

A MARC21 tételfejének és kódolt tartalmi jellemzőinek feldolgozási minősége néhány nemzeti könyvtárban

Egy elemzés tanulságai

Ungváry Rudolf, Király Péter

Előző munkánkban összevetettük a nagyobb nemzeti könyvtárak MARC21 állományában a tartalmi adatmezők (648, 65X) feldolgozottságát. E tanulmányban a kötelezően kitöltendő tételfej és bizonyos kódolt tartalmi jellemző (007 és 008) feldolgozottságával foglalkozunk. Ezekben egy karakteres alfanumerikus jelzetekből („kódkból”) álló enumeratív osztályozási rendszerek szerint kell, illetve kellene minősíteni a forrásokat. Kiderült, hogy az utóbbi két mező pozícióinak használata rendkívül szegényes. Ez egyrészt összefügg a nem kötelező voltukkal, másrészt a MARC vonatkozó osztályozási rendszereinek hiányosságaival, ill. azzal, hogy más adatmezőkhöz képes (megjegyzés mezők és tartalmi adatmezők) tapasztalható redundanciákkal, ami miatt a 00X mezők feldolgozását elhanyagolják. Holott egyrészt kitöltésük az információkeresés szempontjából rendkívül hasznos lenne – ha a könyvtári rendszerek keresőit felkészítenék a keresésben a felhasználóbarát lehetőségekre.

Tárgyszavak: könyvtári katalógus, katalogizálás, szabályzat, osztályozáselmélet, könyvtárügy, szabvány

2020-tól szisztematikusan összevetettük a jelentősebb tudományos és nemzeti könyvtárak katalógusaiban a MARC21-állományok feldolgozását. Ehhez a munkához letöltöttük a vizsgált könyvtárak katalógusainak állományát, elemzésükhöz pedig a QA catalogue nevű, a MARC-rekordok minőségellenőrzését végző alkalmazást használtuk [2][3][4]. Az alkalmazással a MARC21-állományt a számítógépes elemzés eddig nem alkalmazott, új módszerrel tártuk fel a további, tartalmi vizsgálat céljaira. Így olyan adatok tömege vált hozzáférhetővé, melyek alapján elemezhető a katalógusok MARC21 mezőinek kitöltése, e kitöltés munkájának tartalmi jellege. Mindez jelentős segítség lehet a katalógusok minőséget meghatározó strukturális jellemzők megértéséhez és a hibák feltárásához.

2020-ban ennek a feltárásnak a *tartalomra vonatkozó* adatelemeit (045, 072, 043, 045, 052, 072, 080, 082, 084, 085, 505, 520, 6XX) és a velük összefüggő adatelemeket (007, 008, 034, 041) elemeztük [7][8][9] néhány tudományos könyvtárban és a Német Nemzeti Könyvtárban.

A mostani elemzésben a MARC21-katalógusok egy, a szakmai közvélemény előtt lényegében homályban maradó részével foglalkozunk. Ennek a résznek a jellemzőit csak az egyes könyvtárak speciális feldolgozói ismerik igazán, de nekik sincs rálátásuk más nagy könyvtárak hasonló adatainak kitöltési elveire.

Az elemzett rész a rekordfej (06, 07 és 17), továbbá a vezérlő mezőkre (kódolt fizikai jellemzők, 00X) korlátozódik. Az utóbbiak közül is csak a kódolt fizikai jellemzők (007) és a meghatározott jellemzők és információk adatok (008) *tartalomra vonatkozóan fontosabbnak tekintett* adatelemeit vizsgáltuk. Ebben a tanulmányban az egyes nemzeti könyvtárak katalógusaiban a kiválasztott pozíciókon szereplő értékeket hasonlítottuk össze. Az eredményt a gyors áttekinthetőség érdekében vizuálisan buborékdiagrammokkal szemléltettük.

A szóban forgó adatelemek mesterséges nyelvű, enumeratív osztályozási rendszerek egyetlen karakteres értékből álló jelzetei (kódszavai, a szakmai szlengben kódjai). Velük osztályozva a dokumentu-

1. táblázat Az elemzésbe bevont MARC21 mezők és pozícióik

mező	pozíció	elnevezés
rekordfej	06	a dokumentum jellege (=a rekord típusa)
	07	bibliográfiai szint
	17	a leírás jellege
007	00	kódolt fizikai jellemzők, dokumentumkategória
	01	a dokumentum speciális megjelölése: szöveg
	01	a dokumentum speciális megjelölése: térkép
008	18–21	könyvek, illusztráltság
	25–27	könyvek, tartalmi jellemzők
	33	könyvek, műfaj

mok a legkülönbébb feldolgozási mód és dokumentumtípus szerint osztályozhatók. Például: a rekordfej 07. pozícióján azt osztályozták, hogy analitikusan, gyűjteményesen, kötetként, monografikusan stb. dolgozták föl. A rekordfej 17. pozícióján azt, hogy autopsziával, rövidítetten, előzetesen stb. A rekordfej 06. pozícióján a dokumentum fajtáját osztályozták a külső fizikai megjelenés szempontjából: nyomtatott, kéziratos, audiovizuális stb. A 007 adatmező 00 pozícióján a dokumentum fajtáját a vizuális megjelenés szempontjából osztályozták aszerint, hogy szöveg, térkép, grafika stb. (Lásd példaként a 2. és 3. táblázat pozícióhoz kötött osztályait.)

Megvizsgáltuk, milyen összhang áll fenn e mezők között. Például vannak-e nyilvánvaló, vagy azoknak látszó ellentmondások; teszem azt egy mű nyomtatott nyelvi dokumentum (rekordfej/06=a), ezzel szemben a kódolt fizikai jellemzők szerint nem kivethető grafikákhoz tartozó kollázs (007/01=c)? Vagy van-e olyan kézirat, melyet kötetként/részegységként osztályoztak? Ugyancsak kiderül az elemzésből, hogy vannak-e hibás pozícióértékek. Ezeket a diagramokban az „[érvénytelej] kategóriába soroltuk és az előforduló értékeket külön táblázatba foglaltuk össze. Ilyen nagy tömegű összehasonlító elemzésekre megfelelő informatikai eszközök hiányában eddig nem volt lehetőség

A rekordfej és a vezérlő mezők adatai hasznos, és egyszerű kereshetőségeket biztosíthatnának olyan esetekben, amikor a keresésbe a dokumentumok egyes típusait is bevonják. Ezeknek az adatmezőket a mai könyvtári keresőrendszerek mosto-

hán kezelik. Gondoljuk csak meg, hogy mennyire nem egyszerű manapság egy olyan keresést megvalósítani, melynek tárgya mindazon elbeszélések, melyek illusztráltak?

