosított szervezettől szabadon megszerezheti.

(1d) Az (1a)-(1c) bekezdésben meghatározott szabad felhasználás körében bármely nyilvánosságra hozott, írott vagy egyéb jelrendszer formájában rögzített [18. § (1) bekezdés a) pontja és (2) bekezdése] mű felhasználható annak illusztrációival együtt, ideértve ezek digitális és hanganyagga alakított változatait is.”

A szabad felhasználás részletes szabályait külön kormányrendelet határozza meg, amely kitér a *hozzáférhető formátumú példány*, a *jogosított szervezet* és az *olvasási képességet érintő fogyatékossgal élő személy* fogalmára, valamint rendelkezik egy nyilvántartás felállításáról és vezetéséről az említett szervezetekről és az általuk előállított hozzáférhető formátumú művekről és más kapcsolódó jogi teljesítményekről.

Az OSZK a Könyvtári Intézettel és az ELTE Egyetemi Könyvtárral együtt közös ajánlást dolgozott ki a témával kapcsolatban az érdeklét könyvtárak számára.

Az ajánlás az alábbi linken érhető el: <https://www.ki.oszk.hu/dokumentumtar/az-oszk-konyvtari-intezet-es-az-elte-egyetemi-konyvtar-koz-os-ajanlasi-marrakesi>.

---

---

# KÖNYV, KÖNYVTÁR, KÖNYVTÁROS

---

---

30. évfolyam, 10. szám

2021. október

## Tartalom

*Bibliográfiai számbavétel és digitális ökoszisztéma*

<b>Dancs Szabolcs:</b> Bibliográfiai számbavétel 2021 – Könyvtárak átalakuló szerepkörben .....	3
<b>Ilácsa Szabina:</b> Információs korok és a bibliográfiai számbavétel .....	4
<b>Ilácsa Szabina:</b> Mik a szabványok és mire jók? .....	7
<b>Dancs Szabolcs:</b> Összecsengő jövőképek .....	10
<b>Ilácsa Szabina:</b> A bibliográfiai számbavétel újragondolása az IFLA LRM-entitások tükrében .....	16
<b>Hubay Miklós:</b> Entitásalapú szemlélet és új hordozóformátum a katalógizálásban .....	20
<b>Bódog András:</b> Azonosítók modern könyvtári környezetben .....	23
<b>Ilácsa Szabina:</b> „Olyan jó vagyok, mint te”: a szakértelem halála és az entitásmenedzsment az internet korában .....	34
<b>Hubay Miklós:</b> Tudásgráfok és többnyelvű visszakereső felületek .....	37
<b>Bódog András:</b> Röviden a könyvtárak webes láthatóságáról és a schema.org-ról .....	42
<b>Dancs Szabolcs:</b> Metaadat-menedzsment a könyvtári falakon tűlmutató összefogásban .....	45
<b>Bódog András:</b> Az osztályozás és a tárgyszavazás modern könyvtári környezetben .....	47

## From the contents

<i>Szabolcs Dancs:</i>	Bibliographic control 2021 – the transforming roles of libraries (3)
<i>Szabina Ilácsa:</i>	Rethinking bibliographic control in the light of IFLA LRM entities (16)
<i>Miklós Hubay:</i>	Entity-based approach and new communication format in library cataloguing (20)
<i>András Bódog:</i>	Identifiers in a state-of-the-art library environment (23)

## Cikkeink szerzői

*Bódog András*, az OSZK Könyvtári Intézet Könyvtári Szabványosítási Irodájának munkatársa; *Dancs Szabolcs*, az OSZK főtanácsosa, a Könyvtári Intézet Könyvtári Szabványosítási Irodájának irodavezetője; *Ilácsa Szabina*, az OSZK Könyvtári Intézet Könyvtári Szabványosítási Irodájának metaadat-szakértője; *Hubay Miklós*, a Petőfi Irodalmi Múzeum humáninformatikusa

### Szerkesztő:

Tóth Béla István

### Szerkesztőbizottság:

Barátné Hajdu Ágnes, Bánkeszi Lajosné, Dancs Szabolcs, Mezey László Miklós, Németh Márton, Radó Rita, Rózsa Dávid, Szeifer Csaba, Szépvölgyi Katalin, Verók Attila

---

A szerkesztőség címe: 1827 Budapest I., Budavári Palota F épület; Telefon: 224-3797; E-mail: [3k@oszk.hu](mailto:3k@oszk.hu); Internet: [ojs.elte.hu/3k](http://ojs.elte.hu/3k), [ki.oszk.hu/3k](http://ki.oszk.hu/3k), [epa.oszk.hu/3k](http://epa.oszk.hu/3k), [facebook.com/konyvkonyvtarkonyvtaros](https://facebook.com/konyvkonyvtarkonyvtaros)

Közreadja: a Könyvtári Intézet

Felelős kiadó: **Fehér Miklós**, a Könyvtári Intézet igazgatója

Olvasó- és technikai szerkesztő: **Cz. Tóth János**

Borítóterv: **Gerő Éva**

Nyomta a NALORS Grafikai Nyomda, Vác

Felelős vezető: **Szabó Gábor**

Terjedelem: 8,25 A/5 kiadói ív.

Lapunk megjelenését támogatta az

**Emberi Erőforrások Minisztériuma**  
**Nemzeti Kulturális Alap**



Terjeszti a Könyvtári Intézet

Előfizetési díj 1 évre 4800 Ft. Egy szám ára 400 Ft.

**HU ISSN 1216-6804 (Nyomtatott)**

**HU ISSN 2732-0375 (Online)**

# Bibliográfiai számbavétel és digitális ökoszisztéma

## A 2021-es firenzei Bibliographic Control konferencia meghatározó gondolatai<sup>1</sup>

### Bibliográfiai számbavétel 2021 – Könyvtárak átalakuló szerepkörben

Az elmúlt évtizedek történései, technikai előrehaladása a könyvtárakat többféle szemszögből, így például információfeldolgozói és -szolgáltatói szerepkörükben is megújulásra készíti. Utóbbi változások főként azokat az intézményeket érintik, amelyek korábban is jelentős mértékben vették ki részüket a bibliográfiai számbavételből. Az információs források feldolgozásában élen járó könyvtárak hamar felismerték az adatcsere és adat-újrahasznosítás fontosságát, amely a publikációk – technológiai fejlődés nyújtotta – képtelen arányú megszorodásával csak tovább növekedett. A *bibliográfiai ügynökségek* (bibliographic agency<sup>2</sup>) számára a szerepváltozás együtt jár a kiadókkal, könyvterjesztőkkel, a forgalomban lévő könyveket számon tartó úgynevezett books-in-print szervezetekkel való együttműködésre való nyitottsággal. A nemzeti bibliográfiai ügynökségeknek be kell látniuk, hogy képtelenek egyedül megbirkózni a nemzeti kiadványtermés feldolgozásával, főleg, ha azt a kor szellemének megfelelő technológiák és felhasználói elvárások mentén kívánják megvalósítani. Az együttműködés különös haszna lehet számukra a kiadványtermésre való

<sup>1</sup> *Bibliographic Control in the Digital Ecosystem: International Conference*, Firenze, 2021. február 8–12. <https://www.bc2021.unifi.it/>

<sup>2</sup> Korábban „bibliográfiai műhely”-ként is utaltak ezekre.

nagyobb rálátás, míg az együttműködő szervezetek profitálhatnak a metaadatok – könyvtári szaktudás által szavatolt – minőségéből és hitelességéből. A szerepkörükben megújuló bibliográfiai ügynökségek gondozhatják majd többek között azokat a kontrollált (tulajdon- és köz)névállományokat (nemzeti entitásfájlokat, authority állományokat), amelyek felhasználási lehetőségei messze túlmutatnak az intézményi falakon, és talán egészen a mesterséges intelligencia világáig terjednek.

A 2021-es nagyszabású Bibliographic Control konferencia, amelynek előadásai alapján ezt az új helyzetet – a 3K folyóirat eme tematikus számában – az olvasó elé tárjuk, komoly jelentőséggel bírhat a könyvtárak feladatainak újragondolása szempontjából, valamint útmutatóként szolgálhat a nemzeti könyvtárügy céljainak, missziójának megfogalmazásakor. Nem utolsósorban fontos áttekintést adhat a szakmai utánpótlás számára a nemzetközi könyvtárügy jelen tendenciáiról.

**Dancs Szabolcs**

## Információs korok és a bibliográfiai számbavétel<sup>3</sup>

Az emberiség eddigi történetében négy olyan találmány volt, amely teljes mértékben átalakította az információhoz való viszonyulásunkat. Az adott találmánynak minden esetben volt egy azonnali hatása az emberi információközvetítésre és a kulturális emlékezet rögzítésére, illetve volt egy lassabb, de annál mélyrehatóbb következménye, ami az emberi élet minden területére hatással bírt. Ez a négy találmány az *írás*, a *nyomtatás*, a *telekommunikáció* és az *internet*. A találmányok megjelenése alapján az emberiség története öt információs kora osztható. Az egyes információs korok hossza (eddig legalábbis) nagyságren-

<sup>3</sup> Gordon Dunsire *Bibliographic control in the fifth information age* című előadása alapján. Elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 12. <https://www.youtube.com/embed/sPrefNT8wG0?start=1991&end=4057> (2021.10.20.)

dekkel csökkent az azt megelőző korhoz képest. Így a kezdetben több tízezer év mára lecsökkent néhány évtizedre.

	<b>Első információs kor</b>	<b>Második információs kor</b>	<b>Harmadik információs kor</b>	<b>Negyedik információs kor</b>	<b>Ötödik információs kor</b>
<b>A rögzített információ gyakorisága</b>	Ritka.	Gyakori.	Gyakori.	Szükséges.	Elkerülhetetlen.
<b>A kulturális emlékezet rögzítése és terjesztése</b>	Nem rögzítik, csak szóban történik.	Rögzítik ugyan, de a terjesztése még nem megoldott, a szóbeliséget még nem váltja le.	A kulturális emlékezet rögzítése mindennapos, terjesztésére külön szakmák alakultak ki.	A kulturális emlékezetet kódolva, hírközlési csatornákon keresztül továbbítják, majd a célnál dekódolják.	Minden rögzítésre kerül, a terjesztés egyszerű és szinte azonnali.
<b>A rögzített kulturális emlékezet tartalma</b>	Rajzolt ábrák és szobrok.	Többnyire kereskedelmi, jogi és vallási információk kerülnek rögzítésre.	Megjelenik a tudományos kommunikáció is.	Lehetséggé vált audiovizuális tartalmak előállítására és továbbítására.	Minden olvasó egyben szerző is, minden rögzítésre kerül.
<b>Hordozó</b>	A hordozó egyedi, a reprodukció drága.	A hordozó egyedi, a reprodukció egyszerűbb, de még mindig drága.	A hordozók több példányban léteznek.	A hordozó megfoghatatlanná válik, a reprodukció szükség-szerű.	A hordozó többféle digitális formátumban létezik, a reprodukció elkerülhetetlen.
<b>Hozzáférés</b>	Nehézkés, helyhez kötött vagy törekeny információ-hordozók.	A hozzáférés kevésbé nehézkes, de még mindig a helyhez kötött információhordozó dominál.	A hozzáférés egyszerű, a példány könnyen mozgatható a könyvkereskedőtől vagy a könyvtárból.	A hozzáférés térben és időben kötött. Az információ megy a felhasználóhoz.	Bárki, bárhol, bármikor hozzáfér az információhoz.

A nyomtatott kiadványok volumenének növekedésével és könnyű hozzáférhetőségével együtt emelkedett a rögzített emlékezetet megőrizni kívánó egyének és csoportok száma. A kulturális emlékezet szisztematikus gyűjtése már a második információs korban megjelent, ám akkor a kiadványok magas költségei miatt még ritka volt.

Ahogy a gyűjtemények tartalma növekedett, úgy vált egyre inkább szükségessé, hogy valamilyen módon rögzítésre kerüljön az, hogy pontosan mi és hol található meg bennük. Hamar világossá vált, hogy ennek a rögzítésnek az alapja a metaadat. A szöveges metaadatok nem csupán lehetővé tették egy kiadvány beazonosítását, de a nyelvi forma lehetővé tett olyan rendezési és csoportosítási formákat, amelyek megalapozták a jelenlegi információ-visszakeresési algoritmusainkhoz vezető utat. A mostani metaadat-szemléletünk tehát a harmadik és a negyedik információs korban gyökerezik.

Miután rutinszerűvé vált az egyes gyűjtemények feltárása, felmerült az igény arra vonatkozóan, hogy valamiféleképpen rögzíthető legyen a világ dokumentumtermésének az összessége. Kezdetben egy nagy központi adatrögzítésben gondolkodtak, de hamar kiderült, hogy ez a gyakorlatban tarthatatlan lett volna. Így a decentralizált megoldások felé fordultak. Az *Egyetemes Bibliográfiai Számbavétel* (Universal Bibliographic Control – UBC) a nemzeti bibliográfiai ügynökségek felelősségi körébe utalja, hogy az országuk dokumentumtermését feldolgozzák. A nemzeti bibliográfiák összessége pedig kiadja a világ dokumentumtermését.

Ahhoz, hogy ez a decentralizált feltárás működjön, az egyes bibliográfiai ügynökségeknek, ha nem is azonos módon, de legalábbis minimális eltérésekkel kell feltárniuk a saját anyagaikat, ehhez pedig közös szabványok kellene. Erre az igényre adott válaszként dolgozták ki a *Nemzetközi szabványos bibliográfiai leírási szabályzatot* (International Standard Bibliographic Description – ISBD) és a UNIMARC-ot.

Azt még csak találgatni tudjuk, hogy az ötödik információs kor pontosan hogyan fogja átalakítani a bibliográfiai számbavételről alkotott elképzelésünket, de az már most is látszik, hogy a változás jelentős lesz. Tematikus számunk további részeiben azokat a trendeket és jelenségeket gyűjtöttük egybe, amelyek minden bizonnyal kulcsszerepet játszanak majd ebben az átalakulásban.

**Ilácsa Szabina**

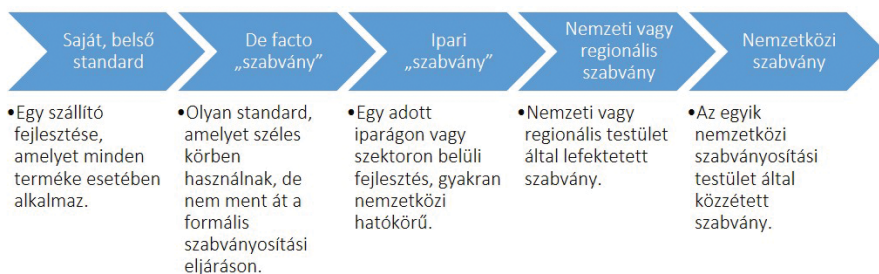


## Mik a szabványok és mire jók?<sup>4</sup>

„A szabványok általános és ismételten alkalmazható eljárásokat és műszaki megoldásokat határoznak meg, amelyeket közmegegyezéssel fogadtak el, és optimális megoldást kínálnak a különböző érdekelt felek számára.”<sup>4</sup>

A szabványok előírásainak betartása teszi lehetővé a szolgáltatások (köztük a könyvtári és szakirodalmi szolgáltatások) megbízhatóságának és hatékonyságának növelését, egységességük és átjárhatóságuk biztosítását.

Az angol *standard* szót gyakran fordítják automatikusan szabványnak, pedig a magyar kifejezés jóval szűkebb értelmű. Magyarul szabványnak csak az a standard nevezhető, amely átesett a szabványosítási folyamaton. Ennek ellenére a de facto és az ipari standardokat is szokás szabványnak nevezni, mivel alkalmazásukat tekintve szabványszerűek, ha jogi értelemben véve nem is azok.



*„Szabványok” típusai a legkevésbé formálistól a leginkább formális felé haladva (David Haynes ábrája alapján)<sup>6</sup>*

<sup>4</sup> Renate Behrens *Standards in a new bibliographic world – community needs versus internationalisation* című előadás alapján. Elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 8. <https://youtu.be/ybUDrILt0kI?t=5224> (2021.10.20.)

<sup>5</sup> Forrás: a Magyar Szabványügyi Testület honlapja – <https://prod.mszt.hu/hu-hu/gyakranismetelt-kerdesek> (2021.10.20.)

<sup>6</sup> Az ábra eredetijének forrása: Haynes, David: *Metadata for information management and retrieval. Understanding metadata and its use*. London, Facet, cop. 2004. 50. p.

Könyvtári területen jellemzően nemzeti vagy nemzetközi szabványokkal dolgozunk, de előfordul olyan is, hogy egy adott könyvtár gyakorlata válik de facto szabvánnyá. Olyan is előfordul, hogy egy más szektorra kidolgozott szabvány a könyvtári területen is fontossá válik. A szabványkövetés nyújtotta átjárhatóság ugyanis nem áll meg a szakterületek határán, segít azt is biztosítani, hogy az egymáshoz kapcsolódó területek közötti információáramlás megfelelő legyen. A könyvtári tevékenységnek szabványok szempontjából két fontosabb határterülete van: ezek a könyvtári források előállítóinak és terjesztőinek szabványai, valamint az informatika. Ez utóbbi az élet minden területére bevette már magát, az informatikai szabványok jelentősége bizonyos szempontból az egyes írásrendszerek betűalakjainak megállapodásszerű használatához mérhető.

## **Kinek van szüksége szabványokra?**

Mindenkinek. Ha nagyon szigorúan vesszük, akkor szabványok nélkül nincsen könyvtár, sőt civilizáció sem. Az írásjelek „szabványosítása” nélkül nem alakulhatott volna ki az írásos kommunikáció. Az írás által lehetőség nyílt arra, hogy tőlünk térben és időben távol lévő dolgokról információt szerezzünk. Az írás a kommunikációt, a szabványosítás megjelenése pedig a munkamegosztás területét forradalmasította. A szabványok segítségével tőlünk térben és esetleg időben távol lévő emberekkel tudunk együtt dolgozni egy nagy, összetett feladat részfeladatain úgy, hogy munkánk a végén hézagmentesen illeszkedni fog egymáshoz.

A szabványok olyannyira körbevesznek bennünket, hogy nehéz elképzelni az életünket nélkülük. Ahhoz, hogy legyen egy balesetvédelmi szempontból megfelelő könyvtárépületünk, több tucat szabványnak kell megfelelnünk. Ám nemcsak a fizikai világban vesznek minket körül szabványok, hanem a számítógépekkel hozzáférhető – és az életünkben egyre fontosabb szerepet játszó – virtuális világban is. Az informatika kitermelt magából legalább annyiféle szabványt, mint amennyi a fizikai életünket könnyíti meg, ha nem többet, ami főleg annak fényében lenyűgöző, hogy végső soron mindegyik csak „nullákra és egyesekre” épül.

A szabványok közül könyvtári környezetben főleg az adatok tárolásával, tartalmával és cseréjével foglalkozó szabványokra támaszkodunk a munkánk során. Ha elektronikus környezetben bármit cserélni vagy akár csak megjeleníteni szeretnénk, szabványokhoz kell alkalmazkodnunk. Annál könnyebben – és ezzel összefüggésben annál költséghatékonyabban – lehet cserélhető, mások által megtekinthető vagy újrafelhasználható tartalmat létrehozni, ha minél kevesebb vetélkedő szabvány van, vagyis mindenki ugyanarra a célra ugyanazt a szabványt használja. Világunk azonban túlzottan sokszínű a fűnyíróelví egységmegoldásokhoz. Ezért meg kell elégednünk azzal, hogy a számunkra fontos közösségekkel használunk közös szabványokat, amelyeket a többi közösség szabványainak megfeleltetünk, már ahol lehetséges. Ezek a közösségek lehetnek nyelvi vagy foglalkozás alapján

szervezettek, de lehetnek tágabban meghatározottak is, például olyanok, amelyek szeretnének információt vagy tartalmat cserélni egymással, esetleg közös céljaik vannak, amiken együtt szeretnének munkálkodni, és fenntartható eredményeket szándékoznának elérni.

A szabványosítás lassú és költséges folyamat. Saját szabvány alkotása esetén a költségek nagy részét az emberierőforrás-ráfordítás teszi ki, míg egy nemzetközi szabvány honosításánál a költségek megoszlanak a szabványosítás intézményesített csatornájának díja és az emberierőforrás-ráfordítás között. A jó szabvány viszont hosszú távon behozza ezeket a költségeket, mivel segítségével hatékonyabban tudunk működni.