A tételfej és a kódolt fizikai jellemzők osztályait a jelentőségükhöz képest a végfelhasználók csak nagyon nehézkesen tudják használni. Ezért eleve lemondanak a használatukról. Munkánk célja, hogy néhány elérhető nemzeti könyvtár vonatkozó állománya esetén a felszínre kerüljenek eme adatmezők kitöltésének jellegzetességei, és az esetleg belőlük fakadó problémák, anomáliák. Végső célunk, hogy mindezzel empirikus adatok álljanak rendelkezésre mind a katalógusok, mind a kezelőrendszerek, mind pedig nem utolsósorban a MARC21 fejlesztéshez.

Már itt utalunk rá, hogy a rekordfej pozícióinak kitöltése kötelező, a vezérlő mezőké nem. Ennek messze ható következményei vannak. Ezeket a tanulmány végén összegeztük.

A vizsgálatba bevont nemzeti könyvtárak a buborékdíagramokban megjelenő sorrendben az alábbiak:

- LC Kongresszusi Könyvtár¹
<http://ddb.qa-catalogue.eu/loc/>
- DNB Német Digitális Könyvtár²
<http://ddb.qa-catalogue.eu/dnb/>
- ÖNB Osztrák Nemzeti Könyvtár³
<http://ddb.qa-catalogue.eu/onb/>
- NFI Finn Nemzeti Könyvtár⁴
<http://ddb.qa-catalogue.eu/nfi/>, Fennica adatbázis (a nemzeti bibliográfia)
- libris Svéd Nemzeti Könyvtár és közös katalógus⁵
<http://ddb.qa-catalogue.eu/libris/>
- BNPL Lengyel Nemzeti Könyvtár⁶
<http://ddb.qa-catalogue.eu/bnpl/>
- LNB Lett Nemzeti Könyvtár⁷
<http://ddb.qa-catalogue.eu/lnb/> (csak a 2017-2020 közötti állomány)

1 Letölthető a <https://www.loc.gov/cds/products/marcDist.php> címről.

2 Az adatokat kutatási céllal a könyvtártól kaptuk meg.

3 Letölthető OAI-PMH-n keresztül a https://eu02.alma.exlibris-group.com/view/oai/43ACC_ONB/request címről.

4 Letölthető a <http://data.nationallibrary.fi/download/> címről.

5 Letölthető SPARQL (<https://libris.kb.se/sparql>) és az Xsearch API-k (<http://libris.kb.se/xsearch>) kombinálásával.

6 Letölthető a <https://data.bn.org.pl/databases> címről.

7 Letölthető a <https://dati.lnb.lv/> címről.

- KBR Belga Nemzeti Könyvtár⁸
<http://ddb.qa-catalogue.eu/kbr/>
 BL Brit Nemzeti Könyvtár⁹
<http://141.5.102.114/bl/>

A tanulmányban mezőnevek dolgában a HUN-MARC 2002. március magyar KSZ/4.1 kiadás terminológiáját követjük [1]. Ha a MARC21-ben [5] eltérés van, vagy az almező új, akkor az angol kifejezés fordítását adjuk meg.¹⁰

1 Rekordtípusok és bibliográfiai szint

Jól látható, hogy az összehasonlításban a bibliográfiai szint egyes típusai elsősorban a nyelvi anyagra jellemzők. Figyelemre méltó a kétdimenziós nem kivetíthető és a zenei hangzó anyagok esete. Ezek között van összefoglaló, alárendelt (kötet, részegység) és sok monografikus szintű feldolgozás. A kétdimenziós anyagok esetében az LC-ben, az osztrák (ÖNB), a lengyel (BNPL), a belga (KBR) és a brit (BL) nemzeti könyvtárakban túlnyomórészt fényképekről van szó. A zenei hangzó anyagok pedig többnyire könnyűzenei szórakoztató dalok, és az LC-re meg DNB-re jellemzők. Hasznos volna pontosabban tudni, mit jelent például kéziratos és nyomtatott zenemű esetében a monografikus, analitikus, alárendelt és integrált forrású bibliográfiai szint? Ugyanezt a kérdést a kartográfiai anyagok, ill. a háromdimenziós ábrázolások (pl. plakátok) esetén is feltehetjük. Mit tartalmaznak a jelentősebb nemzeti könyvtárak belső feldolgozási szabályzatai ezeknek a dokumentumoknak az esetében? Például milyen alapértelme-

zéseket? A rajtuk alapuló gyakorlat szakmai nyilvánosságának hiányában a probléma átfogó, konszenzusos megoldása nagyon korlátozott.

Előfordulnak más figyelemre méltó esetek. Például az LC-ben összefoglaló (időszaki) (rekordfej/07=s) a besorolása egy fényképgyűjteménynek (rekordfej/06 k) csak azért, mert 1985 óta gyűjtik benne online hozzáférhetően a képeket.¹¹ Attól tehát, hogy az egyes fényképek különböző időkben készültek, illetve kerültek az adatbázisba, lett a gyűjtemény időszaki? De akkor egy olyan cikkgyűjtemény-adatbázis, melybe különböző időben megjelent cikkek különböző időkben folyamatosan kerülnek be, szintén időszaki volna? Kérdés, hogy az időszaki kiadványok jelenlegi meghatározása mennyire teszi ezt lehetővé? A német (DNB) nemzeti könyvtárban például nem létezik kétdimenziós (nem kivetíthető) (azaz fénykép) mint időszaki, az osztrákban igen.

Sok más eset is van; álljon itt befejezésül, hogy az osztrák (ÖNB) nemzeti könyvtárban alárendelt (Rf/07=d) a bibliográfiai szintje például három plakátrekordnak, de nincs megadva a teljes gazdadokeletum.

2 Rekordtípusok és a leírás jellege (szintje)

Feltűnő, hogy milyen kicsi a számítógép által kezelt állomány (RF/06=m). Holott az ilyen források elterjedtsége már rendkívül nagy. Erre utal, hogy az 5. *diagramcsoportban* a német (DNB) nemzeti könyvtárban mennyire nagy az elektronikus minősítésű (007/00=c) dokumentumok aránya. Hogy a fenti 2. *diagramcsoport* esetén rekordtípusként mégis annyira kevés, annak fő oka bizonyára a MARC21-formátum szabályozása. Ha ugyanis van egy jelentős szempont, amely miatt a forrás egy másik rekordtípusba is tartozik, akkor az m kód helyett az adott lényeges szempont rekordtípusával kell osztályozni. Így kell például eljárni, ha az állomány például nyelvi anyag, grafika, kartográfiai anyag, hangzó anyag, mozgókép. Más szóval ilyen

8 Letölthető OAI-PMH és Z39.50 API-k kombinálásával. Az API-k végpontjait a könyvtár szívességéből kutatási céllal használhattuk.