Fontos kiemelni a szabványok tekintetében, hogy azok mindig a megalkotásukkor érvényes helyzetre optimalizált megoldást írnak le. Ebből az következik, hogy egy szabványnak fontos hozadéka a stabilitás, amit teremt, de egy teljesen változatlan szabvány nem feltétlenül felel meg a szabványosítás céljának. A környezeti feltételek és a rendelkezésre álló technológiai keretek ugyanis változhatnak, így az is változhat, hogy mi az optimális megoldás. Ezért indokolt a szabványokat időről időre felülvizsgálni.

Jelenleg éppen egy ilyen megújító ciklus közepén vagyunk. A technológiai fejlődés elért odáig, hogy a világhálón lévő adatok legfőbb felhasználói már nem az emberek, hanem az algoritmusok, így az adattögzítésünket ennek megfelelően át kell gondolnunk. Néhány példa erre a megújulásra a 2016-os *Nyilatkozat a nemzetközi katalógizálási alapelvekről* (ICP), az IFLA *könyvtári referenciamodell* és a *3R projekt* utáni RDA (Resource Description and Access). Jelenleg zajlik az adattartalmi szabványok közül az ISBD felülvizsgálata, az adatstruktúra-szabványok közül a MARC 21 és a UNIMARC elemkészletének átdolgozása, de ide sorolhatók a BIBFRAME-mel való kísérleti projektek is.

**Ilácsa Szabina**

<sup>7</sup> *Statement of International Cataloguing Principles (ICP)*. By IFLA Cataloguing Section and IFLA Meetings of Experts on an International Cataloguing Code. 2017. 21 p. [https://www.ifla.org/files/assets/cataloguing/icp/icp\\_2016-en.pdf](https://www.ifla.org/files/assets/cataloguing/icp/icp_2016-en.pdf) (2021.10.20.) Magyar fordítása: Nyilatkozat a nemzetközi katalógizálási alapelvekről (ICP). Közreadja az IFLA Katalógizálási Szekciója és az IFLA Szakértői Tanácskozása a Nemzetközi Katalógizálási Szabályokról (IFLA Meetings of Experts on an International Cataloguing Code); Ford. Gazdag Tiborné, 2018. 21 p. [https://www.oszk.hu/sites/default/files/ICP\\_2016\\_magyarul\\_2019\\_09\\_27.pdf](https://www.oszk.hu/sites/default/files/ICP_2016_magyarul_2019_09_27.pdf) (2021.10.20.)

<sup>8</sup> Riva, Pat – Le Boeuf, Patrick – Žumer, Maja: *IFLA Library Reference Model: A conceptual model for bibliographic information*. IFLA, 2017. 101 p. [https://repository.ifla.org/bitstream/123456789/40/1/ifla-lrm-august-2017\\_rev201712.pdf](https://repository.ifla.org/bitstream/123456789/40/1/ifla-lrm-august-2017_rev201712.pdf) Magyar fordítása: IFLA könyvtári referenciamodell: A bibliográfiai információk elméleti modellje. Ford. Gazdag Tiborné, IFLA, 2018. 99 p. [https://repository.ifla.org/bitstream/123456789/46/1/ifla\\_lrm\\_2017\\_hun\\_v3.pdf](https://repository.ifla.org/bitstream/123456789/46/1/ifla_lrm_2017_hun_v3.pdf) (2021.10.20.)

# Összecsengő jövőképek

## 1. Az Egyetemes Bibliográfiai Számbavétel és a digitális ökoszisztéma: lehetőségek és kihívások<sup>9</sup>

*Mauro Guerrini*, a Firenzei Egyetem professzora, az IFLA számos munkacsoportjának és szekciójának tagja, az Olasz Bibliográfiai és Könyvtári Társaság (SIBB) társalapítója, nem utolsósorban az RDA olasz nyelvre való átültetésének irányítója, akinek neve a szakma hazai képviselői számára is ismerősen cseng. A konferencia bevezetésének szánt felszólalását egy *Hans-Georg Gadamer*-idézettel kezdte, amely szerint „[a] kultúra az emberiség egyetlen olyan vagyona, amely, ha megosztjuk egymás közt, inkább nagyobb lesz, mint kisebb.”

Az Egyetemes Bibliográfiai Számbavétel gondolata mindig is az IFLA egyik központi témája, víziójának fontos eleme volt. Ez irányú elkötelezettségét a nemzetközi szervezet az idők során többször is megerősítette. A UBC mindenekelőtt a katalógizálás történetében tölt be fontos szerepet. A bibliográfiai számbavétel témájának szentelt konferencia azt vizsgálta, hogy miként változtak a számbavétel határai a digitális korban. Az új bibliográfiai modellek az adatcsere határait is kitolják a homogén katalógizálási rendszereken túl a nyitott, kulturális és nyelvi tradíciókon átívelő dialógus irányába. A forma helyett a tartalom válik hangsúlyossá, amely tartalomhoz egy adatkészlet társul, a metaadatok készlete. Így válik a metaadat a bibliográfiai számbavétel központi paradigmájává. E paradigma azután hatással van a rögzített ismeretre. Ilyen hatás például, hogy a mű metaadatainak minősége kihat a mű megismerésére, végső soron annak sikerére. A metaadatok létrehozása nyomon követhető az alapadatokat (pl. címet) rögzítő szerzőtől kezdve a már szabványos azonosítókat alkalmazó kiadón és az intellektuális adatgazdagítást végző bibliográfiai ügynökségen keresztül egészen a keresőrendszerekig. A metaadatolás a nyílt hozzáférés (Open Access) modelljében publikált tartalmaknál sem marad el.

<sup>9</sup> Mauro Guerrini előadása alapján, lásd *Universal bibliographic control in the digital ecosystem: opportunities and challenges*. Elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 8. <https://youtu.be/ybUDrILt0kI?t=2934> (2021.10.20.)

Az adatok újrahasznosulását és összekapcsolását központba állító digitális világban a bibliográfiai számbavétel dinamikussá válik, a nemzeti könyvtárak, illetve a bibliográfiai ügynökségek pedig elvesztik monopolhelyzetüket, jóllehet szerepük továbbra is jelentős, ami – mindenekelőtt – a bibliográfiai szabványok terjesztését, a metaadatok létrehozásában érdekelt szervezetek közötti közvetítést illeti.

Amellett tehát, hogy a könyvtár továbbra is fontos részét képezi a digitális ökoszisztémának, érdemes átgondolni, a digitális átállás miként érinti a könyvtári katalógusokat, illetve a metaadatok létrehozásának folyamatait. Az egyik fontos következmény, hogy a könyvtárak és a könyvtárosok mellett egyéb intézmények és szervezetek képviselői (kiadók, terjesztők, egyetemi dolgozók, közgyűteményi szakemberek) is közreműködnek a metaadatok és authority állományok rögzítésében, adatgazdagításában. További következmény a könyvtárak, kiadók, terjesztők közötti stratégiai partnerségek kialakulása. A nemzeti bibliográfiai ügynökségek a minőséget szavatolják, illetve felelősek az authority kontrollért, amiben olyan eszközökre építhetnek, amilyen a VIAF (Virtual International Authority File = nemzetközi virtuális authority állomány) vagy az ISNI (International Standard Name Identifier = nemzetközi szabványos névazonosító) nemzetközi azonosítók, amelyek egyben a közgyűteményi rendszerek közötti interoperabilitást is szolgálják.

Fontos megemlíteni, hogy jelentősen változott a felfogás a visszakeresés szempontjából megfelelőnek tekintett névalak kiválasztását illetően. A különböző kulturális (nyelvi, írásrendszeri, irodalomtörténeti stb.) kontextusukból következően eltérő névalakok a visszakeresés szempontjából ma ekvivalensnek tekintendők, és a modern szoftveres megoldásoknak köszönhetően egymáshoz kapcsolva egy-egy külön osztályt (klasztert) alkotnak. Korábban a nyugati világ bibliográfiai tradíciója diktált, a mai globális dimenziót öltő kommunikációban evidencia, hogy a nevek megjelenítése annak a kulturális és nyelvi kontextusnak a függvénye, amely a nevet használó adatkészletet jellemzi. Ezt a felfogást támogatja az *IFLA könyvtári referenciamodell*, amely szerint egyetlen *entitás*hoz több, egymástól eltérő, ugyanakkor egyaránt érvényesnek tekinthető *nomen entitás* kapcsolható. A mostani szoftveres megoldások alkalmasak arra, hogy kontextustól függően jelenítsék meg a névváltozatokat.

A metaadatok világában lezajlott változások az együttműködés modelljeinek kialakítását igénylik, tekintettel az eltérő katalogizálási tradíciókra. A szabványokat tehát úgy kell fejleszteni – a használói közösségekkel szoros együttműködésben –, hogy egyfelől megbízható módon szolgálják az interoperabilitást, másfelől kellőképp rugalmasnak bizonyuljanak a változatos használói igényeknek való megfelelés érdekében. Nem kizárt ugyanakkor, hogy ma csak az út elején járunk, ami a metaadat-menedzsmentet és az authority kontrollt illeti. A gépi tanulás vagy a mesterséges intelligencia mindkét területen hozhat újabb meglepetéseket.

## 2. Egyetemes Bibliográfiai Számbavétel a szemantikus webben: a bibliográfiai adatmodellek összehangolásának lehetőségei és problémái<sup>2</sup>

*Tiziana Possemato*, a Firenzei Egyetem munkatársa szerint az Egyetemes Bibliográfiai Számbavétel és a szemantikus web modelljeinek közös célkitűzése, hogy a tudáshoz való hozzáférést az entitások és kapcsolataik köré szervezett dokumentumuniverzum megragadásán keresztül tegyék minél szélesebbé. A szemantikus web paradigmája a relatíve homogén katalogizáló rendszerek közötti rekordcserén túlmutatva az egyes szereplők és rendszerek közti – nyelvi, földrajzi és tematikai korlátokon átívelő – nyílt eszmecsere színtereként szolgálhat. Ez az új dialógus olyan heterogén felhasználói közösségekre is kiterjed, amelyekkel a katalogizálási hagyomány eddig nem számolt, ahogy az újrahaznosítható adatokra épülő újfajta tudásokkal sem. Kérdés, hogy az eltérő katalogizálási tradíciók az új környezetben alkalmasak-e a párhuzamos „együttélésre”, megőrizve korábbi értéküket.

Az 1970-es években az IFLA berkeiben működött a UBCIM Iroda (Universal Bibliographic Control and International MARC = Egyetemes Bibliográfiai Számbavétel és Nemzetközi MARC), amelynek célja a hatékony nemzetközi együttműködés előmozdítása volt, végső soron pedig az, hogy minden dokumentumról egyetlen rekord készüljön (lehetőleg a kiadás országában), amelyet azután a világ összes könyvtárában, illetve bibliográfiai ügynökségében felhasználhatnak. Ehhez mindössze következetesen kell világszerte alkalmazni a bibliográfiai és authority rekordok standardjait. Az 1990-es években azután ráébredtek, hogy a felhasználó nyelve és írásrendszere, kulturális hovatartozása ugyancsak fontos tényező, valamint a webtechnológia új eszközöket kínál az adatmegosztáshoz. Így aztán a 2000-es évek első évtizedében a Kongresszusi Könyvtár munkatársaiban megfogalmazódott, hogy a szöveges leírásokat géppel feldolgozható adatkészletként kell alakítani; az adatelemeket a web kontextusában egyedileg azonosíthatóvá kell tenni; az adatoknak a webtechnológia szempontjából kompatibilisnek kell lenniük, valamint szükség van az interoperabilitásnak egy új, a web valóságára kifejlesztett nyelvére.

<sup>10</sup> Tiziana Possemato *Universal bibliographic control in the semantic web. Opportunities and challenges for the reconciliation of bibliographic data models* című előadása alapján. Elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 8. <https://youtu.be/ybUDrILt0kI?t=6195> (2021.10.20.)

LC control no.: no201816161  
LCN Permink: <https://www.loc.gov/psn/no201816161/>

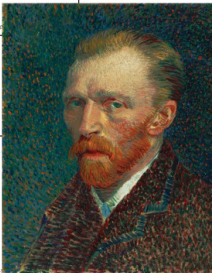
**HEADIN:** Gogh, Vincent van, 1863-1911  
000 001 712 47080 634 830  
001 091 4209  
005 001117071 63 0  
006 1811263 42266616 6 634 c  
010 18 00201816116  
015 18 010201816167562  
040 18 000 0 0 0 0 0 0 0 0 0  
044 18 1300 0 1911  
100 18 18 000 18 000 18 000 18 000  
400 18 18 000 18 000 18 000 18 000  
670 18 18 000 18 000 18 000 18 000

Authority record for Vincent van Gogh, Marc21 (LOC catalogue)

Scheda Unimarc: 1 ▶ Esito ▶ Stampa ▶ Scarico Unimarc

LEADER: 000000 22018161 45  
001 1#CUCJCN#002047  
005 201810000000303.8  
010 \$a0000000120805689  
100 \$a2014012182aaz01 da0  
102 \$aNL  
152 \$aREICAT  
200 18aGogh, Vincent / van  
300 \$a1863-1890 / Pitor, disegnatore e incisore, nato a Zundert (Brabant) e morto a Auvers-sur-Oise (Oise)  
400 \$aVincent van Gogh, Vincent#1#CUCJCN#002047#0  
801 \$a#1863-1890#CUCJCN#002047#01#07  
810 \$aCatalogo in linea della Bibliothèque Nationale de France: <http://catalogue.bnf.fr>  
810 \$aWorld Biographical Index, Internet edition, K. G. Saur Electronic Publishing München; [www.swbi.de](https://www.swbi.de)  
810 \$aCatalogo in linea della Library of Congress <http://catalog.loc.gov>  
810 \$aEnciclopedia italiana di scienze lettere ed arti, Roma, Istituto della Enciclopedia italiana, 1929-

Authority record for Vincent van Gogh, Unimarc (SBN catalogue)



Van Gogh's portrait

001 0x 422 45  
001 0000011927919  
003 <https://www.loc.gov/psn/no201816161/>  
005 20141119  
010 18 0000000120805689 62 6AF 62 20130724  
030 18 00 64 001597979 62 6AA  
035 18 00 64 001446316 62 6AP  
039 18 00 64 1463164 62 20086107  
039 18 00 64 1463167 62 20140417  
100 18 18 000 18 000 18 000 18 000  
101 \$a file  
102 \$a NL  
105 \$a 18630330 18900729  
106 \$a a  
108 \$a 010  
120 \$a 0  
150 \$a 2  
200 18 18 000 18 000 18 000 18 000 18 000  
300 18 18 000 18 000 18 000 18 000  
301 18 18 000 18 000 18 000 18 000  
330 \$a Comu et France sous le nom de "van Gogh". La particule "van" est exclusivement masculine et elle se voit bien que van Gogh est hollandais  
400 18 18 000 18 000 18 000 18 000 18 000 18 000 18 000  
400 18 18 000 18 000 18 000 18 000 18 000 18 000  
400 18 18 000 18 000 18 000 18 000 18 000 18 000  
400 18 18 000 18 000 18 000 18 000 18 000 18 000  
400 18 18 000 18 000 18 000 18 000 18 000 18 000  
400 18 18 000 18 000 18 000 18 000 18 000 18 000  
400 18 18 000 18 000 18 000 18 000 18 000 18 000  
810 \$a Vincent van Gogh par lui-même: recueils de lettres, de dessins et d'œuvres de la correspondance au portrait réalisé par Auguste Bernard, 1960  
810 \$a Gogh, Vincent, 1863-1890. Sa vie. Sa NCL Authority file, 2009  
810 \$a Brief Service japonais SA BN CG, gen.

Authority record for Vincent van Gogh, Unimarc (BNF catalogue)

*Van Gogh, ahogy az authority rekordok láttatják (Részlet Tiziana Possemato előadásának diából)*

A nemzeti bibliográfiai ügynökségek lokális feladata a nemzeti kiadványtermés számbavétele, globális „hivatása” pedig a kiadványokról készült adatok megosztása, széles körű elérhetővé tétele, valamint szabványos hozzáférési pontok meghatározása az adott országhoz kapcsolódó személyek, családok, testületek, művek stb. tekintetében. Mindegyik területen sokáig a (bibliográfiai, illetve authority) rekordé volt a főszerep, mígnem megtörtént az elmozdulás a rekordtól az entitás irányába. Más megközelítésben ez az identitástól, az entitás egyféle reprezentációjától való elmozdulást jelent az azonosításhoz hasznos tulajdonságok és kapcsolatok sorozatából álló entitás felé. Még egy újabb megközelítéssel megtoldva: ez az elmozdulás az üres leírásnak a valós világbeli objektummal való leváltását célozza, vagyis az identitás kifejeződéseinek („rekordba zárt”) szintézise helyett az egyedi és megismételhetetlen entitást tolvá az előtérbe. Possemato szerint mindez arra emlékeztet, mintha egy kétdimenziós, statikus szemléletet egy háromdimenziós, dinamikus szemléletre cserélnénk.

Felmerül persze a kérdés, hogy ebben az új világban jut-e szerep a nemzeti bibliográfiai ügynökségeknek, illetve az eltérő katalogizálási tradíciók képesek-e párhuzamosan létezni egymás mellett. Az előadó válasza mindkettőre egyértelmű igen. Az ügynökségek szerepe elsősorban a hitelesség garantálása miatt lényeges, így az is fontos, hogy milyen közösségek húzódnak meg egy-egy ügynökség mögött. Lényegi szerep jut az adatkezelésben a származási információnak (proveniencia), amely felelősséggel ruházza föl az adatgazdát, egyben eszközt ad neki arra, hogy kövesse, mi történik az általa létrehozott adattal, de azt is lehetővé





*Van Gogh 3D-ben mint a valós világ objektuma (Részlet Tiziana Possemato előadásának diáiból)*

teszi a felhasználó számára, hogy szűrőt állítson be (az alkalmazási profiljában) a számára nem hiteles források blokkolására.

Összefoglalva: az univerzális érthetőség és megoszthatóság érdekében a strukturált adat visszanyerte a valódi és lényegi információs erejét. A rekord szerepe az új kontextusban egyre inkább a lokális választások és szükségletek kifejezésére szűkül, nem gátolja az adatot a globálisan megosztható információs tartalom terjesztésében. A helyi és nemzetközi intézmények szerepe e téren újból megerősítést nyerhet, amennyiben képesek lesznek lépést tartani a gyors tempót diktáló fejlődéssel.

### 3. Konferenciázó gondolatok<sup>11</sup>

Az Egyetemes Bibliográfiai Számbavétel hagyományos modellje a nemzeti bibliográfiai ügynökségekre, azok egyenkénti elkötelezettségére épített, míg a digitális környezetben a bibliográfiai számbavétel feladata megoszlik a csomópontok sokasága közt, ahol minden csomópontnak módjában áll a bibliográfiai információt megosztani, gazdagítani és újrahasznosítani.

A digitális ökoszisztémát, illetve a webet nem szabdalják fel a nemzeti határok, így az egyes csomópontok közvetlenül szolgálhatják a UBC-t. A csomópontok sorában elsőként a *kiadó* említendő, amely a publikált forráshoz metaadatokat,



köztük – a bibliográfiai számbavétel szempontjából különösen fontos – azonosítókat szolgáltatókat.

A kulturális iparág nagyban épít a bibliográfiai metaadatokra, amelyek előállításában a UBC alapvető csomópontjainak tekinthető *könyvtárak* járnak az élen, ahogy a digitális átállásban is vezető szerepet töltenek be. A MARC-formátumot a hatvanas években kreálták. Azt mondhatjuk, hogy azóta mind a digitális, mind a hagyományos források metaadatai digitálisan állnak rendelkezésre. Az utóbbi évek könyvtári kezdeményezései újabb átállást készítenek elő, amelynek célja az új technológiák kihasználása az adatcsere lehetőségének megőrzésével. Az új technológiák elsősorban a szemantikus megoldásokat, illetve a kapcsolt adat (Linked Data) megközelítést fedik le, az átállás várt eredménye pedig, hogy a könyvtári adatkészletek bekerülnek a világhálón található összekapcsolt adatkészletek gyűjteményébe. Az átállás ugyanakkor mindkét irányból problémákba ütközik: ami a kiindulási pontot illeti, a MARC-alapú formátumok sokfélesége nem szavatolja az egységes kiindulási alapot; ami az átállás eredményére vonatkozik: nincs egyetlen, közös célformátum. Jóllehet a BIBFRAME-re ma úgy tekintünk, mint egy alapvető fontosságú ontológiára, a metaadatszabványokról és adatmodellekről szóló diskurzus még korántsem zárult le. Fontos volna, hogy az ontológiák ki tudják békíteni az eltérő kulturális és nyelvi hagyományokban gyökerező, egymástól különböző megközelítéseket.

Csomópontnak tekinthetők a *keresőmotorok* is. Lényegében ma már sokan ezeket használják a bibliográfiai információk visszakeresésére, és a metaadatok készítőinek is módjában áll, hogy javítsák az adatok webes fellelhetőségét – például a *schema.org* vagy egyéb standardok alkalmazásával.

Érdekes azt is megemlíteni, hogy 2012-ben a MARC 21-ben megjelent a 883-as adatmező (Metadata Provenance = metaadatok provenienciája), amely lehetőséget nyújt a számítógép által generált adatok rögzítésére. A mesterséges intelligenciának köszönhetően a gépek egyre inkább alkalmasak ilyen metaadatok létrehozására. Ugyancsak terjed a könyvtári világban a Wikidata és a Wikibase használata. Az IFLA külön munkacsoportban foglalkozik a Wikidatával, és hogy annak ontológiáját megfeleltesse a BIBFRAME-mel, az RDA-val és a MARC-kal. A Wikidata egyébként igazi sikertörténetként fogható fel a könyvtári adatok újrahasonosítása szempontjából.

A bibliográfiai számbavétel szorosan összefügg a hosszú távú megőrzés kérdésével. A metaadatoknak nem sok értelme van, ha elvész a hivatkozott tartalom, ez mindenképp a digitális információforrásokat érinti. A nemzeti könyvtárak

<sup>11</sup> Mauro Guerrini, Giovanni Bergamin és Laura Manzoni *Closing remarks* című közös előadása alapján. Elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 12. <https://youtu.be/sPrefNT8wG0?t=11466> (2021.10.20.)

feladata, hogy a kötelezpéldány-szolgáltatáson keresztül hosszú távon biztosítsák az információhoz való általános és egyenlő hozzáférést. Ugyanez az alapja a bibliográfiai számbavétel törvényi kötelezettségének. A kezelendő információtömeg azonban óriási, és egyre csak gyarapszik. Ez okból a nemzeti könyvtáraknak a jövőben a szelekció és a kooperáció elvén kell megszervezniük ez irányú tevékenységeiket.

Az Egyetemes Bibliográfiai Számbavétel eszménye és víziója ma is érvényes, még akkor is, ha ma a számbavétel maga több csomóponton és e csomópontok együttműködésén is alapul. A nemzeti bibliográfiai ügynökségek magtevékenysége továbbra is az authority kontroll, a minőség-ellenőrzés, a szabványok iránti elkötelezettség, a szótárak és teauruszok építése és karbantartása. A digitális ökoszisztémában zajló tevékenységek nemcsak a könyvtári katalógusokra fejtik ki hatásukat, hanem a kultúra egyéb területeire is, végső soron pedig a mindennapi életünkre. Ha ezen a téren el is veszítették a monopóliumukat a könyvtárak, küldetésüknek hála, továbbra is ők maradnak a hitelesség biztosítékai.

A kulturális források „szociális objektumok”, amelyeken keresztül értelmezzük a világot, és amelyeken keresztül interakcióba lépünk vele. A „szociális objektumokhoz” való hatékony hozzáférést a digitális ökoszisztéma kulcsinfrastruktúrájának tekinthető metaadatok biztosítják. Ez okból a bibliográfiai számbavétel a közösségek fennmaradását szolgálja mind gazdasági, mind szociális, egészségügyi vagy akár környezetvédelmi értelemben.

**Dancs Szabolcs**

## A bibliográfiai számbavétel újragondolása az IFLA LRM-entitások tükrében<sup>12</sup>

„Az UBC (Universal Bibliographic Control = Egyetemes Bibliográfiai Kontroll<sup>13</sup>) [...] [c]élja, hogy biztosítsa az összes országban kiadott bármely kiadvány alapvető bibliográfiai adatainak egyetemes és azonnali hozzáférhetőségét bármely, nemzetközileg elfogadott formában. Mindazonáltal a nemzeti bibliográfiai ellenőrzés tekinthető az Egyetemes Bibliográfiai Kontroll legfontosabb előfeltételének. Így [sic!] minden országban létesíteni kell egy nemzeti bibliográfiai ügynökséget, amely hivatott az adott országban kibocsátott összes publikáció átfogó bibliográfiai feljegyzéseinek elkészítésére, összhangban azokkal a nemzetközi szabványokkal, amelyeket mind kézi, mind gépesített rendszerekre alkalmazni lehet.” – olvashatjuk az Egyetemes Bibliográfiai Számbavételt életre

hívó, 1977-ben Párizsban tartott *Nemzetközi kongresszus a nemzeti bibliográfiákról* konferenciaanyagának bevezetőjében.<sup>14</sup>

Bár kétségkívül nagy előrelépés volt az Egyetemes Bibliográfiai Számbavétel programja, a technológiai környezet és a használói igények változása miatt az eredeti elképzelés felülvizsgálatra szorul. Az egyik jelentős változás a bibliográfiai számbavétel koncepciója kidolgozásának ideje és a jelen között a használói érdeklődésre számot tartó források körének bővülésében és a források számának ebből is fakadó exponenciális növekedésében mutatkozik meg. Ezzel párhuzamosan csökkent a könyvtári feldolgozó személyzet létszáma, bár ennek a jelenségnek a hatását valamelyest mérsékeltek az ugyanebben az időben lefolyó gépesítési projektek. A végeredmény ugyanakkor így is úgy összegezhető, hogy robbanásszerűen megnövekedett mennyiségű anyagot kell egyre kevesebb embernek feldolgoznia.

A használói elvárások is megváltoztak. A web hozzászoktatta a felhasználót ahhoz, hogy az információkat olyan hálózatként értelmezze, amelyben minden mindennel összefügg. Ez azt is jelenti, hogy a könyvtári katalógusok felé is hasonló szemléletmóddal fordulnak. Az adataink integrálása a webbel nagyobb feladat, mint a weben történő megjelenítése. Tehát nemcsak hogy kevesebb embernek kell jóval több anyagot feldolgoznia, de még a feldolgozás mélységén is változtatni kellene. Ahhoz, hogy bírjuk az iramot, meg kell osztani a munkát. A munkamegosztásból következik a rekord felbomlása, mivel nem feltétlenül teljes leírásokat veszünk át, csak részleteket. Illetve felértékelődik annak a szerepe is, hogy az egyes adatelemekhez tudjunk majd provenienciainformációkat kapcsolni.

Ez abba az irányba mutat, hogy alapvető fontosságú lesz az, hogy pontosan lássuk, melyik adatelem miről is szól. Ebben lesz segítségünkre az *IFLA könyvtári referenciamodell* (IFLA LRM).<sup>15</sup> Az IFLA LRM egy magas szintű referenciamodell, amelynek célja, hogy azonosítsa a bibliográfiai univerzum azon entitásait, amelyek használói érdeklődésre tartanak számot.

<sup>12</sup> Françoise Leresche *Rethinking bibliographic control in the light of IFLA LRM entities: the ongoing process at the National library of France* című előadása alapján. Elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 12. <https://youtu.be/sPrefNT8wG0?t=4452> (2021.10.20.)

<sup>13</sup> A hazai szakirodalomban jelenleg *Egyetemes Bibliográfiai Számbavétel*ként szokás utalni rá.

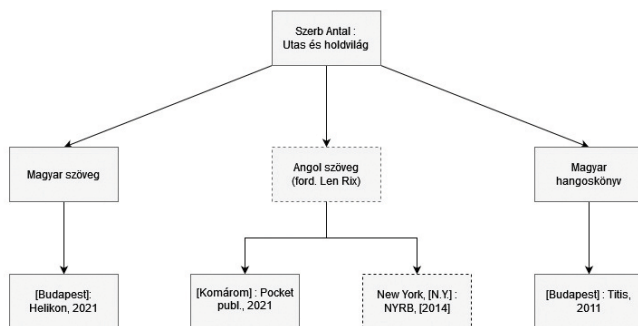
<sup>14</sup> *International congress on national bibliographies = Nemzetközi kongresszus a nemzeti bibliográfiákról*. [Tömörített fordítás a konferencia dokumentumaiból]. Budapest, [s.n.], 1984. 153, 38 fol. 6–7. p. (gépirat)

<sup>15</sup> Riva, Pat – Le Boeuf, Patrick – Žumer, Maja: *IFLA Library Reference Model: A Conceptual Model for Bibliographic Information*. IFLA, 2017. 101 p. [https://repository.ifla.org/bitstream/123456789/40/1/ifla-lrm-august-2017\\_rev201712.pdf](https://repository.ifla.org/bitstream/123456789/40/1/ifla-lrm-august-2017_rev201712.pdf) Magyar fordítása: IFLA könyvtári referenciamodell: A bibliográfiai információk elméleti modellje. Ford. Gazdag Tiborné, IFLA, 2018. 99 p. [https://repository.ifla.org/bitstream/123456789/46/1/ifla\\_lrm\\_2017\\_hun\\_v3.pdf](https://repository.ifla.org/bitstream/123456789/46/1/ifla_lrm_2017_hun_v3.pdf) (2021.10.20.)



- a mű reprezentatív kifejezési formáját megtestesítő első megjelenési forma megjelenési vagy terjesztési helyét;
- a reprezentatív kifejezési forma nyelvét;
- a készítőik nemzetiségét (amennyiben a nyelvi kitétel nem alkalmazható, pl. instrumentális zene vagy állókép esetében, vagy esetleg a nyelvi kitétel kiegészítve).

Ennek a gyakorlatnak a széles körű adaptálása a nemzeti illetőségű művek mellett az azokhoz kapcsolódó eredeti nyelvű, vagy az adott nemzethez tartozó ágens által létrehozott kifejezési formákat sorolná a nemzeti bibliográfiai ügynökségek feltárási kötelezettségi körébe. A következő ábrán látható, hogy a BnF gyakorlata hogyan nézne ki magyar környezetre vetítve. Az ábrán folytonos vonalú téglalapban a magyar nemzeti bibliográfiai ügynökség feladatkörébe tartozó entitások láthatók, míg szaggatott vonalú téglalapban az azon kívül esők vannak.



*A BnF gyakorlata magyar környezetre értelmezve*

Nem mindig egyszerű azonban megállapítani, mi számít nemzeti illetőségű tartalomnak. A személyek nemzetiségi hovatartozásának problémáján kívül néhány kevésbé szembeötlő nehézséggel is szembesülni kellett, főleg az audiovizuális anyagokat illetően. A profi filmek és zenei anyagok jellemzően nemzetközi stáb együttműködésének eredményei, így nemzetiségi besorolásuk problémás lehet, míg az amatőr alkotó kimondottan webes terjesztésű alkotásai esetén a kiadási hely teljesen értelmét veszíti, a nemzetiségi hovatartozást pedig az internet anonimitása miatt nem feltétlenül lehet megállapítani. Világos tehát, hogy a kultúra terjesztési mechanizmusainak átalakulását nem lesz egyszerű feladat lekövetni, ugyanakkor az entitáskatalogizálás egy nagy lépés ebbe az irányba.

**Ilácsa Szabina**

## Entitásalapú szemlélet és új hordozóformátum a katalogizálásban<sup>17</sup>

Az entitásalapú katalogizálást a gyakorlatban is megvalósíthatóvá tévő RDA-szabályzat első változata 2010-től kezdve körülbelül hét éven keresztül funkcionált, majd a 3R-projekt keretében (RDA Restructure and Redesign) dolgozták át az *IFLA könyvtári referenciamodell*ben lefektetett elméleti kereteknek megfelelően. (E változtatások mértékét jól jelzi, hogy a *Program for Cooperative Cataloging* (PCC), vagyis a Kongresszusi Könyvtár közös katalogizálási programja az átalakított RDA implementációját csak 2022 júliusára tervezi.) A fejlesztés kezdete óta a szakemberek egyetértenek abban a tekintetben, hogy a MARC helyett valamilyen új hordozóformátumot kell igénybe venni az RDA-ban meghatározott metaadatelemek megfelelő leírása, illetve az adatcsere folyamatának modernizálása céljából. A megoldás, a BIBFRAME bibliográfiai keretrendszer (lényegében szemantikus elemkészlet, azaz relációsztár) ötlete 2011-ben merült fel, és folyamatos fejlődésen ment keresztül a szakmai közösségtől érkező visszajelzések

<sup>17</sup> A cikk elkészítésekor az alábbi előadásokat vettem alapul:

Sparling, Abigail és Bigelow, Ian: *Control or chaos: embracing change and harnessing innovation in an ecosystem of shared bibliographic data*. Elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 8. <https://youtu.be/ybUDrILt0kI?t=7715> (2021.10.20.)

Bianchini, Carlo – Sardo, Lucia: *Wikidata: a new perspective on the Universal Bibliographic Control*. Elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 10. <https://youtu.be/Yo6Vi72E1T4?t=5792> (2021.10.20.)

Chapman, John: *Building a shared entity management infrastructure: moving from promise to production*. Elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 9. [https://youtu.be/Z\\_MDPpzVRl4?t=11152](https://youtu.be/Z_MDPpzVRl4?t=11152) (2021.10.20.)

Riva, Pat: *The multilingual challenge in bibliographic description and access*. Elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 10. <https://youtu.be/Yo6Vi72E1T4?t=12826> (2021.10.20.)

alaján. Jelenleg a 2.0.1. verziója használható, és több nemzeti könyvtár folytat vele kiterjedt vagy elszigetelt kísérleteket. A BIBFRAME fejlesztésének célja egy olyan közösségi szabvány megalkotása volt, amelynek segítségével az RDA gyakorlati alkalmazásai végre túlléphetnek a MARC korlátain.

Noha az RDA kapcsolt adat környezetben történő alkalmazási lehetőségeinek vizsgálata 2017-től kezdve jelentősen felgyorsult, az átalakulási folyamat várhatóan évekig fog tartani, ezért a munkafolyamatokat és a technikai ökoszisztémát hibrid rendszerek megvalósítására kell felkészíteni. A Kongresszusi Könyvtár ennek megfelelően kidolgozta a MARC-ról BIBFRAME-re, illetve BIBFRAME-ről MARC-ra történő konverzió specifikációját is – ezeknek a konverziós eljárásoknak tekintettel kell lenniük az adatelemek rögzítésének időben változó gyakorlataira is.

Az RDA és a BIBFRAME azonban eltér egymástól az alkalmazott alapvető entitások tekintetében.<sup>18</sup> Előbbi, megörökölve az FRBR<sup>19</sup>-ben és az LRM-ben írtakat, négy központi entitással dolgozik, míg a BIBFRAME csak hármat definiál: a *művet* (Work), annak valamely *előfordulását* (Instance), illetve a *példányt* (Item); a kifejezési forma metaadatelemei ebben a felosztásban a mű jellemzőiként rögzítendőek. Problémát okozhat ugyanakkor, hogy a BIBFRAME, bár definiál *hasExpression*-, illetve *expressionOf*-tulajdonságot, e relációk végeredményben két mű közötti kapcsolatot írnak le. E visszás helyzet megoldására alkották meg az olaszországi @Cult szoftverfejlesztő cég munkatársai az Opus (korábban *SuperWork*), illetve ezzel párhuzamosan a Kongresszusi Könyvtár szakemberei a *Hub* nevű, magas szintű entitást, amely segítségével akár MARC-ból történő konverzió során is kifejezhetővé válnak a művek és kifejezési formák kapcsolatai.

A kapcsolt adatokra épülő új forrásleírási ökoszisztémát, illetve az RDA és a BIBFRAME együttműködését a Sinopia szerkesztőfelületén tesztelik. A regisztráció után szabadon hozzáférhető felület lehetővé teszi egyedi, intézményre szabott alkalmazási profilok, azaz a specifikus igényekre szabott metaadat-leíró űrlapok létrehozását, amelyek beviteli mezői a legtöbb esetben megfeleltethetők a BIBFRAME-szótár relációinak, bizonyos forrástípusok esetében kiegészítő elemkészleteket (pl. a PMO-t, Performed Music Ontology) is fel kell használni.

<sup>18</sup> Li, Xiaoli: *MARC to BIBFRAME (Linked Data)*. Elhangzott: NCTPG 80th Annual Meeting, San Francisco Public Library, 2017. április 28. Az előadás diasora: <https://www.slideshare.net/nctpg1/xiaoli-li-marc-to-bibframe-linked-data> (2021.10.20.)

<sup>19</sup> Functional Requirements for Bibliographic Records – a bibliográfiai univerzum entitásanalízisen alapuló vizsgálata, amely először írta le a könyvtártudomány területén az entitásokat. Magyar nyelven, Berke Barnabásné fordításában elektronikusan elérhető: <https://repository.ifla.org/bitstream/123456789/822/1/ifla-functional-requirements-for-bibliographic-records-frbr-hu.pdf> (2021.10.20.)



Az alkalmazási profilban meg kell határozni a metaadatelem ismételtetésének, illetve kötelezőségének paramétereit, továbbá rögzíteni kell az elem URI<sup>20</sup>-ját és címkéjét. A profil ezen kívül tartalmaz még egy, a metaadatérték szabványos rögzítéséhez használandó RDA-fejezetpontra mutató hivatkozást is. A Sinopia felületén számos intézmény – például az Alberta Egyetem és a Kongresszusi Könyvtár Hálózatfejlesztési és MARC-szabványosítási Osztály (NDMSO) – alkalmazásprofiljai hozzáférhetők, megtekinthetők, kipróbálhatók.

A Sinopia mellett az Alberta Egyetem könyvtárában zajló kísérleti munka fő komponensei a Sirsi integrált könyvtári rendszer, amely a hagyományos könyvtári munkafolyamatok támogatását végzi (alapvetően MARC-ra épül), a Share-VDE discovery-felület, illetve az *@Cult* által fejlesztett *Sapientia* és *J. Cricket* tudástár, amely a forrásleírást kiegészítő authority kontrollban, azaz entitásmenedzsmentben nyújt segítséget. Az egyes rendszerkomponensek között állandó, kétirányú adatkapcsolat áll fenn, amely oda-vissza konverziós fázist is tartalmaz.

Az előadás időpontjában – s ez azóta sem változott – még nem állt rendelkezésre hivatalos megfeleltetés az RDA elemkészlete, illetve a BIBFRAME elemei között. Az RDA kidolgozói ugyanakkor úgy gondolják, hogy a szabályzattal együtt megalkotott elemkészleteknek előnyt kell élvezniük a bibliográfiai leírások gráfként történő formalizálásakor: egy RDA-kompatibilis metaadatleírás-halmaz elkészítéséhez az RDA szótárát – és csak azt – lehet használni, így a BIBFRAME relációival, sőt az RDA úgynevezett korlátozás nélküli (unconstrained) metaadatelemeivel készült halmazok esetében is megfelelőségi problémák merülnek fel. A jövőben az egységes vagy legalábbis ahhoz közelítő közgyűjteményi adattárolás érdekében feltétlenül szükséges ennek a problémának a megoldása, és szintén fontos az egyes konverziós eszközök finomhangolása, valamint ezzel párhuzamosan a BIBFRAME-gráfok metaadatelem-készletének szabványosságá tétele.

**Hubay Miklós**

<sup>20</sup> Universal Resource Identifier, egységes forrásazonosító.



# Azonosítók modern könyvtári környezetben<sup>21</sup>

Azonosító alatt egy olyan valamit értünk, amely révén egyedi módon meghatározható valakinek a kiléte vagy valaminek a mibenléte, legyen szó személyről, fogalomról vagy tulajdonképpen bármiről. Az ujjlenyomatunk, a DNS-ünk és egyéb biológiai jellemzőink egyedisége azonosít minket egyénekként, és az ilyen jellemzőket felhasználó úgynevezett biometrikus azonosítók (pl. ujjlenyomat- és íriszazonosítás, arc-, illetve hangfelismerés) a közelmúlt technológiai fejlesztései révén már a mindennapokban is alkalmassá váltak személyazonosításra.

Az azonosítás iránti igény egyidős az emberi civilizációval. Alapvetően az emberek neve is azonosítónak tekinthető, bár nem túl hatékonyak, hiszen több személynek lehet azonos neve. Ahhoz hogy kétséget kizáróan azonosítani lehessen egy bizonyos dolgot (vagyis entitást), ahhoz olyan egyedi karaktersor, kulcs vagy bármilyen erre alkalmas dolog, tárgy (pl. egy oklevelet hitelesítő pecsét, beléptetéshez használt kártya, felhasználónév és jelszó párosítása) szükséges, amely kizárólag ahhoz az egy bizonyos entitáshoz van hozzárendelve. Személyek példájánál maradván a nevünk megegyezhet másokéval, ám a személyi igazolványunk vagy más okmányunk azonosítószáma már kizárólag minket azonosít. Persze továbbra is megvan az esélye annak, hogy ugyanezzel a számsorral kiadnak valahol a világon valaki másnak is egy ezzel megegyező azonosítójelet, ám (hacsak nem

<sup>21</sup> A cikk elkészítésekor az alábbi előadásokat vettem alapul:

Boulet, Vincent: *Towards an identifiers' policy: the use case of the Bibliothèque nationale de France*. Elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 9. [https://youtu.be/Z\\_MDPpzVRl4?t=382](https://youtu.be/Z_MDPpzVRl4?t=382) (2021.10.20.)