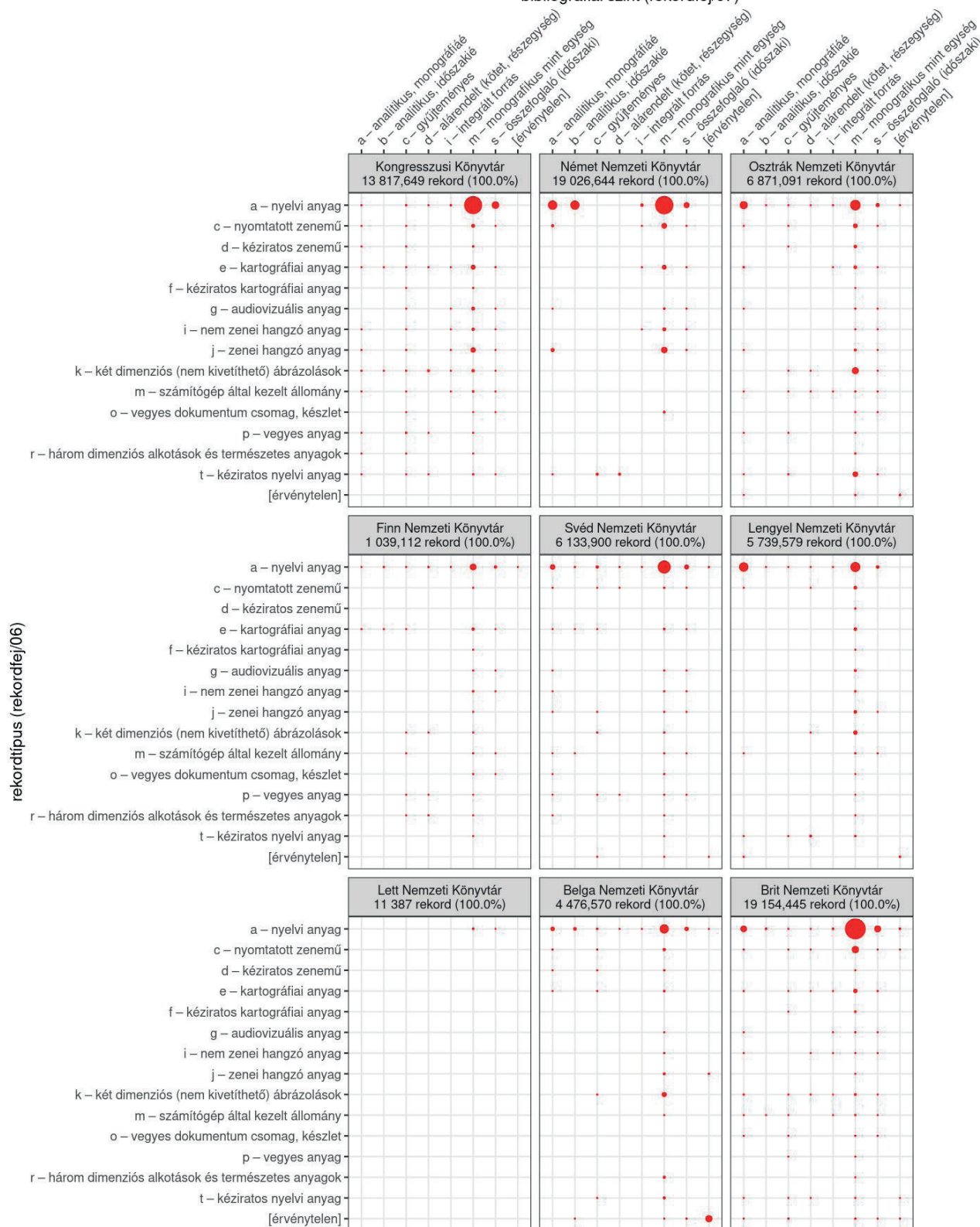
9 Az adatokat kutatási céllal a könyvtártól kaptuk meg.

10 A MARC21 2018-as angol változatról készült egy „MARC 21: a bibliográfiai rekordok adatcsere-formátuma–MUNKA-ANYAG”. című fordítás, mely csak az adatmezők, az almezők és indikátorok fordítását tartalmazza (ford. Kasza Zsófia.) < http://www.oszk.hu/sites/default/files/MARC21_bibliografiai_rekordok_adatcsere_formatuma.pdf >, és nem KSZ kiadás. A kéziratban nincs jelezve, hogy lektorálták volna, a fordításhoz egyáltalán nem vették figyelembe a már meglévő, 2002-es KSZ/4.1 fordítási megoldásait, a kézirat nyilvános vitájából kizárták azokat, akik mindezt megpróbálták szóvá tenni, a mező- és almezőnevek fordítása mereven követi az angol kifejezéseket. A benne alkalmazott magyar terminológiai problémáikról lásd [15].