Putnam, Nathan B.: *VIAF and the Linked Data Ecosystem*. Elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 10. <https://youtu.be/Yo6Vi72E1T4?t=7689> (2021.10.20.)

MacEwan, Andrew: *The International Standard Name Identifier: extending identity management across the global metadata supply chain*. Elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 12. <https://youtu.be/sPrefNT8wG0?t=359> (2021.10.20.)

tévedésről van szó) az adott rendszer kontextusán belül értelmezve – tehát például egy adott ország okmányait figyelembe véve – teljesül az egyedi azonosítás követelménye. Ebből adódóan értelmes szavakkal (elnevezésekkel) és látszólag összefüggéstelen karaktersorozatokkal egyaránt lehet adott dolgokat azonosítani. Utóbbiak hatékonysága könnyen belátható: számok, betűk és akár egyéb karakterek, írásjelek kombinációjával gyakorlatilag végtelen számú azonosítójel állítható elő, amely egyedi és független a beszélt nyelvtől, tehát nemzetközi szinten és gépi környezetben egyaránt alkalmazható. Utóbbi különösen fontos terepe az azonosítóknak, hiszen a modern értelemben vett azonosítás nagyrészt már elképzelhetetlen a háttérben futó informatikai megoldások nélkül, legyen szó egy könyvtári integrált rendszer katalógusát működtető adatbázisról, egy elektronikus beléptetőrendszerrel, online ügyintézésről vagy akár csak egy jelszót igénylő számítógép-használatról.

A szerzők és a kiadványok azonosítása nem új keletű, a könyvtártörténet egészen az ókori ninivei könyvtár katalógusáig vezet vissza ennek gyakorlatát,<sup>22</sup> ám a kiadványazonosítás szükségességét a könyvkereskedelem igénye alapozta meg. Az egyesült államokbeli *szabványos könyvazonosító szám* (Standard Book Number – SBN), majd ennek módosításával a *nemzetközi szabványos könyvazonosító szám* (International Standard Book Number – ISBN)<sup>23</sup> bevezetésével vált lehetővé adott könyvek azonos kivitelű (kötéstáblás, fűzött) és azonos árú példányainak azonosítása.<sup>24</sup>

Egy modern könyvtár már jóval több, mint nyomtatott dokumentumok gyűjtőhelye. Elektronikus gyűjteményei, tájékoztató információforrásai és egyéb szolgáltatásai révén az online világ szerves részét képezi. A bibliográfiai feldolgozás területén is történt egy hasonló szintlépés, ahogy a könyvtári bibliográfiai feltáró eszközök először elektronikus formátumra váltottak, majd az internetes összeköttetésnek köszönhetően egymással is interakcióba kerültek. A kapcsolható-környezetben jelenleg használt azonosítók olyan egyedi karaktersort jelölnek, amely egy adott dolog egyértelmű meghivatkozására szolgálnak, ezért nem meg-

<sup>22</sup> Holl András – Bilicsi Erika: *ORCID – egy újabb szerzői azonosító tudományos közleményekhez.* = KönyvtáriFigyelő, 27. (63.) évf. 2017. 3. sz. 346–350. p. [http://epa.oszk.hu/00100/00143/00348/pdf/EPA00143\\_konyvtari\\_figyelo\\_2017\\_03\\_346-350.pdf](http://epa.oszk.hu/00100/00143/00348/pdf/EPA00143_konyvtari_figyelo_2017_03_346-350.pdf) (2021.10.20.)

<sup>23</sup> *ISBN útmutató*, Összeáll. Csirmazné Rezi Éva és Szabó Erika Zita. Budapest, Országos Széchényi Könyvtár, 5. átd. kiadás, PDF-szöveg, 2012. 44 p. <https://www.isbn-international.org/sites/default/files/Hungarian%20Manual.pdf> (2021.10.20.)

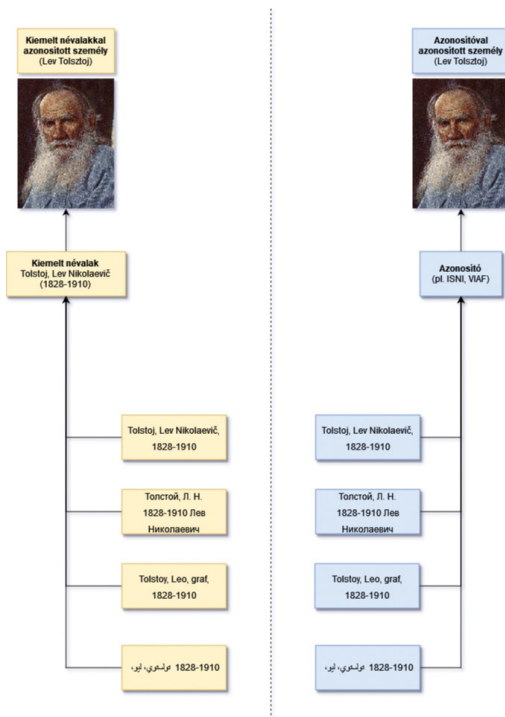
<sup>24</sup> Ungváry Rudolf – Vajda Erik: *Nemzetközi (és más) azonosító számok és kódok*. In: Uők.: *Könyvtári információkeresés*. Budapest, Typotex, 2002. 119. p. A MEK-en található elektronikus verzió tördelése nem egyezik meg a nyomtatott könyvével. Itt a PDF-dokumentum 80–81. oldalán található a hivatkozott rész. <https://mek.oszk.hu/05000/05030/> (2021.10.20.)

lepő hogy az azonosítók használatának előnyei számítógépesített környezetben mutatkoznak meg igazán.

Az azonosítók használatának egyik legfontosabb hozadéka, hogy a különböző alakban írt fogalmak (gondoljunk itt például a szinonimákra vagy a nyelvi, írásképi különbségekre) általuk kezelhetővé válnak egy nagyobb (akár nemzetközi) informatikai rendszerben. Ha a bibliográfiai adatelemek beírása helyett az adott adatelemre hivatkozó szabványos azonosítót használunk a feldolgozás során, akkor minden erre alkalmas rendszer azt az adott dolgot (fogalmat, nevet) fogja érteni alatta, amit mi feldolgoztunk, míg kézi bevitelnél jóval nagyobb a tévedés (pl. névazonosságból eredő) lehetősége, nem is beszélve a nyelvi korlátokról és az adatcsere nehézségeiről.

Egy személy neve sokféle formában létezhet, különösen, ha figyelembe vesszünk más nyelveket és írásrendszereket. A könyvtári adatrendszerező munka egyik fő feladata, hogy rendet teremtsen az efféle bábeli zűrzavarban. Az erre a célra szolgáló kitüntetett névalak a név egységesített formája, a többi közül kiemelt szabványosított névváltozat, amely egyezményesen jelöli az adott entitást. *Lev Tolsztojt* példaként említve ez az egységesített kitüntetett névalak az Országos Széchényi Könyvtárban és egyúttal a Magyar Nemzeti Bibliográfiában a *Tolsztoj, Lev Nikolaevič (1828–1910)*. A magyar nyelvnek idegen forma a könyvtártudományban alkalmazott cirill–latin karakterek transliterációjára vonatkozó szabvány (MSZ ISO 9:1997) alkalmazásának a következménye. A szabványosított névalak és a születési-halálozási évszámok együttes feltüntetése felfogható egy természetes nyelven megfogalmazott azonosítónak, amely egy adott rendszeren belül (a szláv nevek kitüntetett alakjai esetében az ISO 9 szabványt alkalmazó könyvtárak köre) egyértelműen a klasszikus orosz íróra utal. Az ilyen egységesített besorolási adatnak kijelölt személynév formátuma szintén szabványosítás következménye (Magyarországon ezt az MSZ 3440/2-79 határozza meg), azonban a gyakorlatban országonként más és más szabvány szabályozza a kiemelt névalakot (lásd a következő ábrán feltüntetett különböző kitüntetett névalakokat a világ tájairól), így a webes információcsere gépesített korszakában célszerű egy nyelvfüggetlen azonosítót hozzárendelni az adott személyhez. Az ilyen azonosítók képesek nemzetközi szinten az eltérő kitüntetett névalakok meghivatkozására, így például Lev Tolsztoj esetében egy korszerű rendszer akkor is tudja, hogy az orosz író a hivatkozott személy, ha épp anyanyelvén, cirill betűkkel vagy a számtalan latin betűs transliteráció egyikével tüntetik fel hivatkozandó kitüntetett névalakként.

Az azonosítókat sokféleképpen csoportosíthatjuk. Tágabb értelemben az azonosító lehet birtoklásalapú (pl. igazolvány, olvasójegy, pecsét, kulcs, rádiófrekvenciás azonosítás – RFID), viselkedésalapú (pl. aláírás), biometrikus (pl. ujjlenyomat, arcfelismerés, íriszazonosítás, DNS) és tudásalapú (pl. jelszó, PIN-kód, informatikai azonosítók). Az azonosítók több szintűek, több csatornásak is



*Azonosítás és kiemelt/kitüntetett névalak használata  
(Utóbbi rendszerként elterjedt, míg az azonosító globálisan alkalmazható)*

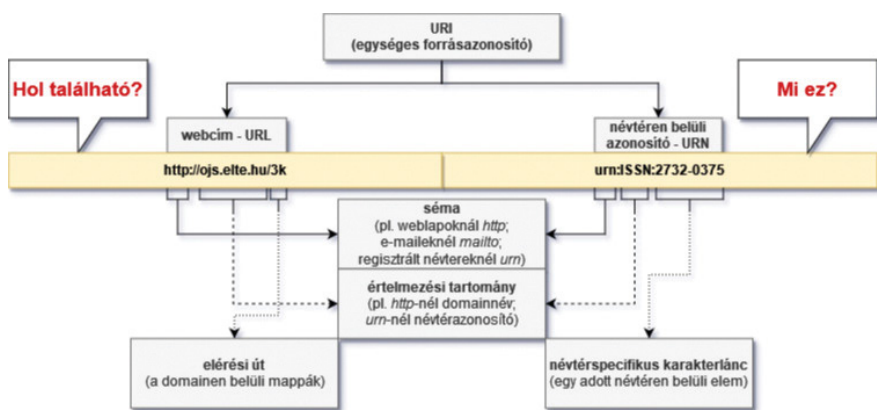
lehetnek, különösen, ha valamilyen biztonsági céllal alkalmazzák. Online banki ügyintézésnél például a számlaszámunk (ez már egy azonosító) fiókjához eleve felhasználónév és jelszó párosával, valamint SMS-ben küldött jóváhagyó kóddal vagy biometrikus azonosítással férhetünk hozzá.<sup>25</sup>

Webes és számítógépes hálózati környezetben alapvető azonosítónak tekinthető az URI, azaz az egységes forrásazonosító (Uniform Resource Identifier),<sup>26</sup> amely egyedi karaktersorozatként egy olyan webcím (URL) vagy névtérséma (URN) felettes fogalma, amely egyedi módon azonosít egy webes forrást pontosan meghatározott séma szerint. Az URI az IANA (Internet Assigned Numbers Authority) által központilag regisztrált sémákat tartalmazhat. Például a webol-

<sup>25</sup> Folláth János – Huszti Andrea – Pethő Attila: *Az azonosítók fajtái*. In: Uő.k.: *Informatikai biztonság és kriptográfia*. Debrecen, Kempelen Farkas Digitális Tankönyvtár, 2010. 237 p. <https://gyires.inf.unideb.hu/KMITT/b09/ch06s02.html> (2021.10.20.)

<sup>26</sup> *Uniform Resource Identifier (URI): Generic Syntax*. The Internet Society, 2005. <https://datatracker.ietf.org/doc/html/rfc3986> (2021.10.20.)

dalaknál ilyen séma a HTTP-protokoll, de maga az URN is ennek tekinthető.<sup>27</sup> Röviden összefoglalva: az URL mindig megadja, hogy mit és hol találunk meg a hálózaton, az URN pedig megnevezi a hálózaton lévő, de a valóságban nem feltétlenül létező, elvont fogalmat is. Webcímek esetében fontos megemlíteni, hogy sokszor sajnos nem tartósak, így gyakorta belefuthatunk már inaktív URL-ekbe. A hálózati tartalmak fenntartóinak ezért törekedniük kell a könnyen felismerhető, logikus webcímek kialakítására és azok tartós fenntartására, csak így használhatók hosszú távon azonosítóként.<sup>28</sup>



*Példák a Könyv, Könyvtár, Könyvtáros folyóiratot azonosító URI-akra  
(Az URL az ELTE OJS-tartományon belüli 3K-kezdőoldali webcímével, míg az URN  
a folyóiratot az online ISSN-számával azonosítja)*

A bibliográfiai feldolgozó munka kontextusában az azonosító egyedi karakter-sort jelöl, így a továbbiakban csak ilyen értelmezésben foglalkozunk a témakörrel. Még a bibliográfiai területre leszűkítve is igen népes az azonosítók családja. Az azonosítók csoportosításához sorvezetőként a BnF azonosítási irányelve szolgált, amelyet *Vincent Boulet* ismertetett a BC2021 konferencián.

A bibliográfiai munkában is használt azonosítók első csoportjának a nemzetközi szabványok (ISO) által szabályozott azonosítókat tekinthetjük. Egyik

<sup>27</sup> Sütthető Péter: *Digitális tartalmak azonosítása, bitelesítése, biteles tartalomszolgáltatás*. = Könyvtári Figyelő, 29. (65.) évf. 2019. 4. sz. 535. p. [https://epa.oszk.hu/00100/00143/00358/pdf/EPA00143\\_konyvtari\\_figyelo\\_2019\\_04\\_527-544.pdf](https://epa.oszk.hu/00100/00143/00358/pdf/EPA00143_konyvtari_figyelo_2019_04_527-544.pdf) (2021.10.20.)

<sup>28</sup> McMurry, Julia A., et al.: *Identifiers for the 21st Century: How to design, provision, and reuse persistent identifiers to maximize utility and impact of life science data*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press, March 20, 2017 Publicly Available Content Database. DOI: <http://dx.doi.org/10.1101/117812> (2021.10.20.)

ismertetőjelük a nemzetközileg használt rövidítésükben általában feltűnő IS (International Standard – Nemzetközi szabvány) betűpár, ám előfordul kivétel is, például a nemzetközi szabványként elfogadott, digitális objektumok azonosítására szolgáló DOI-azonosítók (ISO 26324) esetében.<sup>29</sup> A teljesség igénye nélkül a nemzetközi szabványos azonosítók sorába tartozik a könyvek azonosítására szolgáló ISBN, az időszaki kiadványokéra szolgáló ISSN,<sup>30</sup> a nevekére az ISNI,<sup>31</sup> az audiovizuális művekére az ISAN,<sup>32</sup> maguknak a könyvtáraknak az azonosítására az ISIL.<sup>33</sup> Más területek is rendelkeznek hasonló nemzetközi azonosítókkal. A nemzetközi szabványos azonosítók használatának legnagyobb előnye a szabványosításból eredő normatív jelleg, azaz hogy pontosan meghatározott módon és célra kell használni ezeket az azonosítókat, amelyek fenntartásáról és regisztrálásáról az erre a célra létrehozott szervezetek gondoskodnak. Ez a rendszerszintű háttér garantálja, hogy globális értelemben vett egyedi azonosító társuljon az azonosított entitáshoz.

A második csoporthoz a „kváziszabvány” azonosítók tartoznak. Ezeket nagyívű, országos vagy nemzetközi projekteken és egyéb vállalatokban hozzák létre, és épp olyan széleskörűen felhasználhatók különböző célokra, mint a nemzetközi szabványos azonosítók. Hasonló infrastruktúra áll mögöttük, szintén nemzetközi felhasználói réteget és közönséget céloznak meg, ám az ezen vállalatokban történő részvétel lazább feltételekhez kötött, kevésbé szabályozott, mint a nemzetközi szabványosítás folyamata. Az sem precedens nélküli, hogy egy ilyen jellegű vállalásból származó azonosító válik nemzetközi szabvánnyá, ahogy az történt a DOI esetében is. A tartalomszolgáltató iparágak által használt azonosítók is ide sorolhatók (pl. a mozifilmek és televíziós műsorszámok azonosítására szolgáló EIDR<sup>34</sup>), de a jelentős könyvtári projekteken használt azonosítók is, legyen szó a nemzetközi authority állomány VIAF-azonosítóiról<sup>35</sup> vagy a Kongresszusi Könyvtár világszerte alkalmazott authority szolgáltatásának (Library of Congress/NACO<sup>36</sup>) azonosítóiról. Napjaink egyre inkább előtérbe kerülő, feltö-

<sup>29</sup> Lásd a DOI-szervezet honlapján elérhető DOI Handbookot: <https://www.doi.org/hb.html> (2021.10.20.)

<sup>30</sup> Az Országos Széchényi Könyvtár honlapja ISSN-aloldala: <https://www.oszk.hu/issn> (2021.10.20.)

<sup>31</sup> Az ISNI honlapjának kezdőoldala: <https://isni.org/> (2021.10.18.)

<sup>32</sup> Az International Standard Audiovisual Number honlapjának kezdőoldala: <https://www.isan.org/> (2021.10.20.)

<sup>33</sup> Az Országos Széchényi Könyvtár honlapja. *ISIL-kódok*: <https://www.oszk.hu/orszagos-konyvtari-szabvanyositas/isil-kodok?mobile=off> (2021.10.20.)

<sup>34</sup> Az EIDR honlapja. Kezdőoldal: <https://www.eidr.org/> (2021.10.20.)

<sup>35</sup> Az VIAF honlapja. Kezdőoldal: <https://VIAF.org/> (2021.10.20.)

rekvő nemzetközi projektje a Wikidata,<sup>37</sup> egy szabadon hozzáférhető nyílt tudásbázis, amely (kezdetben) a Wikipédia-szócikkekben leírt ismereteket dolgozta fel és rendezte emberi és gépi olvasásra egyaránt alkalmas strukturált adattá, amely így hozzáköthető más szabványos felépítésű nyílt kapcsoltadat-tárakhoz. A Wikidata mára jócskán túlnötte a Wikipedia és társszolgáltatásainak szócikkszámát (ez jelenleg több mint 57 millió<sup>38</sup>), jelen sorok írásának pillanatában több mint 95 millió tartalmi (nem adminisztratív célú) cikket jelez a statisztika.<sup>39</sup> Az ezen cikkek tárgyát képező dolgok mind rendelkeznek egyedi azonosítóval (ez az úgynevezett Q-azonosító vagy Q-ID), így a Wikidata jó úton halad afelé, hogy a különféle dolgok szemantikus weben történő azonosításában emberi és gépi értelmezésre egyaránt alkalmas, egyetemesen használható szolgáltatássá váljon.<sup>40</sup>

Az azonosítók harmadik nagy csoportját a belső munkafolyamatokban használatos helyi azonosítók képezik. Ide tartozik minden olyan azonosító, amelyet kizárólag valamilyen szervezeten (vagy együttműködő szervezetek körén) belül használnak. Példának említhetők az interált rendszerekben és egyéb adatbázisokban alkalmazott azonosítók, illetve kézenfekvő példa egy adott dokumentum példányazonosítója.

BnF-funkció	Nemzetközi ISO-szabvány azonosítók	Nemzetközi közönséggel rendelkező azonosítók	Helyi azonosítók
Hozzájárulás vagy regisztráció	 <b>isni</b>	 <b>ARK Alliance</b>	<b>FRBNF</b>
Alkalmazás, szolgáltatás épül rá	 <b>isbn</b>	<b>EAN</b>	
Megfeleltetéseket készít hozzá		<b>LC SH, MESH, GND, datos.bne.es, VIAF, NOMISNA, Geonames, Agrovoc, Wikidata</b>	
Munkafolyamatokba integrált azonosítók	 <b>ISAN</b> <small>INTERNATIONAL STANDARD AUDIOVISUAL NUMBERING</small>	<b>EIDR</b>	

*Azonosítóhasználati irányelv a Francia Nemzeti Könyvtárban (BnF)  
(Vincent Boulet prezentációja alapján)*

<sup>36</sup> A Library of Congress honlapja. *NACO – Name Authority Cooperative Program*: <https://www.loc.gov/aba/pcc/naco/> (2021.10.20.)

<sup>37</sup> A Wikidata kezdőoldala: [https://www.wikidata.org/wiki/Wikidata:Main\\_Page](https://www.wikidata.org/wiki/Wikidata:Main_Page) (2021.10.20.)

<sup>38</sup> WikiMedia List of Wikipedias: [https://meta.wikimedia.org/wiki/List\\_of\\_Wikipedias](https://meta.wikimedia.org/wiki/List_of_Wikipedias) (2021.10.20.)

<sup>39</sup> Wikidata Statistics: <https://www.wikidata.org/wiki/Special:Statistics> (2021.10.20.)

<sup>40</sup> Van Veen, Theo: *Wikidata: From „an” identifier to „tbe” identifier*. = Information Technology and Libraries (Online), 38. vol. 2019. 2. no. 72–81. p. In: Library & Information Science Collection; Publicly Available Content Database. DOI: <https://doi.org/10.6017/ital.v38i2.10886> (2021.10.20.)



Ahogy láthattuk, már magukban a könyvtári munkafolyamatokban is igen széles az elérhető azonosítók köre. A hatékony alkalmazás érdekében alaposan meg kell gondolni, hogy adott munkafolyamathoz melyik azonosítót használjuk. A francia gyakorlat szerint alapvetően négyféle funkció szerint használhat egy könyvtár a fenti csoportosítás bármelyik szintjébe tartozó azonosítókat:

- A szervezet maga is részt vehet az azonosítók regisztrációjában és bővítésében, nemzetközi fenntartású azonosítók esetén követve a nemzetközi szabályozásokat és irányelveket. Ennek keretében nemzetközi együttműködésben osztható ki országon belül azonosító. Erre példa az Országos Széchényi Könyvtárban működő Magyar Nemzeti ISSN Központ<sup>41</sup> és Magyar ISBN és ISMN Iroda,<sup>42</sup> amelyek a hazai kiadású időszaki kiadványok és könyvek, illetve kották azonosítóinak kiosztásáért felelnek. A francia példában szereplő digitális művek azonosítására is szolgáló ARK Alliance-azonosító is jellemzően ide sorolható, hiszen több mint 838 szervezet együttműködésében (beleértve számos nemzeti könyvtárat) épül.<sup>43</sup> A szervezet belső használatra szánt és saját maga által előállított azonosítóit értelemszerűen e funkció szerint használják.
- A szervezet alkalmazhat már létező azonosítókat, és ezekre építhet szolgáltatásokat. Talán ez tekinthető a leggyakoribb azonosítófelhasználási formának, hiszen gyakorlatilag arról van szó, hogy egy adott dologhoz egy már akár nemzetközi szinten is hozzárendelt azonosítót társítunk, például könyvekhez az ISBN-számot, időszaki kiadványokhoz az ISSN-számot. A mások által is használt azonosítók felhasználásával az ilyen jellegű azonosítóval rendelkező entitások révén a szervezetünk részévé válik egy nagyobb rendszernek. Minél több általánosan használt azonosítót alkalmazunk, annál inkább integrálódik könyvtárunk ebbe a szisztémába. Gondoljunk csak a könyvtárközi kölcsönzést biztosító rendszerre, amely működése lehetetlen lenne szabványos vagy legalábbis az érdekelt felek által közösen használt azonosítók nélkül.
- A harmadik funkció a különböző azonosítórendszerek közti megfeleltetések készítését takarja. Ez szintén a fenti rendszerek (és könyvtárak) közötti interoperabilitást hivatott elősegíteni. Ez a munka jelentheti egy könyvtár sa-

<sup>41</sup> Az Országos Széchényi Könyvtár honlapja. *Magyar ISSN Nemzeti Központ (ISSN Hungary)*. <https://www.oszk.hu/book/magyar-issn-nemzeti-koezpont-issn-hungary> (2021.10.20.)

<sup>42</sup> Az Országos Széchényi Könyvtár honlapja. *Magyar ISBN és ISMN Iroda*. [https://www.oszk.hu/isbn\\_ismn\\_iroda](https://www.oszk.hu/isbn_ismn_iroda) (2021.10.20.)

<sup>43</sup> Ark Alliance honlapja. Community: <https://arks.org/community/> (2021.10.20.)



ját belső azonosítóihoz társított entitások megfeleltetését a nemzetközi azonosítók által identifikált ugyanazon entitásokhoz (pl. az integrált rendszer belső azonosítója által reprezentált könyvhöz nem csupán az ISBN-számot lehet hozzárendelni, hanem meg lehet feleltetni nemzetközi azonosítóknak, sőt más könyvtárak belső azonosítóinak is). A VIAF-szolgáltatás egyik fő terépe ezen azonosítók megfeleltetése kapcsoltadat-környezetben (lásd lentebb a VIAF-ról szóló bekezdést). A nemzeti könyvtárak és a nemzetközi könyvtárszakmai szervezetek hasznos munkája nemcsak az egyes entitásokhoz rendelt különböző azonosítók megfeleltetését takarja, hanem a különböző szótárak, tárgyszórendszerek, ontológiák kifejezéseinek egymáshoz történő megfeleltetését is.

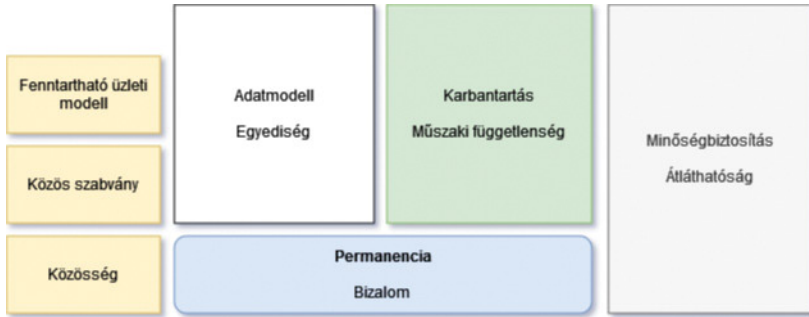
– Végül a szervezetek a munkafolyamataikba is integrálhatnak azonosítókat.

Az azonosítók világában a permanencia (azaz a tartósság, állandóság) kulcsfontosságú követelmény. Ennek fényében az azonosítóknak négy alapkövetelménynek kell megfelelniük:

- nagy létszámú és stabil közösség tartsa fent: ez szabványosítási vagy stratégiai szintű folyamat (pl. jelentős digitalizálási projekt) részeként létrejött azonosítókat jelent (lásd nemzetközi szabványos azonosítók, pl. ISBN, ISSN, ISNI, de a szabványos ipari azonosítók is idesorolhatók);
- egyértelmű alkalmazási irányelvekkel kell rendelkezniük;
- műszakilag fenntarthatónak kell lenniük;
- politikai és műszaki (platform-) függetlenség jellemezze, vagyis az azonosítót működtető közösségnek szavatolnia kell, hogy a mindenkori műszaki-informatikai infrastruktúrától függetlenül működjön, a mindenkori politikai vezetés beleszólásának komoly veszélye nélkül.

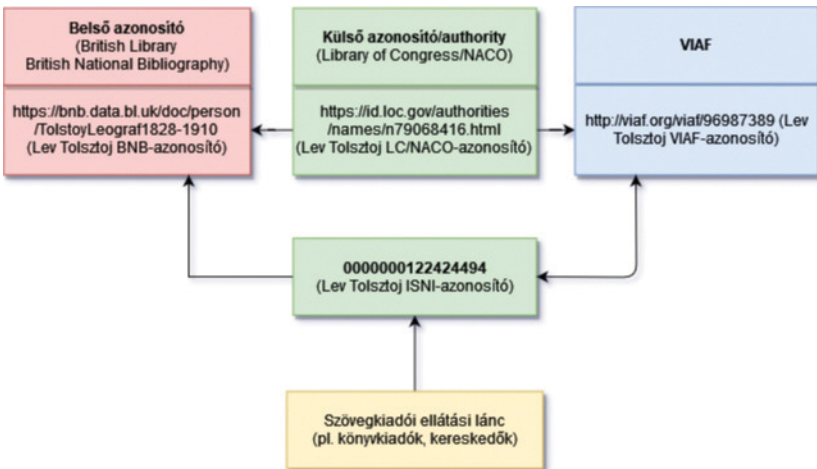
A különböző szakmai területek, ellátási láncok azonosítási egységesítésének szempontjából mérföldkönek tekinthető az ISNI.<sup>44</sup> A könyvtári metaadatok környezetében az ISNI már most jelentős megalapozottsággal bír személyek, csoportok és szervezetek azonosításában, emellett a zeneiparban is jelentős az alkal-

<sup>44</sup> Adamich, Thomas: *Linked Data Identifiers: Part 1 - International Standard Name Identifier (ISNI)*. = Technicalities, 34. vol. 2014. 1. no. 6–9. p. Library & Information Science Collection. ISSN 0272-0884



*Az azonosítás alapkövetelményei  
(Vincent Boulet prezentációja alapján)*

mazása az alkotók azonosítása terén. Használja már többek között a YouTube, az Apple és a Spotify is. A könyvkiadásban is terjed, német nyelvterületen például a *Verzeichnis Lieferbarer Bücher* (VLB)<sup>45</sup> nevű forgalomban lévő könyvekre vonatkozó (books-in-print) katalogizálási szolgáltatás egyben ISNI-regisztrációt is magában foglal.



*Az ISNI mint híd a tartalomszolgáltató ipar és a könyvtárszakma között brit példával  
(Az LC/NACO authorityt alkalmazó brit nemzeti bibliográfia rekordállománya a VIAF közbeiktatásával  
feleltethető meg a másik szektor által is használt ISNI-azonosítónak)*

<sup>45</sup> A VLB kezdőoldala: <https://vlb.de/> (2021.10.20.)

A nemzeti könyvtárak és dokumentációs szervezetek jellemzően a szövegkiadási szektorban alkalmazott nemzetközi szabványos azonosítók regisztrációs intézményei, mivel feladatkörük az adott nyelvterület dokumentumtermésének regisztrálása és megőrzése. Az ISBN, az ISSN vagy az ISNI azonosítóként egyaránt hídként funkcionál a könyvtárszakma és a tartalomipar között, megalapozva egy globális információs rendszer alapjait, amely megszünteti a különböző szektorok (például a könyvtárak) információtárának önálló szigetként való működését. Nemzetközi szinten az egyes azonosítók fenntartásáért felelős nemzetközi szervezetek gondoskodnak az azonosítási alapelvek érvényesüléséről és az azonosítók rendszerének működtetéséről. Specializáltabb szerepkörű azonosítók országos szintű regisztrálásáért jellemzően az adott szakterület információs szervezetei felelősek.

Ahogy már említettük, a nemzetközi szakmai szintéren az azonosítók megfeleltetésének egyik eszköze a VIAF. Ez egy olyan szolgáltatás könyvtárak és könyvtárhasználók számára, amely nemzetközi kapcsolt adat authority állományok széles választékához nyújt közös felületen hozzáférést.<sup>46</sup> Az OCLC nonprofit könyvtárszakmai<sup>47</sup> szervezet által működtetett szolgáltatást a hozzájáruló szervezetek (köztük nemzeti könyvtárak és nemzetközi azonosítók regisztrációs szervezetei) látják el authority állományokkal, amelyeket a VIAF összekapcsol és halmazokba (klaszterekbe) rendszerez. A szolgáltatás lehetővé teszi nevek, helyszínek, művek és kifejezési formák azonosítását, megtartva a regionális nyelvi sajátosságokat kiejtés és írás szintjén egyaránt. Több mint harminc országból több mint ötven hozzájáruló építi a VIAF állományát, amelyet a hozzájáruló intézmények küldöttei által alkotott VIAF Tanács felügyel. A szolgáltatás segítségével a könyvtári katalógusban tárolt metaadatok nemzetközi azonosítókkal (személyek ISNI-vel, könyvek ISBN-nel és a különböző entitások olyan „kváziszabványos” azonosítókkal mint pl. a Library of Congress/NACO) köthetők össze, megépítve az előző bekezdésben említett, a könyvtárakat a tartalomiparral összekötő hídhoz szükséges pilléreket.

**Bódog András**

<sup>46</sup> Dancs Szabolcs: *Bemutkozik az ISNI, a nemzetközi szabványos névazonosító.* = Tudományos és Műszaki Tájékoztatás, 64. évf. 2017. 4. sz. 199–207. p. <https://tmt.omikk.bme.hu/tmt/article/view/951> (2021.10.20.)

<sup>47</sup> Az OCLC kezdőoldala: <https://www.oclc.org/en/home.html?redirect=true> (2021.10.20.)

## „Olyan jó vagyok, mint te”: a szakértelem halála és az entitásmenedzsment az internet korában<sup>48</sup>

Ma már közhelynek számít az a megállapítás, hogy az interneten minden megtalálható. Ami pedig nem, az sok ember számára nem is létezik. Világossá vált tehát, hogy a könyvtáraknak is jelen kell lenniük az interneten, és ennek a kíváncsornak eleget is tettek. Az interneten elérhető tartalmak robbanásszerű növekedése azonban egy újabb helyzetet teremtett, aminek ismételten meg kell felelnünk.

A web 2.0 lehetővé tette az egyszeri internethasználó számára is, hogy tartalmat állítson elő és publikálja azt a weben. Nincs semmiféle szűrő vagy ellenőrzés, csupán csak a technológiát kell tudni használni. A közösségi média elterjedése pedig az információmegosztást alakította át radikálisan. Az internet gyökeresen változtatta meg azt, ahogyan informálódunk. Ez a változás viszont nem hozta el a vágyott információs Kánaánt, sőt a kialakult helyzet inkább disztópikus. A jelenséget több névvel is illeték: a szakértelem halála, igazság utáni társadalom.

Ahogy *Tom Nichols* írja a 2017-ben megjelent *The Death of Expertise* (A szakértelem halála) című könyvében: „a legnyilvánvalóbb probléma, hogy bárki bármit szabadon posztolhat online, elárasztva a világot hibás információkkal és félkész gondolatmenetekkel. Az internet mezején milliárdnyi virág virágozhat, ám ennek többsége bűzlik, kezdve random bloggerek üres gondolataival vagy zsémbes emberek összeesküvés-elméleteivel egészen a bizonyos csoportok és kormányok

<sup>48</sup> Philip Schreur *“I’m as good as you!”: The death of expertise and entity management in the age of the internet és Gordon Dunsire Bibliographic control in the fifth information age* című előadásai alapján. Előbbi elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 8. <https://youtu.be/ybUDrILt0kI?t=9985> (2021.10.20.) Utóbbi elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 12. <https://www.youtube.com/embed/sPrefNT8wG0?start=1991&end=4057> (2021.10.20.)

által indított kifinomult dezinformációs kampányokig. Az interneten lévő téves információ több okból lehet téves. Van, ami amiatt az, mert a szerzője felületes volt, van, ami amiatt az, mert jó szándékú emberek megosztanak valamit, amiről azt hiszik, hogy helyes, de tévednek, ám az is előfordul, hogy amiatt téves, mert kapzsiságból vagy rosszindulatból szándékosan közöltek téves információt. A médium maga mindenféle szerkesztői közbeavatkozás és kommentár nélkül egyenlőként jeleníti meg ezeket. Az internet csak egy tároló, nem bíráló. [...] A könyvtárak vagy legalábbis a tájékoztató szolgálatuk és a szakirodalmi szekciójuk, valaha az első szűrő szerepét játszotta, ami csökkentette az [információs] piac zaját. A könyvtárlátogatás már önmagában tanulás volt, főleg, ha az olvasó a könyvtárossal is interakcióba lépett. Az internet azonban nem hasonlít egy könyvtárra. Az internet inkább egy repozitórium, ahova bárki bármit betehet az első fóliánstól egy hamisított fényképig, egy tudományos értekezéstől a pornográfiaig, rövid hivatalos közleményektől a semmitmondó elektronikus graffitiiig. Ez egy szinte teljesen szabályozatlan közeg, ami lehetővé teszi, hogy a marketing, a politika és alulinformált laikusok döntései irányítsák szakértők véleménye helyett.”<sup>48</sup>

A könyvtárak által régebben betöltött szerepre tehát nagyobb szükség lenne, mint valaha, a szakma azonban még nem találta meg a módját, hogy a megváltozott környezetre át tudja értelmezni a feladatát. Ennek a feladatnak több alkotóeleme is van, ám most csak a könyvtári feldolgozó munkáról fogunk szót ejteni.

Ugyancsak közhelyszámba megy az a megállapítás, hogy a könyvtárak magas minőségű adatot állítanak elő a feldolgozás során, és az idő múlásával csupán a létrehozás és az adattovábbítás technikai részei változtak meg. A könyvtári világban kívül azonban már más a helyzet. Ott van például a Wikipédia, a szabad enciklopédia, amit az olvasói szerkesztenek. Ugyanez vonatkozik a Wikidata-ra és annak metaadatokkal való gazdagítására, illetve az azonosítókiosztásra. A Wikidata szerkesztéséhez semmilyen képzettség vagy szakismeret nem szükséges. Ennek megvan a maga pozitív és negatív oldala. Pozitív oldalon ott van például a költséghatékonyság, vagy hogy könnyebb a speciális tudásmorzsákat begyűjteni. Negatív oldalon pedig azt találjuk, hogy a laikusok nem feltétlenül tudnak jó minőségű (meta)adatokat előállítani, és az internet természete miatt nem tudhatjuk, hogy az adott adat megbízhatónak tekinthető-e. Márpedig a Google és a Wikidata sokak szemében autoratív információforrássá nőtte ki magát, így az is kiemelten fontossá vált, hogy milyen információ jelenik meg ezeken a felületeken.

<sup>49</sup> Nichols, Tom: *The death of expertise. The campaign against established knowledge and why it matters*. New York, Oxford University Press, 2017. 108–109; 110 p.

Ennek a problémakörnek a lehetséges megoldása az adatok eredetének megadása (adatproveniencia) lehet. Az adatproveniencia a minőségbiztosítás egy formája. Ha tudjuk, hogy ki hozta létre az adott (meta)adatot, akkor az segít abban, hogy el tudjuk különíteni a képzett szakemberek által létrehozott adatot a laikusokétól. Az adatproveniencia azonban több mint a létrehozó megnevezése. Könyvtári közegben idetartozik például az is, hogy mikor és milyen szabványokat követve hozták létre a metaadatot. Ez is nagyon fontos, hiszen a múltban készült leírás megtévesztő vagy használhatatlan is lehet a jelen kontextusban. Idővel változhatnak ugyanis a szabványaink is, vagy egy újonnan felfedezett információ átírhatja azt, amit igaznak gondoltunk. Kiderülhet például, hogy valakinek a halálzási évét rosszul ismertük.

A wikik rendelkeznek egy sajátos minőségbiztosítási mechanizmussal, amely különösen hasznos a többszerzős wikik esetében. Ha egy szerző állításait egy másik szerző átírja, akkor az első szerző állításai nem tűnnek el nyom nélkül, hanem mindig visszakövethető egy állítás teljes előzménytörténete, sőt egy állítás visszaállítható bármely korábbi állapotára is. Az adatproveniencia felértékelődésével várható ehhez hasonló mechanizmusok megjelenése a könyvtári adatszolgáltatásokban is. De nem a verzió- és előzménytörténet az egyedüli, amit érdemes eltanulnunk a wikiktől, hanem például az azonosítóalapú entitáskezelés is előrelépés a karakterlánc-alapúhoz képest.<sup>50</sup>

Egyre inkább azt látjuk kirajzolódni, hogy a könyvtáraknak szemléletükben érdemes közeledniük a különböző Wikimedia-termékekhez, ám egy ilyen közeledés várhatóan nem lesz problémamentes, mivel a kétféle adatmenedzsment-megközelítés között elég nagy a különbség. A hagyományos authority rendszer építéskor a könyvtárak úgy szoktak eljárni, hogy a rendszer egy ellentmondásmentes egészzé álljon össze, ahol például egy személyhez csakis egy leírás kapcsolódik, és egy leíráshoz csakis egy személy. Az ellentmondásmentesség és az egy-az-egyhez elv fenntartása érdekében minden authorityvel kapcsolatos művelet a behasonlításal kezdődik. A Wikidata ennél rugalmasabb munkafolyamatot tesz lehetővé azáltal, hogy az azonosítók és a leírások könnyedén összekapcsolhatók vagy összeolvaszthatók, ha szükséges. Ez a fajta rugalmasság segítette elő az adatok robbanásszerű sokasodását. Még nem látszik, hogy ez az ellentmondás feloldható-e, és sikerül-e egymást kiegészítő identitásmenedzsment-eszközzé integrálni az eltérő személet alapján épülő rendszereket, de már folynak ezzel kapcsolatos kísérletek. Ha ezek pozitív eredménnyel zárulnak, akkor – egy részterületen legalábbis – a könyvtár elkezdheti visszafoglalni a régi szerepét a társadalom életében.

**Ilácsa Szabina**

<sup>50</sup> A témáról bővebben az *Azonosítók modern könyvtári környezetben* című fejezetben olvashatnak.