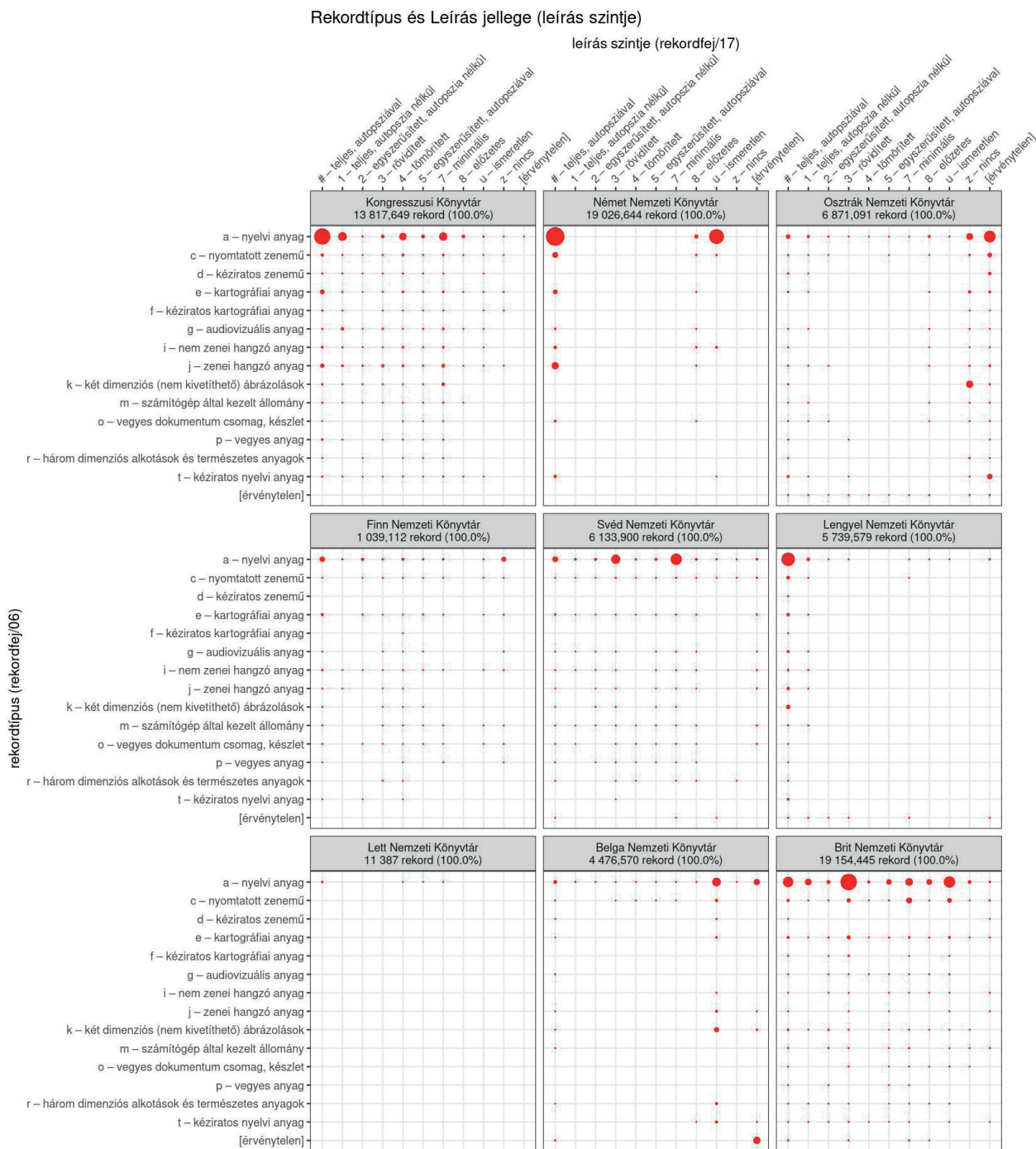
11 „Collection of digitized photographs of buildings contributed by SBSE members, formerly available as a subset of slides under the SBSE Slide Exchange Program. Slides date from October 1985 to the present, and are intended to be used as a teaching resource by building science educators. Includes image indexes and author annotations accompanying the slides.”

Rekordtípus és Bibliográfiai szint

bibliográfiai szint (rekordfej/07)



1. diagram Rekordtípusok és bibliográfiai szint

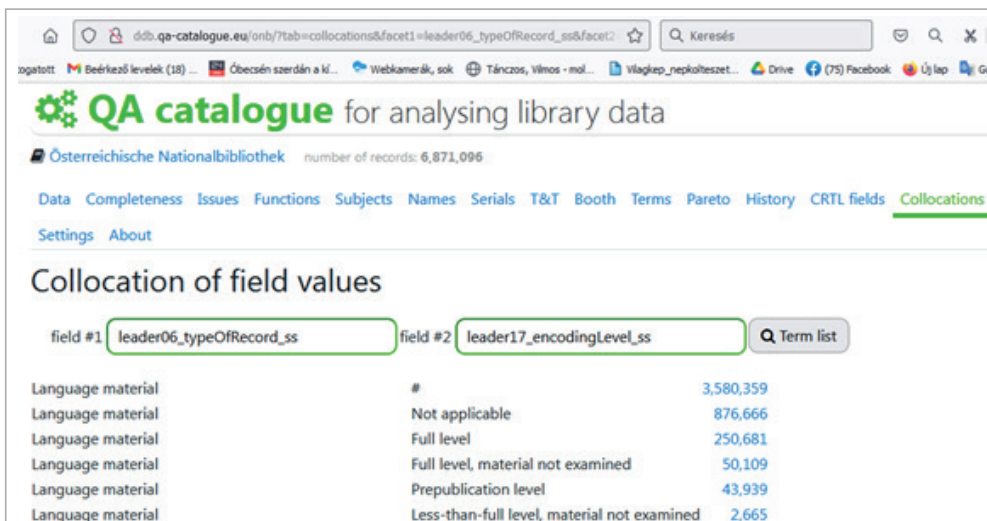


2. diagram Rekordtípusok és a leírás jellege (szintje)

esetekben nem lehet több szempont szerint osztályozni, ami az egy dimenziós osztályozási rendszerek alapvető hátránya.

A Kongresszusi (LC) könyvtárban, a brit (BL) és a svéd (libris) nemzeti könyvtárakban minden jel szerint változatosan, a különféle leírasi szinteket figyelembe véve osztályoznak. A német (DNB), osztrák (ÖNB) és lengyel (BNPL) nemzeti könyvtárakban feltehetően az alapértelmezések az uralkodók.

Figyelemre méltó, hogy az osztrák (ÖNB) nemzeti könyvtárban a leírás szintje szempontjából a rekordtípusok nagy része a nem megengedhető kettős keresztet (#) tartalmazza (358036 rekord), ezért az érvénytelen minősítést kapta (1. ábra). Félreértésre adhatott okot, hogy a MARC-szabványban a # jelet a szóköz jelölésére használja. Ezt a tényt csak a szabvány bevezetőjében közlik, az egyes adatelemek leírásakor nem. Lehetséges volna, hogy



1. ábra Az Osztrák Nemzeti Könyvtár (ÖNB) # értékű rekordfej 017 pozíciói

Bibliográfiai szint és Leírás jellege (leírás szintje)



3. diagram Bibliográfiai szintek leírasi szintjei

a feldolgozók úgy értelmezték, hogy a szabványban a kettős kereszt (#) nem szóközt (tehát Rf/17=_), hanem valóban kettős keresztet jelent? Vagy lehetséges, hogy a # jelét a kitöltési szabályzatban írják elő? Ez nyilvánosan hozzáférhetetlen.