## Tudásgráfok és többnyelvű visszakereső felületek

A BIBFRAME mellett a kapcsoltadat-univerzum másik kiemelkedő, lassan megkerülhetetlenné váló pontja a Wikidata adatbázis, amely egyre inkább helyet kér és kap a bibliográfiai számbavétel és az authority kontroll területén.

A Wiki-szolgáltatási univerzum részeként 2012-ben meginduló *Wikidata* a szemantikus, vagyis az adatok értelmezett kapcsolataiból épülő világháló különleges szereplője. Az adatbázis, amely bárki számára térítésmentesen igénybe vehető (akárcsak a Wiki-univerzum többi tagja), a Wikipédiához hasonlóan tudást tárol, alapegységei azonban nem szócikkek, hanem minden esetben saját URI-val és egyedi Wikidata-azonosítóval rendelkező elemek, amelyek egymást követő, rövid és tömör kijelentésekből, metaadat-közlésekből épülnek fel. Ezek az adatok számos Wikimedia-szolgáltatásba, így a Wikipédiába is átemelhetők, így a Wikidata voltaképpen a Wiki-univerzum központi adattáráként funkcionál. Ugyanakkor külső szolgáltatásokba is képes adatokat továbbítani, amelyeket egyebek mellett a VIAF, illetve a Google tudásgráfja (Google Knowledge Graph) is felhasznál – ez utóbbinak markáns jele a személyek, testületek keresésénél a találati képernyő jobb oldalán megjelenő tudáspanel, infobox.

A Bibliographic Control konferencia idején, 2021 februárjában az adatbázis 91 millió elemet tárolt – hét-nyolc hónappal későbbre ez a szám már csaknem 96 millióra emelkedett. Ezek jelentős hányada (februári adatok alapján a teljes mennyiség körülbelül egyharmada) különféle tudományos publikációk adatait rögzíti, ezért a Wikidatát tudományometriai szolgáltatások működtetésére is felhasználják, ilyen például a Scholia, amely a Wikidatából aktuálisan, a keresés pillanatában lekérdezett adatmennyiséget – szerzők, testületek, helyszínek, illetve publikációk legfontosabb adatait és azok sokszínű kapcsolatrendszerét – igen sok összefüggésben, számos megjelenítésmóddal (lista, grafikon, térkép) képes bemutatni a használnak.<sup>51</sup> Az elemek további egytizede személyek leírását adja.

<sup>51</sup> A szolgáltatás hozzáférhető a <https://scholia.toolforge.org/> címen.

Egy elem mindig tartalmaz legalább egy címkét, tehát az elem megnevezését egy adott nyelven. Ilyen címkék felvételére korlátlan számú nyelvi változatban van lehetőség: a többnyelvűség a visszakeresés folyamatát támogatja, mivel az egyes visszakereső, megjelenítő felületek automatikusan válogathatnak a rögzített címkeváltozatok közül a felhasználó által megadott nyelvi beállításoknak megfelelően. A címkék mellett lehetőség van rövid értelmező leírást adni, ez az adott megnevezés egyértelműsítésére, az azonos alakú névformák egymástól való megkülönböztetésére szolgál, és magában a Wikidatában történő keresést is megkönnyíti, mivel rögzítése esetén részét képezi a keresőmezőbe gépelés közben érkező keresési javaslatoknak. Végül minden nyelvi címkeváltozathoz tartozhatnak névvariánsok is.

Ezt követi az elem tulajdonságainak (properties) megadása, amelyek közül kiemelkednek az úgynevezett azonosító típusú tulajdonságok (identifiers). Ezeket keresztül a Wikidatában előforduló entitás-előfordulás leírása összekapcsolható az azonos tartalmú, de más forrásban található adathalmazokkal. A hagyományos, közgyűjteményi authority állományok rekordjaira mutató azonosítók mellett szótárak, enciklopédiák, biográfiák és más források (ISNI-adatbázis, Discogs, Europeana, IMDb vagy akár az Instagram, Twitter, YouTube stb.) felé is létrehozhatók hivatkozások, így a keresett entitás-előfordulásról átfogó információ szerezhető; bizonyos elemek leírása akár 90–100 külső hivatkozást is tartalmaz.

A Wikidatára irányuló figyelem az utóbbi években a hagyományos technológiákon és módszertanon alapuló bibliográfiai forrásleírás krízise miatt erőteljesen megnőtt. Ez a krízis jelenti egyrészt az elektronikus bibliográfiai források nehézkes kezelését, ugyanakkor jelenti az RDA megjelenésével egyre nagyobb szerephez jutó entításalapú feldolgozással kapcsolatban felmerülő kompatibilitási problémákat is – az ilyen leírások támogatására a hagyományos csereformátum nem, vagy csak nagyon korlátozottan alkalmas. A Wikidatában tárolt forrásleírások azonban már korszerűbb alapokra, a kapcsoltadat-technológiára, vagyis az adatok összekapcsolódására, az úgynevezett tripletekre és a belőlük épülő gráfokra épülnek, amelyek segítségével – legalábbis részben – megoldható az egyes entításokra vonatkozó metaadatközlések elkülönítése. A külső azonosítók társításával, az egyes adatelemek minősítésével (kitüntetett, elavult stb.), valamint az adatelemek forrásainak rögzítésével átfogó, informatív leírások készíthetők az egyes forrásokról.<sup>52</sup>

<sup>52</sup> A leíráshoz felhasználható metaadatelemek, illetve külső azonosítók forrástípusok szerint csoportosított listáját lásd a [https://www.wikidata.org/wiki/Template:Bibliographic\\_properties](https://www.wikidata.org/wiki/Template:Bibliographic_properties) című oldalon, az egyes relációk entítások szerinti, alapszintű csoportosítását pedig a [https://www.wikidata.org/wiki/Wikidata:WikiProject\\_Books](https://www.wikidata.org/wiki/Wikidata:WikiProject_Books) című oldal közli.



A Wikidata adatbázis használata a bibliográfiai számbavétel szempontjából elméleti, illetve gyakorlati síkon egyaránt értékelhető. Az előadók szerint a számbavételt, illetve az azt támogató authority kontroll folyamatát úgy kell kezelni, mint az általános emberi tudás rögzítésének egy szeletét, s a Wikidata felületének újszerű megközelítése révén megvalósulhat ez az összeolvadás. A közösségi, megosztott/megosztható munkavégzésnek köszönhetően az identitáskezelés (identity management), azaz az entitás-előfordulások adatainak gondozása összhangban van a nemzetközi katalogizálási alapelvekben leírt két követelménnyel, amelyek a felhasználó kényelmére, illetve az általános szóhasználatra vonatkoznak.

Gyakorlati szempontból ki kell emelni a Wikidatában elkészíthető forrásleírások maximális testreszabhatóságát, az alkalmazási profilok elkészítésének tekintetében tapasztalható igen nagy rugalmasságot. Az előadók úgy vélik, az alulról felfelé történő építkezési módszernek és a közösségi munkavégzésnek hála, megvalósulhatnak az UBC-program céljai, amelyeket a nemzeti bibliográfiai ügynökségek hosszú évtizedek óta tartó együttműködésével sem sikerült teljes mértékben megvalósítani.

A Wikidata, illetve az azt működtető Wikibase szoftver használata, valamint az egymással összekapcsolódó adatok előállítása, kezelése és megjelenítése az OCLC tevékenységében is fontos szerepet kapott. *John Chapman* számolt be arról, hogy a kapcsoltadat-technológiával összefüggő kísérletek, projektek már 2009 óta jelen vannak az OCLC tevékenységében, elég ha a FAST vagy a VIAF adathalmazára, vagy éppen a WorldCat találataihoz kapcsolódó,<sup>53</sup> a schema.org szótár segítségével megjelenített állításokra gondolunk. A 2010-es évek első felében *EntityJS* néven megjelenítő eszközt fejlesztettek ki, ahol a keresőkifejezésekre érkező találatok entítások – személyek, testületek, fogalmak, helyek, események, illetve művek – alapján válogatva jelentek meg. Egy entitás-előfordulást, például egy konkrét személyt kiválasztva a hozzá kapcsolódó más entitás-előfordulások listáját kapta a felhasználó. A szolgáltatásplatformot adatminőség-javító segéd-eszköz is kiegészítette, amely az OpenRefine-hoz hasonló adattisztítási lehetőségeket biztosított a felhasználóknak (CONTENTdm Metadata Refinery).

A Wiki-univerzum workflowba illesztésére már 2017–18-ban, a *Project Passage* néven ismert kezdeményezésben is sor került, amelynek célja közgyűjteményi kapcsoltadat-adathalmazok előállítása, illetve ezek külső azonosítókkal történő, széles körű adatgazdagítása volt a Wikibase felhasználásával. Az idén befejeződő, két évig futó Entity Management Infrastructure program célja az ekkor megalkotott adathalmaz további bővítése, és olyan szolgáltatássá terebélyesítése, mely más

<sup>53</sup> Az OCLC közlése alapján ez a szolgáltatás 2021 februárjáig állt fenn: [https://help.oclc.org/Discovery\\_and\\_Reference/WorldCat-org/Troubleshooting/What\\_happened\\_to\\_the\\_linked\\_data\\_that\\_previously\\_displayed\\_on\\_records\\_in\\_WorldCat.org](https://help.oclc.org/Discovery_and_Reference/WorldCat-org/Troubleshooting/What_happened_to_the_linked_data_that_previously_displayed_on_records_in_WorldCat.org)

termékeket is támogathat a jövőben. A projekt keretében közösségi tudásgráf épül, amely közgyűjteményi authority állományok, a WorldCatben tárolt műinformációk és ellenőrzött szótárak adattartalmát gyűjti egybe, kiegészítve azok forrásaira vonatkozó adatokkal. A gráfot API-kon és lekérdezési végpontokon keresztül szolgáltatják.

Egy tudásgráf megalkotásának és karbantartásának folyamata alapvetően négy hívószó köré csoportosítható: *tudásalkotás*, amely az adatforrások integrálását, illetve a tripletek megalkotását (semantic lifting), a tudás kinyerését (extracting) jelenti; *a tudástárolás*, amely fázisban az adathalmazt valamilyen szemantikus repozitóriumban (gráfadatbázis, triplestore) helyezik el; *tudáskarbantartás*, amely az adatok helyessége és teljessége érdekében végrehajtott tevékenységeket összegzi, hiszen egy ilyen gráf soha nem egy statikus produktum; végül a tudásbevetés, amikor a gráf információtartalma alkalmazásokban hasznosul.

Az OCLC-n belül épített tudásgráf, ahogy már említettem, külső adatforrások (VIAF, WorldCat, sőt a Wikidata) anyagának importálásával és konverziójával bővül. A kinyert tudást nTriples-szintaxisban tárolják, majd entitásfelismerési-adatgazdagítási folyamatnak vetik alá, melynek során összekötik a forrásadathalmazokban található azonos entitás-előfordulások adatait. A kibővített tudásgráf már JSON-szintaxisban áll rendelkezésre, és fontos eleme az adatok többnyelvű tárolása. Az entitás-előfordulásokról rögzítendő adatokat a Minimum Viable Entity (MVE) dokumentáció, avagy adatminőségi modell tartalmazza – ez a Wikidata adatmodellje, illetve az LRM és a BIBFRAME relációinak tanulmányozása során készült. Az adatminőségnek a pontossághoz, az egyértelműséghez, a megbízhatóságához, illetve a szerkezeti megfeleléshez kapcsolódó paraméterei egyaránt vannak.

A tudásgráf a legtöbb metaadatok felhasználatában használható: ha a feldolgozási folyamat során autorizált adatokra van szükség, az entitáskezelő rendszerben kell a szükséges entitás-előfordulásra rákeresni (lesz felhasználói interfész, de alapvetően API közbeiktatásával), s az arra mutató azonosítót elhelyezni a metaadatok közt. A WorldCat rekordjai például művekre, személyekre, illetve földrajzi helyekre mutató, az Entity Manager felé vezető hivatkozások sokaságát tartalmazhatják.

A szemantikus web technológiáján alapuló bibliográfiai adatfeldolgozás és -tárolás egyik legnagyobb lehetősége (és kihívása) a metaadatok és a hozzájuk rendelt értékek többnyelvűsítése. E kihívással kapcsolatos gondolatait *Pat Riva*, a könyvtári referenciamodell egyik kidolgozója osztotta meg előadásában. Gondolatmenetét a UBC-programban vállalt feladatoknál, illetve a hozzáférési pontok nemzetközi egységességének fejtegetésénél kezdte. Az egyes nemzetek bibliográfiai gyakorlata eltérően határozza meg ugyanazon névhordozó (pl. Platón) nevének autorizált alakját, így a globális egységesség nem valósul meg, helyette igen heterogén névváltozatok jelennek meg számos nyelven, amelyeket manapság

előszeretettel kapcsolnak össze és deklarálják azok egyenlőségét. Ennek a koncepciónak – tudniillik hogy minden, a világ közgyűjteményeiben kitüntetettnek minősített névformát össze kell gyűjteni – közismert példája a VIAF.

A bibliográfiai adatok több nyelven történő visszakereshetőségéről és megjelenítéséről azonban szükséges tovább gondolkodni. A felhasználók részéről már hosszú ideje jelen van a multilingvális hozzáférés igénye, mivel a közgyűjtemény használói populációja – egészét tekintve – többnyelvű, az egyes emberek is beszélhetnek több nyelven, nem is beszélve a gyűjtemények és az egyes bibliográfiai források nyelvi heterogenitásáról. Napjaink katalógusai azonban egy nyelv használatára épülnek, s jóllehet az RDA-szabályzat már demokratikusabban kezeli a leírás nyelvének kérdését, kevés útmutatást nyújt az egynél több kitüntetett nyelv (preferred language) használatára – így jó gyakorlatból is viszonylag kevés áll rendelkezésre.

A két nyelven történő bibliográfiai forrásleírás szép példája Kanada. Az ország könyvtárai a francia nyelvű forrásokat francia, az angol nyelvűeket angol nyelven írják le, a kétnyelvű forrásokról pedig két különálló, eltérő nyelvű rekord készül. Ezek a különböző nyelveken készülő forrásleírások extrém esetben két katalógusfelületen keresztül férhetők hozzá, silószerű architektúrát megvalósítva ezzel. A több nyelvet rendszeresen, napi szinten alkalmazó használóknak azonban hasznosabb, ha egyetlen kereséssel, egyetlen felületen kapják a releváns találatokat, függetlenül azok nyelvétől. Ezt az igényt igyekszik kiszolgálni a Quebec tizen-nyolc egyetemét tömörítő konzorcium 2020 nyarán indult közös katalógusrendszere, a Sofia.<sup>54</sup> Ebben a rendszerben egy rekord tárolódik, amelynek tartalma a felhasználói felület nyelvi beállításainak megváltoztatásával egyidőben átalakul. A helyes működéshez a metaadatelemek megnevezéseinek (a mezőcímkéknek), illetve az értékszótárak elemeinek fordítása szükséges, s a külső helyről beemelt autorizált névalakok is az adott nemzet gyakorlatának megfelelő formában jelenhetnek meg, hála az RDF-ben rögzíthető számos címkének. A többnyelvűség igénye a tárgyszavakra történő keresés során is jelentkezik, ekkor a tartalomjelölő kifejezések nyelvi változatait vagy maga az authority állomány tartalmazza, vagy valamilyen külső fordítószolgáltatást kell igénybe venni az információkeresési folyamat végrehajtása során. Ehhez arra is szükség van, hogy a felhasználó keresésének a nyelve egyértelműen megállapítható legyen, jó példa erre az *information* kifejezés, amely angolul és franciául egyaránt értelmezhető.

<sup>54</sup> Hozzáférhető a következő címen: <https://sofia-biblios-uni-qc.org/en/>.

A feldolgozó könyvtárosoknak régóta ismerősek a párhuzamos adatokkal vagy többes címoldallal rendelkező kiadványok, ezekben az esetekben a főcím kiválasztásakor alkalmazási sorrend tekintetében egymással fel nem cserélhető szabályokat kell alkalmazniuk. Vannak azonban olyan kiadványok is – az úgynevezett tête-bêche elrendezésűek – amelyek egy művet két fordításban tartalmaznak, s ezek egymáshoz képest fordított állásban (fejfel lefelé) helyezkednek el. Ilyenkor az elsődleges nyelv kiválasztása önkényes, ugyanakkor felmerül a kérdés, hogy egy vagy két leírás készüljön, s amennyiben a feldolgozó az utóbbi mellett dönt, hogyan lehet egymással összekötni ezeket a leírásokat? – tette fel a kérdést Pat Riva.

Hubay Miklós

## Röviden a könyvtárak webes láthatóságáról és a schema.org-ről<sup>55</sup>

Richard Wallis kapcsolattartó-szakértő a Bibliographic Control konferencián a *schema.org*<sup>56</sup> nyílt szótár lehetséges könyvtári felhasználását ismertette. A strukturált adatok hasznossága elsősorban a keresőmotorok tükrében nyilvánul meg, hiszen részletes (és szabványos) leírásokkal szolgál a keresett dologról. Számptalan korai kezdeményezést követően a 2011-ben színre lépett *schema.org* egyre inkább a keresőmotorokat támogató, a webes láthatóságot elősegítő strukturált adatok szótárává válik a weblapok HTML-kódjába beágyazott formátumban, ami Microdata, RDFa vagy JSON-LD lehet. A *schema.org* jelentős penetrációval rendelkezik, a weboldalak közel felénél már alkalmazásban van. Google, Bing, Yahoo!, Facebook, Apple – csupán néhány példa a szótárt felhasználó nagy szolgáltatókra.

<sup>55</sup> Richard Wallis *Follow me to the library! Bibliographic data in a discovery driven world* című előadása alapján. Elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 11. <https://youtu.be/EoCt3ZYBmWI?t=5667> (2021.10.20.)

<sup>56</sup> A Schema.org kezdőoldala: <https://schema.org/> (2021.10.20.)

<b>Könyvtárak</b>	<b>Web</b>
Kapcsolt adatok	Kapcsolt adatok
Könyvtárszakmai standard szótárak (RDA, BIBFRAME)	Közös globális szótárak (schema.org)
Célirányos katalógusfelület (discovery)	Az adatok célirányos előkészítése a keresőmotoroknak
Entitásalapú keresés és részleges adatmegjelenítés	Rich result (adatgazdag találatok vizuális megjelenítése)
Csak könyvtárak (lokális)	Globális (web)

*Eltérő adatfelhasználás a könyvtárak és a világháló esetében*

Ha a könyvtári források webes láthatóságát vizsgáljuk, akkor megállapítható, hogy a web csak a katalógus (vagy discovery) felületét látja, magukat a feldolgozott forrásokat már nem. Ez azért probléma, mivel az átlagos használó közvetlenül a forrást keresi és nem a könyvtárat, ahol a keresett forrás jó eséllyel elérhető. A könyvtárak számára tehát kézenfekvő lenne láthatóvá tenniük az őrzött forrásokat a könyvtári rendszer adatbázisán kívül is, tehát magán a weben. Erre kézenfekvő módszer lehetne, ha schema.org-adatokkal bővítenék ki a discovery jellegű katalógusok felületét. Az alábbi összehasonlítás a kapcsolt adatokat veti össze a könyvtári standard BIBFRAME és a webes standard schema.org tükrében.

<b>BIBFRAME</b>	<b>Schema.org</b>
Könyvtári standard	Globális webes standard
A MARC utódja	Nem könyvtári célra készült
A Kongresszusi Könyvtár fejleszti	Nem informatikai vállalatok (Google, Apple stb.) támogatják
A könyvtárinformatikai fejlesztők foglalkoznak vele	A könyvtárinformatikai fejlesztők még nem foglalkoznak vele
A könyvtári felületet korszerűsíti	A globális szintű felfedezést/hivatkozást segíti elő
Jelentős előrelépés	Újfajta megközelítés
Fejlesztési cél	Hasznos eszköz

*Példa a kapcsolt adatok felosztására könyvtári és webes felhasználás szerint*

Richard Wallis szerint a két megközelítés párosítása lenne az ideális. E célból jött létre a két szótár megfeleltetésére szolgáló *bibframe2schema.org*<sup>57</sup> kezdeményezés.

Összegezve mindazon tényezőket, amelyek a könyvtárak webes láthatóságához szükségesek:

- megfelelőnek kell lenniük a belső kapcsolt adatoknak – a BIBFRAME megoldás erre;
- viszont ne várjuk el, hogy az átlag felhasználók is könyvtárszakmai szótárakat alkalmazzanak;
- a webes kimenetnek de facto szabványos szótárnak kell lennie – erre a *schema.org* lehet a megoldás, de emellett szakmai kontrollált szótárakra is szükség van;
- legyen könnyen használható és konzisztens rendszer a fejlesztőknek és a megvalósítóknak egyaránt (MARC 21 to BIBFRAME<sup>58</sup> és *Bibframe2schema.org* kezdeményezések);
- a közösségi részvétel fontossága.