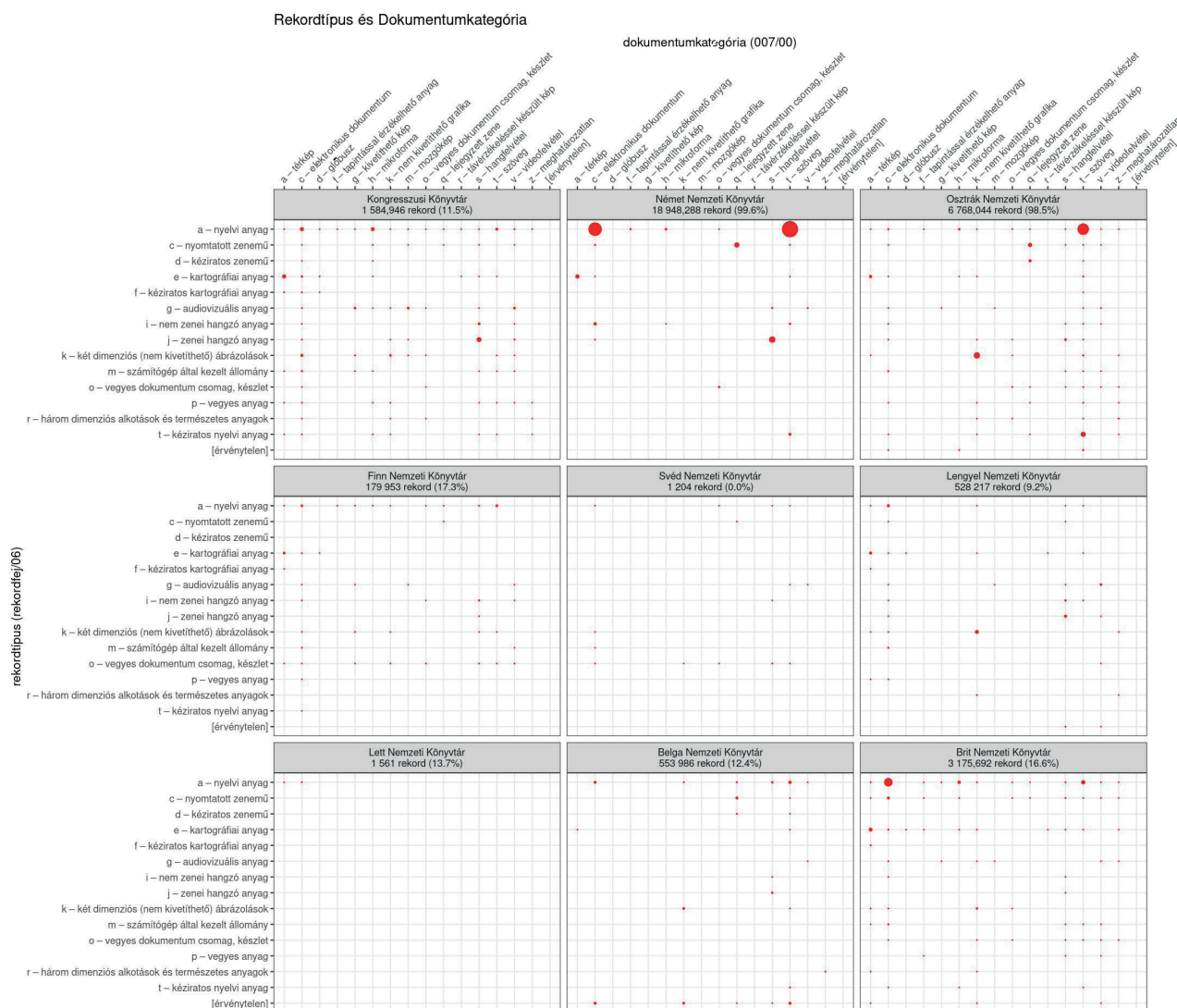
Az ismeretlen minősítés is gyakori. Mindez visszaköszön a későbbi, 3. buborékdiagram-csoportban is.

A belga (KBR) és a brit (BL) nemzeti könyvtárban a rekordok egy részét a leírás szintjét illetően ismeretlennek (Rf/17=u) nyilvánították. A brit nemzeti könyvtárban sok a rövidített leírási szint is (Rf/17=3). A svéd (libris) nemzeti könyvtárban a rekordok többségének leírási szintje minimális (Rf/17=u) és rövidített (Rf/17=3). Feltehetően léteznek mögöttük különféle szabályzatok, melyek ezt előírják.

A leírás szintje szempontjából szinte minden könyvtárban jelentős a nyelvi anyag feldolgozottsága, amint az várható volt. A kongresszusi (LC) és a német nemzeti könyvtárban (DNB) kitűnik még a zenei hangzó anyag is.

Az angol nyelvű és inkább brit orientációjú könyvtárakban észrevehető inkább, hogy a többi (nem szöveges nyelvi) rekordtípusban is megjelenik a leírás szintjének minősítése.

A legkiegyensúlyozottabb helyzet az LC-ben van. Ott feltűnő a zenei anyag és a fényképek minimális jelenléte. A zenei anyagok és a grafikák vonatkozásában hasonló a helyzet az osztrák (ÖNB) nemzeti könyvtárban is. A legegyszerűbb igényeket kielégítő rövid zenék (slágerek, népszerű fogyasztásra



4. diagram Rekordtípus és dokumentumkategória

készült dalok stb.), illetve rajzos, többnyire történelmi eseményeket vagy helyeket ábrázoló grafikák minősítése a „nincs” (rekordfej/17=z). Más könyvtárakban az ilyen dokumentumokat talán nem dolgozzák föl egyáltalán, és ezért nincs nyomuk. Holott legalábbis a fényképeknek idővel többnyire történelmi értéke lesz.

3 Bibliográfiai szintek leírás szintjei

Nem meglepő, hogy a monografikus egységek leírási szintje (Rf/07=m) a legtöbb nemzeti könyvtárban következetesen a teljes leírás (Rf/17=szóköz, ill. =1). Néhányban viszont, mint a svéd (libris), belga (KBR) és brit (BL), a leírás szintje jelentősebb számban minimális vagy rövidített (007/17=7, ill. =3). Feltűnő, hogy a német (DNB), az osztrák (ÖNB), a brit (BL) és a belga (KBR) nemzeti könyvtárakban az analitikus kiadványokat és sok monografikus egységet, mint „ismeretlent” (Rf/17=u) minősítettek. Ezekben a könyvtárakban talán inkább alapértelmezésekkel elégedtek meg. A svéd (libris) feldolgozás esetén ehhez társul a minimális (Rf/17=7) minősítés is. Kérdés, hogy ez minek tulajdonítható? Belső szabályzatoknak?

Ahogy az előző, 2. diagramcsoportban, úgy itt is különösen feltűnő, hogy az osztrák (ÖNB) nemzeti könyvtárban a bibliográfiai leírás szintje ugyan csak nagy számban „érvénytelen” (Rf/17=#), hiszen a # explicit használata nem megengedett. Mindez megint csak vagy félreértésre, vagy arra utal, hogy egy következetes, és talán belső szabályzatban dokumentált esetről van szó.

Összefoglalóan a három diagramcsoportról elmondható, hogy *ha az összehasonlításba csak a rekordfej pozíciói vannak bevonva, akkor a feldolgozottság 100 %-os.* Nem véletlenül, mivel a rekordfej minden pozíciójának kötelező értéket adni. A 00X vezérlő mezők esetében, ahogy ezt a megvizsgált 007-es és 008-as mezők mutatják, ez már koránt sincsen így, ami számos, ma még nem eléggé közismert problémát vet föl.

4 Rekordtípus és dokumentumkategória (Rf/06 vs. 007/00)

A diagramcsoport (ahogy a következő is) ékesen példázza, hogy mekkora eltérések vannak az egyes nemzeti könyvtárak között a 007-es mező kitöltésében.

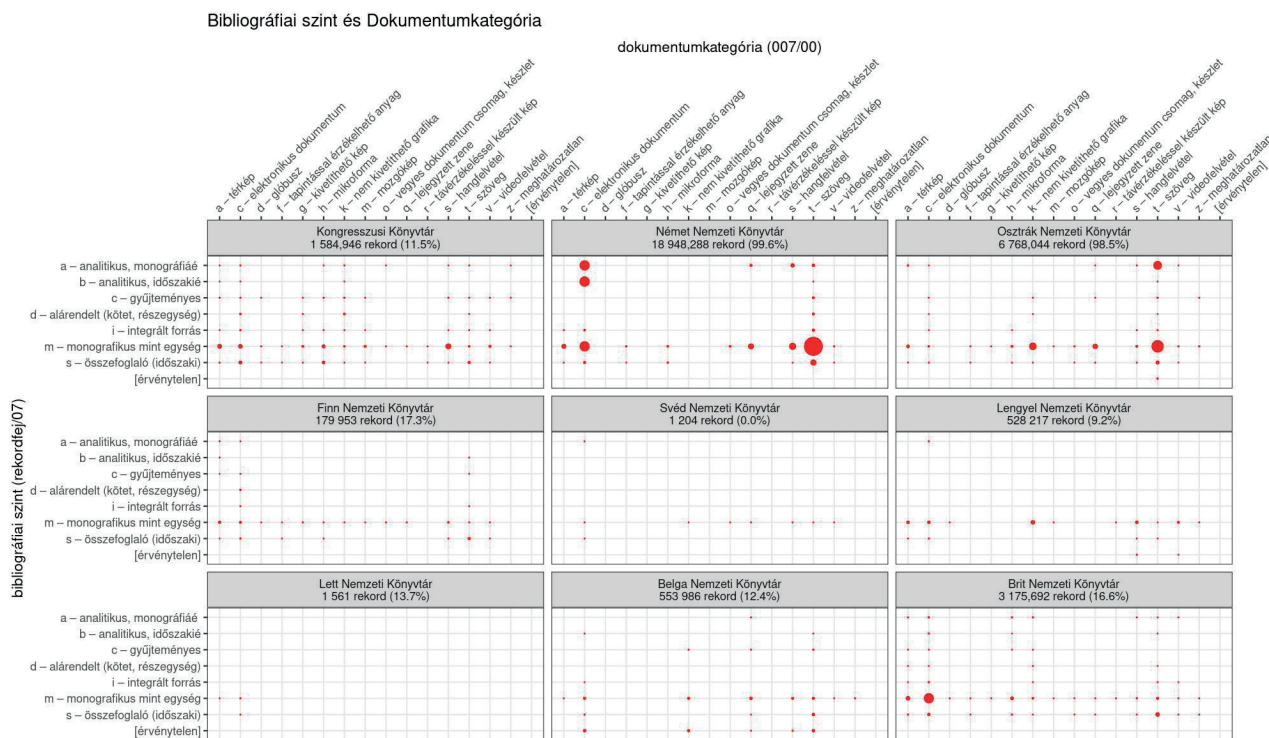
A legfeltűnőbb, hogy csak a két német nyelvű nemzeti könyvtárban (DNB és ÖNB) közel száz százalékos a feldolgozottság, az összes többiben nagyon alacsony, főleg a lett (LNB) esetében mondhatni véletlenszerű (1561 előfordulás). Mindez eleve következik abból, hogy a vezérlő mezők kitöltése nem kötelező.

A 007 és 008 mezők pozícióiba egy karakteres osztályozási jelzeteket kell rögzíteni. Ezek a dokumentumok típusait képviselik, tehát formai jelzetek. A pozíciók és az egy karakteres jelzetek korlátozott száma rendkívül szűkre szabott rendszer. Az osztályozónak gyakorlatilag gúzsba kötve kell tudnia táncolni. A 655-ös tárgyszó mezőben a természetes nyelvű formai tárgyszavak sokkal szélesebb választéka áll rendelkezésre

A nem kötelező vezérlő mezők kezelése a könyvtárak belső szabályzatain múlik. Ezekről nem sok tudható nyilvánosan, és nagyon különbözhetnek a kitöltési előírásaik. A buborékdiagramok tanulsága szerint többnyire vagy egyáltalán nem foglalkoznak a kitöltésükkel, vagy (mint a fenti 007/00 esetben – legalábbis a két német nyelvű nemzeti könyvtárban) egységesen alapértelmezéseket adnak meg. A többi esetében még ennél is kevesebb állapítható meg a külső elemző számára.

Ugyancsak nagyon feltűnő, hogy a német (DNB), az osztrák (ÖNB) és a brit (BL) nemzeti könyvtárak kivételével mennyire nem emelkedik ki a szöveges (007/00=t) nyelvi anyag feldolgozottsága; ráadásul a német és a brit könyvtárakban az elektronikus dokumentumok is (007/00=c) is bekerültek a nyelvi anyagok közé, miközben ennek a többi könyvtárban alig van nyoma. Az LC esetében van csak valamivel fontosabb szerepe a nyelvi anyagnak, itt a mikroformákkal együtt (007/00=h).