**Bódog András**

<sup>57</sup> A *bibframe2schema.org* kezdőoldala: (<https://bibframe2schema.org/index>) (2021.10.20.)

<sup>58</sup> A Library of Congress weboldala. *MARC 21 to BIBFRAME 2.0 Conversion Specifications* (<https://www.loc.gov/bibframe/mtbf/>) (2021.10.20.)

## Metaadat-menedzsment a könyvtári falakon túlmutató összefogásban<sup>59</sup>

A bibliográfiai számbavétellel foglalkozó víziók többségében markánsan jelen van a könyvtári falakon túlmutató együttműködés a metaadatok kezelése (készítése, hitelesítése, terjesztése) terén.<sup>60</sup>

*Michele Casalini* a rendszerek közötti együttműködést azonosítja a jelenkori könyvtári világot jellemző egyik legfontosabb tendenciaként. Az authority állományok építése is ezen az új megközelítésen alapul. Egyszerre támaszkodik ugyanis az adatgazdagítás és -validálás manuális és automatikus folyamataira, illetve a tudásbázisokra mint a hitelesség forrásaira, és ezáltal magasabb szintre emeli a tudás terjesztésében érdekelt nemzetközi intézmények és szervezetek közötti kooperációt. Az adatok töménytelen mennyisége mellett is fennáll ugyanis az igény a hozzáférési információk minőségének és stabilitásának egyfajta garantálására, amelyhez az eltérő nyelvi és kulturális kontextusokban alkalmazott korábbi technológiák és standardok nem jelentenek biztos alapot. Az adatok újrafelhasználását előtérbe állító új felfogás az adat, illetve az authority kontroll mibenlétének újragondolására készíti bennünket, és a fejlődés eme természetes vonalán jutunk el a rekordtól az entitásig. Az entítások azonosítása az entításkapcsolatok meghatározásával egy, a korábbinál rugalmasabb látásmódot feltételez, valamint heterogén szervezeti együttműködést, ahol az eltérő forrásokból származó adatok összehasonlítása és integrációja gondoskodik az információk minél nagyobb teljességéről és hitelességéről. A közgyűjtemények az adatok létrehozásában és gazdagításában is át kell lépjenek korábbi korlátaikat. Az interoperabilitáshoz ugyanakkor az entítások egyedi elnevezésekkel, kódokkal való ellátására van szükség, amely szerepet az egységes forrásazonosító (URI) tölti be, az entítások-

<sup>59</sup> Michele Casalini *The future of bibliographic services in light of new concepts of authority control* című előadása alapján. Elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 12. <https://youtu.be/sPrefNT8wG0?t=6230> (2021.10.20.)

<sup>60</sup> A könyvpiari értéklánra kiterjedő metaadat-menedzsmentről lásd: Attanasio, Piero: Metadata Management in the Book Supply Chain. Elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 9. [https://youtu.be/Z\\_MDPpzVRI4?t=1974](https://youtu.be/Z_MDPpzVRI4?t=1974) (2021.10.20.)

ról való közlések normalizált nyelve pedig a forrásleíró keretrendszer (Resource Description Framework – RDF). Az RDF-nyelven megfogalmazott állítások elemeit különböző forrásokból is vehetjük (pl. RDA, BIBFRAME) összekapcsolva ezzel eltérő standardokat és modelleket.

A 2016-ban indult *Share Virtual Discovery Environment* (Share-VDE vagy SVDE) projekt célja az új technológiák és standardok gyakorlatba való átültetése egy közös (osztott) katalogizálási környezet létrehozásával. A kezdeményezés egyik alapítója a bibliográfiai ügynökségeként is ténykedő *Casalini Libri*, amely műszaki testvérvállalatával, az *@Cultal* közösen régóta tesztelik a kapcsolattad-technológián alapuló könyvtári megoldásokat. E tapasztalatokra építve célozták meg az SVDE-projekten belül a bibliográfiai rekordok authority kontrolljának újszerű megközelítését olyan eljárások létrehozásával, amelyek megújítják a könyvtárak authority szolgáltatásait, valamint előmozdítják átállásukat a kapcsolt adatok alkalmazására.

A projekt kiindulási alapja a könyvtárak állandó igénye, hogy bibliográfiai és authority rekordjaikat hiteles forrásokból tudják frissíteni MARC-formátumban vagy a BIBFRAME kapcsolattad-struktúráját követve. A nemzeti és tudományos könyvtárak aktív részvételével megvalósuló kezdeményezés főbb céljai:

- MARC-rekordok adatgazdagítása azonosítókkal külső forrásokból (pl. ISNI, VIAF);
- entitások egymásnak való megfeleltetése, klaszterezése;
- MARC-RDF-konverzió a BIBFRAME szótára és egyéb ontológiák alkalmazásával;
- automatizált authority szolgáltatások, adatfrissítési és adatterjesztési eljárások;
- manuális entitáskezelő eszköz (J. Cricket);
- adatpublikálás entításalapú felhasználói felületen ([www.share-vde.org](http://www.share-vde.org)).

Az interoperabilitással kapcsolatos egyik legfontosabb kihívás, hogy az adatmodelleinket úgy szükséges egymással megfeleltetnünk, hogy az ne menjen az adatrészletesség rovására. Az SVDE-projektben ezért jóváhagytak egy SVDE Opus Osztályt, amely egyszerre felel meg a BIBFRAME és az SVDE műentitásának. Ahogy látszik, a projekt voltaképp az észak-amerikai BIBFRAME-modell entitásmegközelítését próbálja kibékíteni az Európa-központú IFLA LRM-ben megfogalmazódó megközelítéssel. Mindez jól szolgálja az eltérő katalogizálási gyakorlatok közötti átjárást.



A projekt keretében készülő tudásbázis többféle módon épül, és többféle forrást ötvöz. A mechanikus algoritmusok révén létrejött adattartam manuálisan finomítható-pontosítható, például a könyvtárak munkatársai által, míg az eltérő eljárások szerint épülő VIAF-ból és Wikidatából származó információk egyszerre biztosítják a minőséget és a nyitottságot a széles felhasználói közösségek felé. A tervek szerint a jelenleg J. Cricketként ismert tudásbázis-szerkesztő a könyvtári authority szolgáltatás eszközévé növi ki magát, alkalmassá válva a kapcsolattal alapú entitáskezelésre.

A Share-VDE-típusú projektek eredményeinek mielőbbi implementálásával a könyvtárak authority szolgáltatásai idővel nyitottabbá és kooperatívabbá válhatnak.

Dancs Szabolcs

## Az osztályozás és a tárgyszavazás modern könyvtári környezetben<sup>61</sup>

A könyvtári feltáró munka bibliográfiai adatok rögzítését követő szintje az osztályozás. A világot leíró fogalmakat azonos ismérvek mentén tudjuk csoportosítani, osztályokba rendezni, majd ezen osztályok jelzeteivel meghatározni egy adott dokumentum tartalmát. A könyvtártudomány fejlődése során számos osztályozási rendszer alakult ki, többnyire a 19. század közepétől kezdődően. Napjainkban a leggyakoribb osztályozási rendszerek a *Melvil Dewey*-féle *tízdeses osztályozás* (Dewey Decimal Classification – DDC), ennek továbbfejlesztéseként az *Egyetemes Tízdeses Osztályozás* (Universal Decimal Classification – UDC, magyarul ETO), illetve a Kongresszusi Könyvtár osztályozási rendszere (LCC).

<sup>61</sup> A cikk elkészítésekor az alábbi előadásokat vettem alapul:

Lucarelli, Anna: *Thesauri in the digital ecosystem*. Elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 11. <https://youtu.be/EoCt3ZYBmWI?t=319> (2021.10.20.)

Mödden, Elisabeth: *Artificial intelligence, machine learning and bibliographic control*. Elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 10. <https://youtu.be/Yo6Vi72E1T4?t=125> (2021.10.20.)

Suominen, Osmo – Kirjasto, Kansallis: *Annif and Finto AI: Developing and implementing automated subject indexing*. Elhangzott: Bibliographic Control in the Digital Ecosystem: International Conference, Firenze, 2021. február 10. <https://youtu.be/Yo6Vi72E1T4?t=1894> (2021.10.20.)

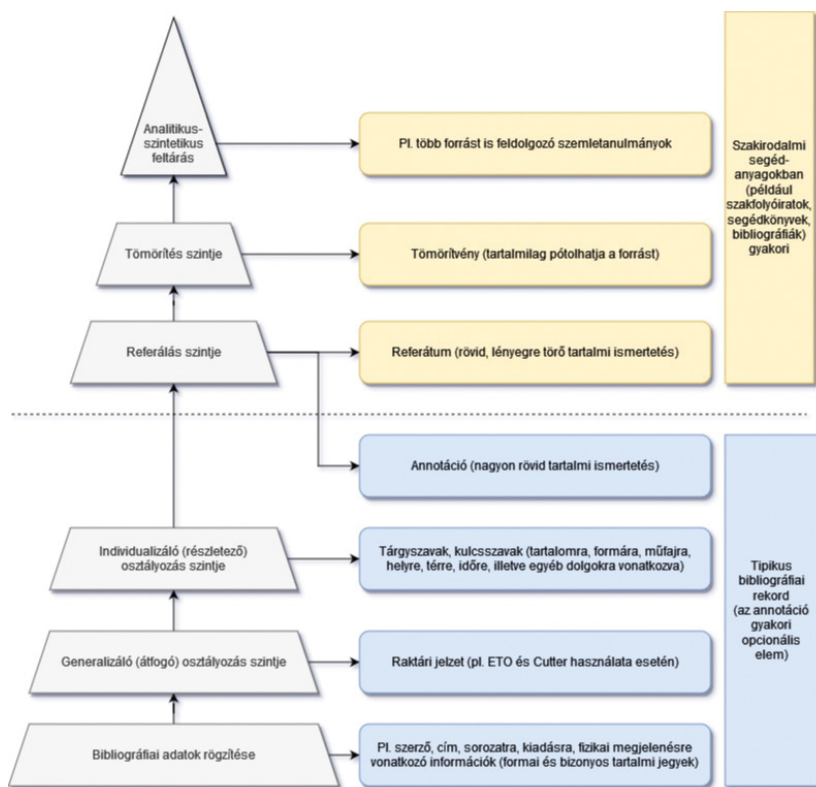
Dewey a világot leíró fogalmakat tíz főosztályba sorolta 0-tól 9-ig terjedő számozással (ezért nevezzük tizedes osztályozásnak). Hierarchikus osztályozási rendszer révén ez további alárendelt fogalmakkal bővíthető. A Dewey-féle rendszert *Paul Otlet* és *Henri La Fontaine* fejlesztették az emberiség teljes tudását átfogni képes osztályozási rendszerré. Az így létrejött ETO a különböző fogalmak összekapcsolásával új fogalmak jelzetelesét is lehetővé teszi, és a fogalmak közötti viszonyok kifejezésével (összekötés, összefoglalás, mellé- és alárendelés, csoportosítás) már közös tulajdonságok (fazetták) szerinti csoportképzésre is alkalmas. Mindkét tizedes osztályozás fő előnye, hogy nyelvfüggetlen rendszer, az egyes fogalmakat számokkal jelöli, így nemzetközi szinten is könnyen használható. Az ETO-t rendszeresen fejlesztik és frissítik, így rugalmasan követi a világ történéseit.<sup>62</sup> Az LCC mélyfeltárással kevésbé alkalmas, ám új fogalmakkal könnyen bővíthető osztályozási rendszer, amelyet elsősorban angol nyelvterületen alkalmaznak, mivel a Kongresszusi Könyvtár az Egyesült Államok nemzeti könyvtáraként e rendszer szerint tárja fel az állományát, és sok esetben a kötetek eleve ilyen jelzettel kerülnek forgalomba.<sup>63</sup>

A tartalmi feltárás mélysége függ a dokumentum és a könyvtár típusától (például az adott terület szakkönyvtára részletesebb tartalomfeltárást végez, mint egy közkönyvtár), a használói igényektől, de az adott könyvtár infrastruktúrájától és munkatársainak kapacitásától is. A dokumentum könyvtári feltárást összegző, a visszakeresést segítő szurrogátum (amely a dokumentum azonosítóját, formai és tartalmi ismérveit tükrözi) az osztályozás szintjén általában generalizáló (átfogó) és individualizáló (részletező) osztályozási jelzetekeket tartalmaz.<sup>64</sup> Előbbi tágran csoportosítja egy könyvtár állományát különböző témák szerint egy osztályozási (pl. az ETO) és a raktári rend kialakítását lehetővé tevő rendszer (pl. a Cutter-szám) jelzeteivel (együttesen alkotják a raktári jelzetekeket), míg a részletező egy-egy dokumentum tartalmát tárja fel részletesebben, általában tárgyszavakat alkalmazva. Még mélyebb tartalmi feltárást már annotálás (nagyon rövid összefoglaló) és referálás (rövid szöveges összefoglaló), tömörítés (az eredeti szakirodalmi szöveget kiváltani képes összefoglaló), illetve analitikus-szintetikus feltárás (több mű szintetizálása) szintjén lehet megvalósítani.<sup>65</sup>

<sup>62</sup> Piros Attila: *Az Egyetemes Tizedes Osztályozás húsz éve. Az ETO változásai 2000 után.* = Tudományos és Műszaki Tájékoztatás, 66. évf. 2019. 10. sz. 587–607. p. <https://tmt.omikk.bme.hu/tmt/article/view/12396> (2021.10.20.)

<sup>63</sup> Horváth Tibor – Sütheő Péter: *A tartalmi feltárás: 2.2. Tudományfelosztáson alapuló, hierarchikus osztályozási rendszerek.* In: *Könyvtárosok kézikönyve 2. kötet – Feltárás és visszakeresés.* Szerk. Horváth Tibor, Papp István, Budapest, Osiris, 2001. [64]–71. p.

<sup>64</sup> Horváth Tibor: *A feldolgozás egységei és a feltárás célja.* In: *Könyvtárosok kézikönyve 2. kötet – Feltárás és visszakeresés.* Szerk. Horváth Tibor, Papp István, Budapest, Osiris, 2001. 20–23. p.



### A feltárás szintjei

A tárgyszó az úgynevezett információkereső nyelvvel (az a közös nyelvrendszer, amely a könyvtári feltárásban és a kívánt információ visszakeresésénél egyaránt használatos)<sup>66</sup> kifejezett adott ismérv szabályozott formája. A tárgyszó utalhat a dokumentum tartalma mellett más ismérvekre is, például műfajra, formára, helyre, időre, de lehet név is. Az általános vagy egy-egy szakterületet részletező tárgyszórendszerek magukon a tárgyszavakon kívül még utalásokat (lásd, lásd még) és magyarázó-értelmező szövegeket tartalmaznak.<sup>67</sup> A tárgyszavakat a szintén információkeresésre alkalmazott, ám szabadon kiválasztott szöveg- és címszavaktól (vagyis a dokumentum szövegében, illetve címében előforduló szavaktól) a szabályozottság ténye különíti el, ugyanis a tárgyszavakat kontrollált

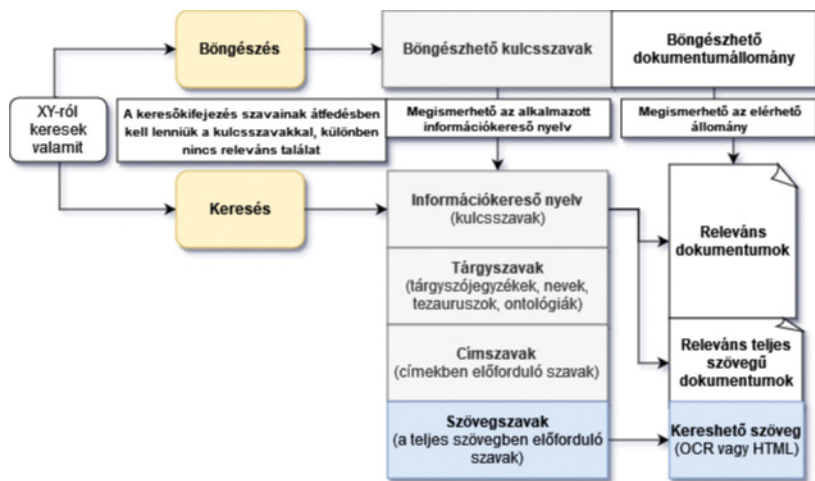
<sup>65</sup> Eszenyiné Borbély Mária – Salgáné Medveczki Marianna: *Komplex könyvtári feldolgozó munka gyakorlata* [online]. Budapest, Kempelen Farkas Hallgatói Információs Központ, 2010. <https://dtk.tankonyvtar.hu/xmlui/handle/123456789/13012> (2021.10.20.)

<sup>66</sup> Horváth Tibor: *i. m.* 28. p.

<sup>67</sup> Horváth – Sütheő: *i. m.* 73–76. p.

szótárakból (pl. tárgyszójegyzékekből, tezauruszokból) kell kiválasztani. Kulcsszavak alatt értjük azokat a keresőszavakat, amelyek segítenek megtalálni a kívánt információt, tárgyszavakat, szöveg- és címszavakat egyaránt magukban foglalva.<sup>68</sup> Alternatív információkereső eszköz a katalógusban feldolgozott kulcsszavak és dokumentumrekordok böngészhető (esetleg fazettákba rendszerezett) listája.

A tartalmi feltárás nagy dilemmája és állandó kihívása, hogy a használók által alkalmazott keresőkifejezéseknek átfedésben kell lenniük a feltárás során használt leírókifejezésekkel. A könyvtártudomány fejlődése során számos eszközt fejlesztettek ki a különböző névalakok, szinonimák kezelésére. A probléma megoldásának legalapvetőbb eszköze az utalók használata. Szinonimák esetén a *lásd* utaló vezet el a feltárás során használt szabályozott névalakhoz, míg a *lásd még* utalók különböző kapcsolatokra utalnak az adott kulcsszó vagy tárgyszó viszonylatában. Az egy-egy szűkebb vagy tágabb terület teljességre törekvő szókészletét tartalmazó kontrollált szótárakat a benne foglalt fogalmak közötti fogalmi összefüggések és relációk rendszerbe foglalásával testesítik meg az információs tezauruszok és olyan más tudásszervezési rendszerek, mint a taxonómiák, az ontológiák és a nyelvi-szemantikai értelemben vett névterek, illetve authority állományok.<sup>69</sup>

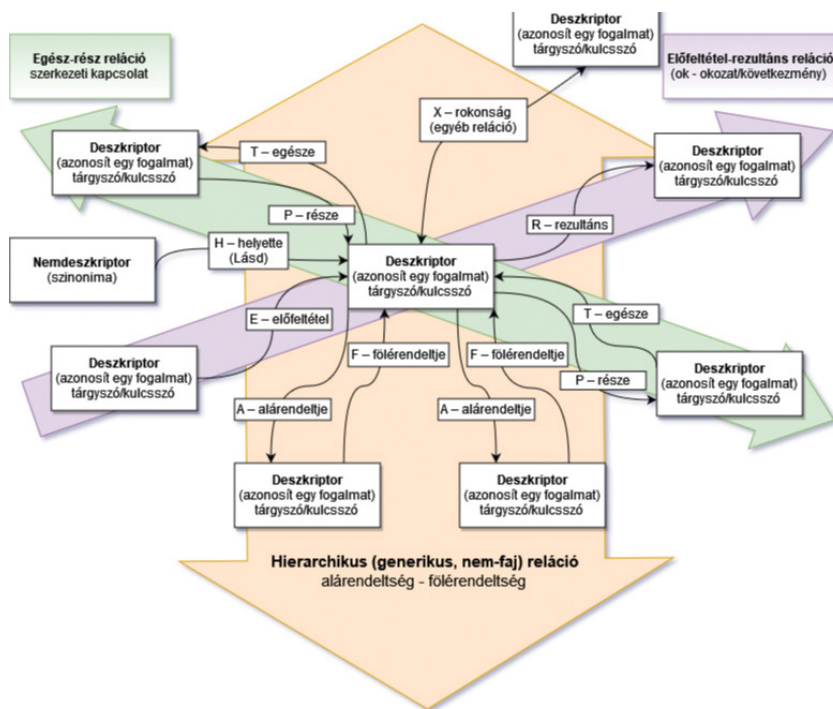


*Tartalomkeresés könyvtári környezetben*

<sup>68</sup> Ungváry Rudolf – Vajda Erik: *Az információkeresés szavai.* = Tudományos és Műszaki Tájékoztatás, 50. évf. 2003. 12. sz. 486–489. p. <https://tmt.omikk.bme.hu/tmt/article/view/2087> (2021.10.20.)