Érdekes, hogy a kétdimenziós nem kivetíthető ábrázolásokat (lényegében a fényképeket, Rf/06=k) csak a kongresszusi (LC), az osztrák (ÖNB), a lengyel (BNPL), és kisebb mértékben a belga (KBR) és brit (BL) könyvtárakban minősítették a 007/00=k pozíción nem kivetíthető grafikaként. Pedig ez a többi könyvtárban is magától értetődő lenne. (Vagy nem, ha a kettő együttes minősítését egy belső szabályzat redundánsnak, ezért feleslegesnek minősítene. A szabályzatok nem ismertek, a MARC21 pedig erre vonatkozóan nem ad eligazítást.)



5. diagram Bibliográfiai szint és dokumentumkategória

A rekordfej 06=k pozícióján (kétdimenziós nem kivetíthető ábrázolások) és a 007/00 dokumentumkategoróriák között szereplő g=kivetíthető kép és k=nem kivetíthető grafika sajátos ellentmondás is előfordul. Mind a Kongresszusi Könyvtárban (LC), mind a finn (NFI) nemzeti könyvtárban van olyan 1144 illetve 337 db rekord, melyben a kétdimenziós nem kivetíthető ábrázolások (rekordfej (06=k) egyben a kivetíthető kép (007/00=g) besorolást is megkapták. A feldolgozottság mértéke azonban olyan kicsi, hogy valójában nem lehet érvényes megállapításokat tenni. Ez a későbbi, 7.–12. diagramok összehasonlításai esetén már minden könyvtárra érvényes.

Az LC-ben a kétdimenziós nem kivetíthető ábrázolások többnyire fényképekről vagy grafikákról készült diapozitívok. A finn (NFI) nemzeti könyvtárban filmlapok és ugyancsak diák. Mivel több száz rekordról van szó, ez talán már nem lehet tévedés. Inkább azt bizonyítják, hogy a pozícióhoz kötött egy-karakteres osztályozási rendszer jelzetei elégtelenek.

5 Bibliográfiai szint és dokumentumkategória (Rf/07 vs. 007/00)

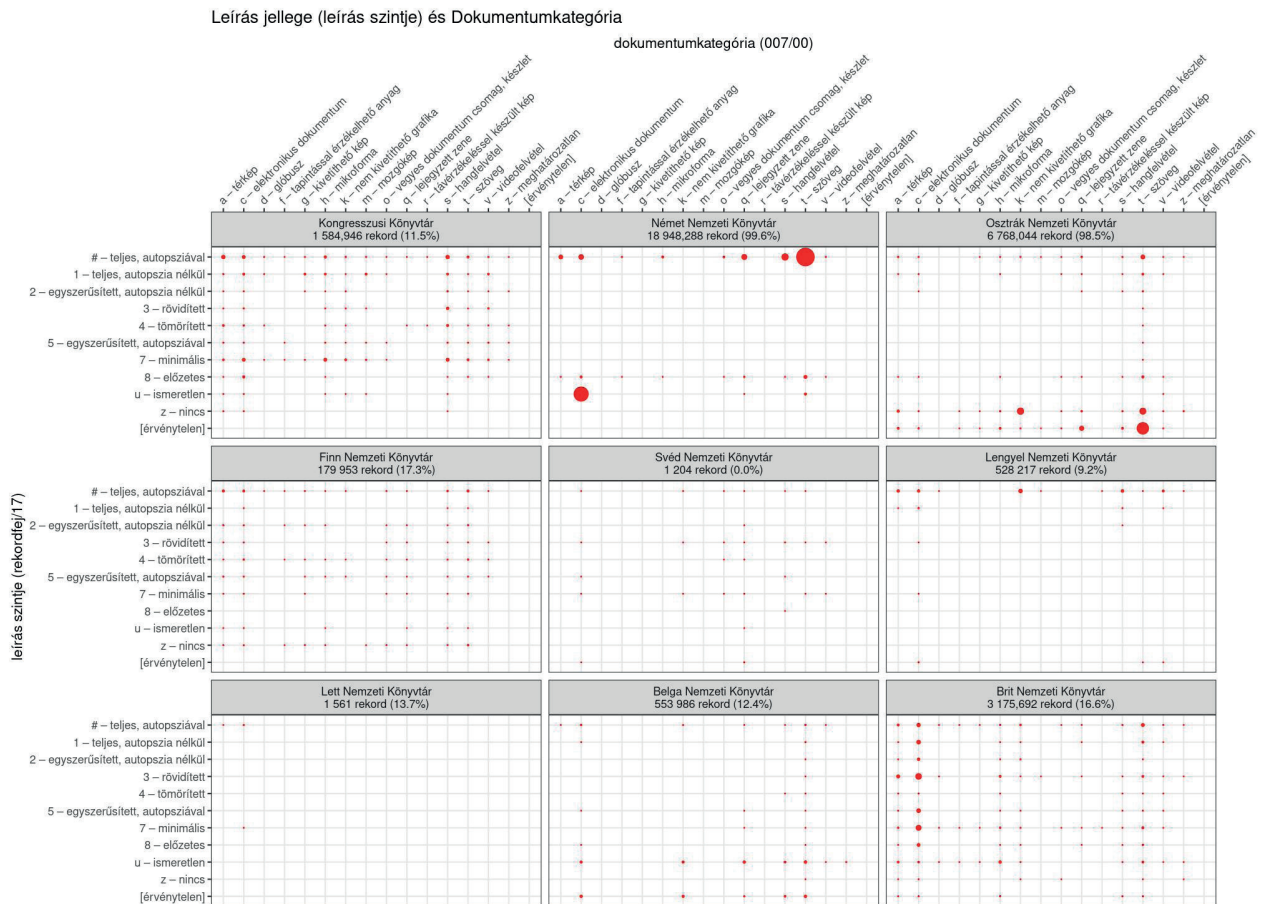
Hasonlóan a 3. diagramcsoporthoz, a német nyelvű nemzeti könyvtárakban a feldolgozottság közel száz százalékos. A diagramok is emlékeztetnek

a 3. diagramcsoportra, mert itt is a monografikus egységek (Rf/07=m és az analitikus monográfiák (Rf/07=a) feldolgozása az uralkodó.

Vizsgont meglepő módon a német (DNB) és részben a brit (BL) nemzeti könyvtárban nagyon jelentős az elektronikus dokumentumok (007/00=c) feldolgozása. Főleg monografikus dokumentumok esetén, de még időszaki formában is (Rf/07=a, ill. b, ill. m). Ennek a többi könyvtárban csak nyomai vannak. Ahogy arra a 2. diagramcsoport esetében utaltunk, ez meglehetősen ellentmondásban van a MARC21 ama szabályozásával, hogy ha van egy jelentős szempont, amely miatt a számítógéppel kezelt forrás egy másik rekordtípusba is tartozik, akkor annak rekordtípusával kell osztályozni.