<sup>69</sup> Ungváry Rudolf: *A névterek értelme. Filozófiai-szerkezeti jellemzők.* = Tudományos és Műszaki Tájékoztatás, 65. évf. 2018. 1. sz. 1–14. p. <https://tmt.omikk.bme.hu/tmt/article/view/1720> (2021.10.20.)

A teauruszok a fogalmakat szabályozott kifejezésekkel, az úgynevezett deszkriptorokkal reprezentálják. A kapcsolódó szinonimákat a nemdeszkriptorok jelölik. A teaurusz hierarchikus és mellérendelő rendszer,<sup>70</sup> ezen felül képes része vagy egésze, rezultáns és előfeltétel, illetve rokonsági relációkat feltüntetni a fogalmak között,<sup>71</sup> így egy ilyen rendszer fogalmaival tárgyszavazott katalógusban a használók elméletileg jóval könnyebben tudnak keresni és navigálni. Kapcsolatad-környezetben különösen jól kamatoztatható ez a fogalmak közötti viszonyrendszer.



*Lehetséges kapcsolatok egy teaurusz fogalmai között*

<sup>70</sup> Egyszerre van jelen benne az alá-fölrendeltségi viszony és az azonos szinten lévő fogalmak használata.

<sup>71</sup> Horváth – Sütheő: *i. m.* 96–107. p.

A Bibliographic Control konferencián *Anna Lucarelli*, a Firenzei Nemzeti Központi Könyvtár (BNCF) osztályozási szakembere ismertette a tezauruszok szerepét a modern digitális környezetben. A Bázeli Egyetem BARTOC adatbázisa<sup>72</sup> ezeket – köztük a magyar közkönyvtárak által használt Köztauruszt – követi nyomon. A tezauruszok és az újabb tudásszervezési rendszerek közötti határvonal a közelmúltbeli szabványosításoknak (ISO 25964 szabványcsalád,<sup>73</sup> illetve az RDF/SKOS<sup>74</sup>) köszönhetően elmosódott, mivel megteremtődött az átjárhatóság a szabványos tezauruszok és más szótárak (például a szemantikus web tudásszervezési rendszerének számító SKOS) között. A szabványok ezen felül lehetővé teszik a különböző nyelvek közötti megfeleltethetőséget, így módon létrehozva többnyelvű tudásszervezési rendszereket. A más rendszerekkel kompatibilis, komplex fogalmi kapcsolatokat feltüntető tudásszervezési rendszerek sok tekintetben a szemantikus web „legjobb barátai”, mivel képesek RDF-formátumban megjeleníteni a metaadatokat, amelyek ezáltal könnyen újrafelhasználhatók. Az interneten elérhető nyílt kapcsolt adatok között sok kontrollált szótárt találunk, a konferencián bemutatott olasz *Nuovo soggettario* tezaurusz<sup>75</sup> is például már a SKOS része. A hasonló nyíltadat-szolgáltatásoknak és a szabványos RDF-formátumnak megfeleltetett, a nemzeti könyvtárak gondozásában álló szótárak (összeköttetésben különféle katalógusokkal) további nemzetközi kapcsolatok létesítéséhez képeznek alapot, így különböző nemzeti kontrollált szótárak fogalmai (és az ezekhez tartozó források) is összeköttetésben lehetnek egymással. Mindez azonban nagy kihívás: ugyanaz a fogalom más nyelven eltérő jelentéstartalommal bírhat. Közvetítőnyelvként használva ehhez nyújthat segítséget például az ETO, amely nyelvfüggetlen, mesterséges nyelvként közbeiktatható az eltérő nyelvek közötti kommunikációban,<sup>76</sup> így egy adott nyelv fogalmához párosított ETO-jelzet megfeleltethető egy másik nyelv ugyanazon jelzethez társított fogalmának.

<sup>72</sup> A BARTOC.org keresőfelületét lásd a következő oldalon: <https://bartoc.org/> (2021.10.20.)

<sup>73</sup> A NISO weboldala. *ISO 25964 – the international standard for thesauri and interoperability with other vocabularies*, lásd <http://www.niso.org/schemas/iso25964> (2021.10.20.)

<sup>74</sup> W3C weboldala. *Introduction to SKOS*. lásd <https://www.w3.org/2004/02/skos/intro> (2021.10.20.)

<sup>75</sup> BNCF *Nuovo soggettario* weboldalát lásd [https://thes.bncf.firenze.sbn.it/index\\_eng.html](https://thes.bncf.firenze.sbn.it/index_eng.html) (2021.10.20.)

<sup>76</sup> Dancs Szabolcs: *Ismét az ETO hasznáról – az új elektronikus kiadás előkészületeinek fényében*. = Könyv, Könyvtár, Könyvtáros, 25. évf. 2016. 4. sz. 4. p. [https://epa.oszk.hu/01300/01367/00277/pdf/EPA01367\\_3K\\_2016\\_04\\_003-008.pdf](https://epa.oszk.hu/01300/01367/00277/pdf/EPA01367_3K_2016_04_003-008.pdf) (2021.10.20.)



Szótár (kifejezéstár)	Szinonimaszótár	Tárgyszójegyzék	Taxonómia	Tezaurusz	Ontológia
Magyarzó-értelmező szöveg (többértelműség-kontroll)	Utalók (szinonimakontroll)	Magyarzó-értelmező szöveg (többértelműség-kontroll)	Magyarzó-értelmező szöveg (többértelműség-kontroll)	Magyarzó-értelmező szöveg (többértelműség-kontroll)	Magyarzó-értelmező szöveg (többértelműség-kontroll)
		Utalók (szinonimakontroll)	Utalók (szinonimakontroll)	Utalók (szinonimakontroll)	Tulajdonságok
			Hierarchikus kapcsolatok	Hierarchikus kapcsolatok	Szemantikus kapcsolatok
				Asszociatív kapcsolatok	Osztályok

*A kontrollált szótárak néhány alapvető típusának összehasonlítása, az egyszerűtől a bonyolultig  
(Heather Hedden ábrája alapján, hedden-information.com)<sup>77</sup>*

Az informatika fejlődésével párhuzamosan a tartalmi feltárás során is megmutatkozott az igény az automatizálásra, de miért fontos a gépesítés? *Elisabeth Mödden*, a Német Nemzeti Könyvtár (DNB) szakértője szerint elsősorban az online publikációk számának rohamos növekedése említhető meg indokként. *Osma Suominen*, a Finn Nemzeti Könyvtár rendszerspecialistája szerint a gépi tárgyszavazó rendszerek fejlesztésének fő indoka az, hogy a manuális tartalmi feltárás rendkívül idő- és munkaigényes feladat, különösen, ha nagyméretű gyűjtemények feldolgozásáról van szó. A DNB-nél a gépi osztályozást először csak egy szűk területen vezették be, orvostudományi publikációknak osztott ki a rendszer rövidített DDC-jelzetet<sup>78</sup> A tapasztalatokat kamatoztatva 2017-ben a gyakorlatot kiterjesztették a nemzeti bibliográfiában feltüntetett nyomtatott monográfiák osztályozásához, a jelenlegi munka célja pedig az, hogy a gépi rendszer teljes hosszúságú jelzeteket állítson elő a németek által használt Dewey-féle tizedes osztályozáshoz.

<sup>77</sup> Hedden, Heather: *What is a thesaurus and what is it good for.* = Hedden Information Management weboldal, 2020. november 22. (<https://www.hedden-information.com/what-is-a-thesaurus-and-what-is-it-good-for/>) (2021.10.20.)

<sup>78</sup> Dewey-féle tizedes osztályozás szerinti jelzet.

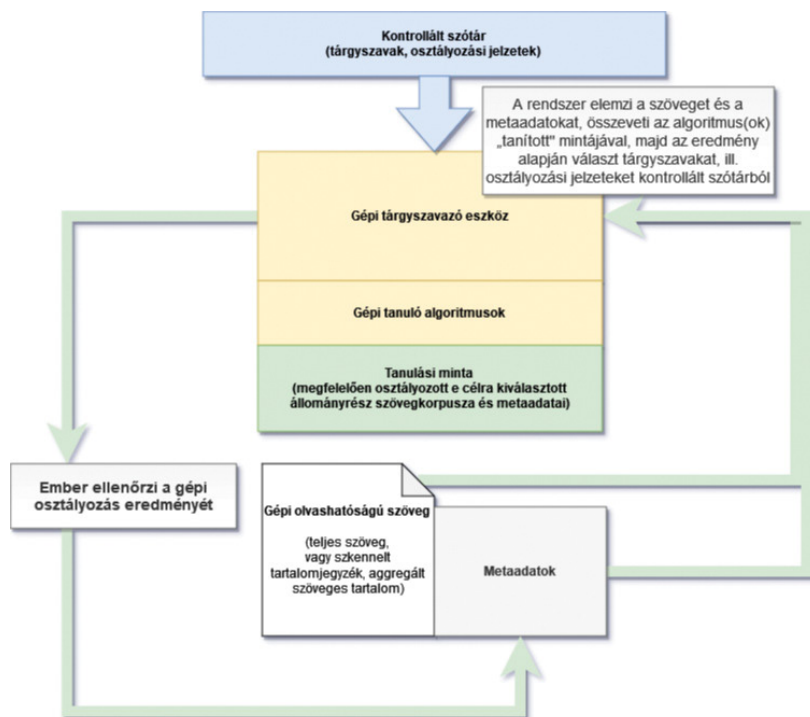


Milyen előfeltételek szükségesek a gépi osztályozáshoz, illetve a gépi tárgyszavazáshoz? Ehhez manapság elengedhetetlenek a bibliográfiai feldolgozás során nyert metaadatok mellett a gépi olvashatóságú szövegrészek, legyenek azok teljes szövegű források vagy akár nyomtatott dokumentumok digitalizált tartalomjegyzékei. Emellett kellenek kontrollált szótárak (pl. ETO- vagy DDC-mutatók adott jelzetekhez, tárgyszórendszerek, teauruszok, ontológiák) az osztályozási jelzetekhez és a tárgyszavakhoz, illetve egy – vagy akár több – gépi tanulást alkalmazó fejlett számítógépes algoritmus, amely elvégzi magát az automatikus jelzetelést és tárgyszókiosztást.

A német gyakorlatban a gépi tartalmi feltárás során a szolgáltatás visszakeresi a könyvtár katalógusában az adott dokumentum már létező metaadatait, illetve a repozitóriumban hozzákeresi a feltárandó dokumentum teljes szövegét vagy a digitalizált tartalomjegyzékét. A szoftver természetes nyelvű szövegfeldolgozásért felelős algoritmus a nyelvészeti és statisztikai módszerek alapján elemzi a fenti információforrásokból nyert szöveget, majd a feladatra „betanított” algoritmus az így kapott eredményhez kiválasztja a legmegfelelőbb osztályozási jelzetet és a tárgyszavakat. Utóbbiak forrása kontrollált szótárként a német integrált authority állomány (Gemeinsame Normdatei – GND<sup>79</sup>). A MARC 21-gyel kompatibilis tárgyszavak és osztályozási jelzetek gépi kiosztását könyvtáros szakemberek felügyelik, és szükség esetén javítják. A gépi tanulás lényege, hogy az algoritmust „betanítják” egy erre a célra kijelölt állományrész adatainak a segítségével a tárgyszavazásra. A „tanulás” szakértők által javított eredménye szolgál mintaként a majdani gépi osztályozási-tárgyszavazási munkafolyamat során. A feltárandó szövegrészek tartalmi ismérveit az algoritmus egy F-score-nak nevezett érték szerint kategorizálja és felelteti meg a tanult minta alapján az egyes tárgyi kategóriáknak. A módszerből következik, hogy minél részletesebb az algoritmus képzési modellje, annál pontosabb lesz az eredmény. Így nem véletlen, hogy elsősorban jól körülhatárolható szűk szakmai területen működik kiválóan a gépi tárgyszavazás. Tágabb területen, pláne szépirodalmi művek feltárásánál kevésbé hatékony, mivel a gépi tanuláshoz nehéz olyan megfelelő elemszámú képzési anyagot összeállítani, amely alapján pontosan ki tudja választani az algoritmus a megfelelő tárgyszavakat. A német szakember további kihívásnak tartja a témát illetően, hogy a könyvtárszakmában nincs még egységes minőségbiztosítási szabvány e területen. A német szakemberek a munkát nemzetközi együttműködés keretében kívánják folytatni, ezért alapos szakmai előkészítést követően a finn Annifhoz történő csatlakozás mellett döntöttek.

<sup>79</sup> A Deutsche Nationalbibliothek honlapja. *The Integrated Authority File (GND)*, lásd [https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd\\_node.html](https://www.dnb.de/EN/Professionell/Standardisierung/GND/gnd_node.html) (2021.10.20.)





*A gépi osztályozás/ tárgyszavazás működésének egyszerűsített ábrája*

A többnyelvű Annif gépi tárgyszavazó eszközt<sup>80</sup> *Osma Suominen* mutatta be. A közösségorientált módon fejlesztett nyílt forráskódú Annif metaadatainak bázisát a *finna.fi*<sup>81</sup> közgyűjteményi aggregátorportál képezi. A gépi tanuló algoritmusok képzésére és tesztelésére ennek a metaadatait, valamint a *Kérdezd a könyvtárost!* online szolgáltatás beérkező kérdéseit és az azokra adott válaszokat, a Jyväskylä Egyetem szakdolgozatait és disszertációit, kiadói könyvleírásokat, valamint a nemzeti könyvtár e-könyveit használták fel. Az algoritmus által a tanulási minta nyomán javasolt tárgyszavak pontosságát a könyvtárosok manuálisan adott tárgyszavaival összevetve tesztelték. Az eszköz már napi szintű használatban van a Jyväskylä Egyetem *JYX* nevű repozitóriumában,<sup>82</sup> ahol a rendszer a hallgatók által feltöltött szakdolgozatokhoz és disszertációkhoz javasol tárgyszavakat, így a szerzők az üres mezők kitöltése helyett a gépi javaslat listájából választhatják ki a megfelelőnek ítélt tárgyszavakat a saját maguk által megadott kulcsszavak

<sup>80</sup> Az Annif kezdőoldala: <https://annif.org/> (2021.10.20.)

<sup>81</sup> A Finna.fi kezdőoldal: <https://finna.fi/> (2021.10.20.)

<sup>82</sup> A Jyväskylä University Digital Repository: <https://jyx.jyu.fi/> (2021.10.20.)

mellé. A projekt újítása a *Finto AI* gépi tárgyszavazó eszköz és alkalmazásprogramozási felület (API).<sup>83</sup> Ezen a webfelületen a rendszer automatikus szövegelemzést követően a tetszőlegesen beillesztett angol, finn és svéd nyelvű szöveg tárgyszavazását végzi el a finn nemzeti ontológia (*YSO*) teauruszát<sup>84</sup> használva. A Finn Nemzeti Könyvtár a gyakorlatban is alkalmazza már a *Finto AI*-t: az elektronikus tárhelyek építése során állítanak elő a segítségével tárgyszójavaslatokat, de együttműködnek egy könyvesboltokat és könyvtárakat ellátó vállalattal is. Ebben az esetben az *Annif/Finto AI* a kiadóktól kapott leírásokat elemezve javasol tárgyszavakat.

A gépi osztályozás/tárgyszavazás igen hasznos segítség, és – bár gyorsan fejlődik – a technológia még gyerekcipőben jár. A gépi tanuló algoritmusok emberi szemnek gyakran banális dolgokat értelmeznek félre. Azonban meg kell említeni, hogy a tartalmi feltárás a viszonylagos szabályozottság (kontrollált szótárak, házi szabályzatok) ellenére jelentősen függ a feldolgozó szubjektivitásától, ugyanis két könyvtáros a szabályos munkafolyamatokat követve is tárgyszavazhatja másként ugyanazt a dokumentumot. A gépi feltárás legnagyobb előnye, hogy a tárgyszavak kiosztása megfelel a „tanult” mintának, így egy automatikus rendszer által osztályozott/tárgyszavazott állományrész jóval konzisztensebb feltárást tesz lehetővé, és ami nagyon fontos, a hozzáértő könyvtáros felügyelete mellett teszi. Az *Annif/Finto*hoz hasonló kezdeményezések (pláne ha nyílt forráskóddal, nemzetközi együttműködés keretében fejlesztik) igazi előrelépésnek tekinthetők a tartalmi feltárás fejlődésében, együttműködésben más könyvtárszakmai fejlesztésekkel. Gondoljunk csak a szövegesen kereshető elektronikus könyvtárak, repozitóriumok egyre bővülő választékára és a különböző forrásokat aggregáló keresőfelületek térhódítására. Ezek a gyűjtemények a szöveges keresőeszközök fejlesztésével a webes környezethez szokott könyvtárhasználó számára is lehetővé teszik az egyszerű, több szempontú visszakeresést: a weben megszokott egymezős kereséssel kutathat az aggregált, teljes szövegű tartalmakban. A könyvtárak bevált formai és tartalmi feltáró eszközeinek köszönhetően a tapasztalt használóknak alternatívaként továbbra is rendelkezésre áll az ennél jóval komplexebb információvisszakeresés lehetősége. Ahhoz, hogy a felhasználók valóban ki is tudják használni az elérhető minőségi forrásokat, a könyvtáraknak ki kell törniük a mai „elszigetelt” létállapotból, és be kell vezetniük ezeket a korszerű tartalomfeldolgozó újításokat, amelyek keresőfelülete a jelenleginél könnyebb használhatóságot és egyértelműbb kezelőfelületet kínál.

**Bódog András**

<sup>83</sup> A *Finto AI* felület: <https://ai.finto.fi/?locale=en> (2021.10.20.)

<sup>84</sup> *Finto* honlap. *Finnish Tesaurs and Ontology Service*. Lásd <https://finto.fi/ysa/en/> (2021.10.20.)

# Megjelent a nyílt archívumi információs rendszer (OAIS) szabványának magyar nyelvű változata

A digitalizáció és ezzel párhuzamosan az elektronikus dokumentumok világában elengedhetetlenül fontos feladattá vált a folyamatosan termelt gigászi mennyiségű elektronikus és fizikai információ hosszú távú, tartós megőrzése. Az eredetileg űrkutatási adatok tárolására szánt nyílt archívumi információs rendszer (OAIS – Open Archive Information System) szabványa egy olyan elméleti referenciamodell és keretrendszert ír le, amely alapján a gyakorlatban is létrehozhatók e tevékenységre szánt archívumok. A szabvány meghatározza mindazon fogalmakat is, amelyekre a nem archiválással foglalkozó szervezeteknek is szükségük van ahhoz, hogy hatékony résztvevői legyenek a megőrzési folyamatnak. Az OAIS-ban használt 'nyílt' kifejezés arra utal, hogy ez a szabvány, illetve az ehhez kapcsolódó releváns szabványok fejlesztése nyílt fórumokon történik. A referenciamodell az aktuális technológia korlátozta implementációktól függetlenül ad útmutatást a hosszú távú megőrzés elméletéhez.

A „hosszú táv” olyan pontosan meg nem határozott időtartam, amely alatt jelentős technológiai változás történik, adatformátumok, sőt műszaki megoldások szűnnek meg és újak jelennek meg, gondoljunk csak a kazetták és optikai lemezek gyors elterjedésére majd eltűnésére néhány évtized alatt, illetve a különböző fájlformátumok gyors változására. A referenciamodellben meghatározott OAIS-archívumnak ilyen változó környezetben kell megőriznie és a célközönsége számára szolgáltatnia a digitális vagy fizikai információt.

Az eredetileg az ezredforduló tájékán megjelent űrkutatási területre szánt ajánlás 2012-ben változtatások nélkül vált ISO-szabvánnyá. Mivel az Országos Széchényi Könyvtár (OSZK) egyik missziója a hosszú távú megőrzés, és a Könyvtári Szabványosítási Iroda feladata az OSZK és a hazai könyvtárszakma szabványügyi támogatása, ezért az Országos Könyvtári Szabványosítási Bizottság feladatául tűzte ki e fontos szabvány magyar nyelvre történő átültetését. Az MSZ ISO 14721:2022 szabványt a Magyar Szabványügyi Testület (MSZT) MB 508 Információ és dokumentáció nemzeti szabványosító bizottsága honosította.

A honosítás azzal a nem titkolt céllal történt, hogy egységes terminológiai kiindulópontként szolgáljon a hazai közgyűjtemények hosszú távú megőrzést érintő tevékenységében. A magyar nyelvű szabvány nemzeti előszavában az MSZT értelmezési segédlettel szolgál a szabványalkalmazás megkönnyítésére. A szabvány – az összes többi MSZ-szabványhoz hasonlóan – elérhető az MSZT szabványboltjában, illetve a mintapéldány kölcsönözhető az OSZK-n belül működő Könyvtártudományi Szakkönyvtár gyűjteményéből.



Nemzeti Kulturális Alap