Megint csak feltételezhető, hogy a következetes gyakorlat mögött belső, dokumentált szabályozás húzódik meg. Ezzel az a probléma, hogy szakmai nyilvánosság hiányában ezek a gyakorlatok se nem egységesülhetnek, se nem termékenyíthetők meg a szükséges fejlesztéseket.

A két német nyelvű könyvtáron kívüli könyvtárakra itt is a kis feldolgozottság a jellemző, ezért ezeknek a gyakorlatáról csak nagyon korlátozottan lehet következtetéseket levonni. Mindez az összes következő diagramcsoportra érvényes.



6. diagram Leírás jellege (szintje) és dokumentumkategória

A két német nyelvű nemzeti könyvtár kivételével a többinél bizonyára azért nincs rekordok több, mint 80%-a, olykor több mint 90%-a földolgozva, mert egyszerűen kihagyják ezeket. Azaz alapértelmezésként még csak az z=meghatározatlan minősítés megadását se tartották szükségesnek megadni. Ez azért is különös, mert arra utal, mintha nem is lenne belső szabályzat a 007 nem kötelező adatmező kitöltésére – vagy a szabályzatban az szerepel, hogy ezzel a mezővel nem is kell foglalkozni. Akkor viszont miért jelennek meg nagy ritkán mégis értékek?

6 Leírás jellege (szintje) és dokumentumkategória (Rf/17 vs. 007/00)

A leírás szintje (más szóval a leírás teljességének mértéke) szempontjából az lenne várható, hogy elsősorban a szöveges dokumentumokra (007/00 t) jellemző a teljes (Rf/17=# és =1) feldolgozás. Ez a német (DNB) nemzeti könyvtár esetében, ahol a feldolgozottság közel teljes (99.6%), kitüntetetten igaz. A többi esetében nem dönthető

el, mert annyira jelentéktelen a 007/00 vezérlőmező használata.

Hogy a DNB-ben az elektronikus dokumentumok 007/00=c leírás szintje ismeretlen (Rf/17=u), talán azzal függ össze, hogy különleges módon, helyi szabályzat szerint írják ezeket le.

Itt is jellemző, hogy a két német nyelvű nemzeti könyvtár kivételével a többinél rekordok több, mint 80–90%-a egyáltalán nincs földolgozva. Az alapértelmezésként még azt se tartották szükségesnek megadni, hogy a dokumentumkategória minősítése meghatározatlan (001/00=z. Ez a jelenség az összes további diagramcsoport esetében érvényes.

A 4., 5. és 6. diagramcsoportra érvényes, hogy a tételfaj dokumentumtípusának (Tf/06), továbbá a 007 mező 00 pozíció a dokumentumkategória jelentőségét csökkenti, hogy a 254-es mező \$a almezőjében a kották, a 256-os mező \$a almezőjében pedig az elektronikus dokumentumok típusa nagyon pontosan, ugyanakkor szabadon megadható. A 6. diagramcsoport esetében ez még kiegészül az 516-os megjegyzés mezővel, melyben az

elektronikus dokumentum típusa adható meg. Ezek azonban egyrészt szövegesen kezelt almezők, másrészt a katalogizáló rendszerek keresőit nem készítik föl arra, hogy böngészve vagy célirányosan kereshetők legyenek.

7 Rekordtípus és dokumentum speciális megjelölése szöveges dokumentumok esetén (Rf/06 vs. 007/01 t)

A rekordtípus (Rf/06) és a szöveges dokumentumok speciális fizikai megjelenéseinek (007/01=t) összehasonlítása is rendkívül kevés adatot tartalmaz. Érdekes megnézni, hogy – például az LC és a DNB esetében – milyen a megoszlás az egyes rekordtípusok, és milyen a nem kötelező speciális dokumentummegjelölések között szöveges dokumentumok esetén (lásd a 2. és 3. táblázatot).

A szöveg rekordtípus (Rf/06=t) adatai alapján ugyan egyértelmű, hogy a legtöbb dokumentum nyelvi anyag (007/=a). A két német nyelvű nemzeti könyvtár (DNB és ÖNB) adatai szerint túlnyomó többségük normál nyomtatás. A Kongresszusi Könyvtárban (LC) alig minősítettek normál nyomtatással (007/01 t=a) ezért ugyanez csak feltételezhető. Ahogy azt se volna érdektelen elemezni, hogy milyen rekordtípusból mennyi a vakírás vagy Braille-írás. A fenti összehasonlítással azonban ilyen adatok megállapítása nem lehetséges, mert a 007 pozícióit a könyvtárak lényegében nem töltik ki.

Az ember nem gondolná, hogy mind a Német Nemzeti Könyvtárban (DNB), mind az angolban (BL) több forrást dolgoztak föl, mint a Kongresszusi könyvtárban! BL = 19,154.445 rekord, DNB = 19,026.644 rekord, LC = 13,817.649 rekord.

2. táblázat A rekordtípus (Rf/06) értékei a Kongresszusi Könyvtárban és a Német Nemzeti Könyvtárban

Rekordtípusok megoszlása (rekordfej/06)	LC 13817649 = 100 %		DNB 19026644 = 100 %	
a – nyelvi anyag	12548611	91.0 %	16903888	88.0 %
c – nyomtatott zenemű	136248	1.0 %	578352	3.4 %
d – kéziratos zenemű	8383	~0.1 %	–	–
e – kartográfiai anyag	333929	2.4 %	291148	1.5 %
f – kéziratos kartográfiai anyag	2848	~0.1 %	–	–
g – audiovizuális anyag	120526	0.9 %	10469	~0.1 %
l – nem zenei hangzó anyag	58991	0.4 %	137321	0.7 %
j – zenei hangzó anyag	450795	3.3 %	1008219	5.3 %
k – kétdimenziós (nem kivethető) ábrák	127809	~0.1 %	–	–
m – számítógép által kezelt állomány	11017	~0.1 %	–	–
o – vegyes dokumentumcsomag, készlet	19	~0.0 %	20761	~0.1 %
p – vegyes anyag	11844	~0.1 %	–	–
r – háromdimenziós alkotások	46	~0.0 %	–	–
t – kéziratos nyelvi anyag	6583	0.5 %	76433	0.4 %
[érvénytelen, hibás jelzet]	–	–	–	–

3. táblázat A rekordtípus (Rf/06) és a szöveges dokumentumok speciális fizikai megjelenéseinek (007/01=t) metszeteinek rekordszámait a Kongresszusi Könyvtárban és a Német Nemzeti Könyvtárban

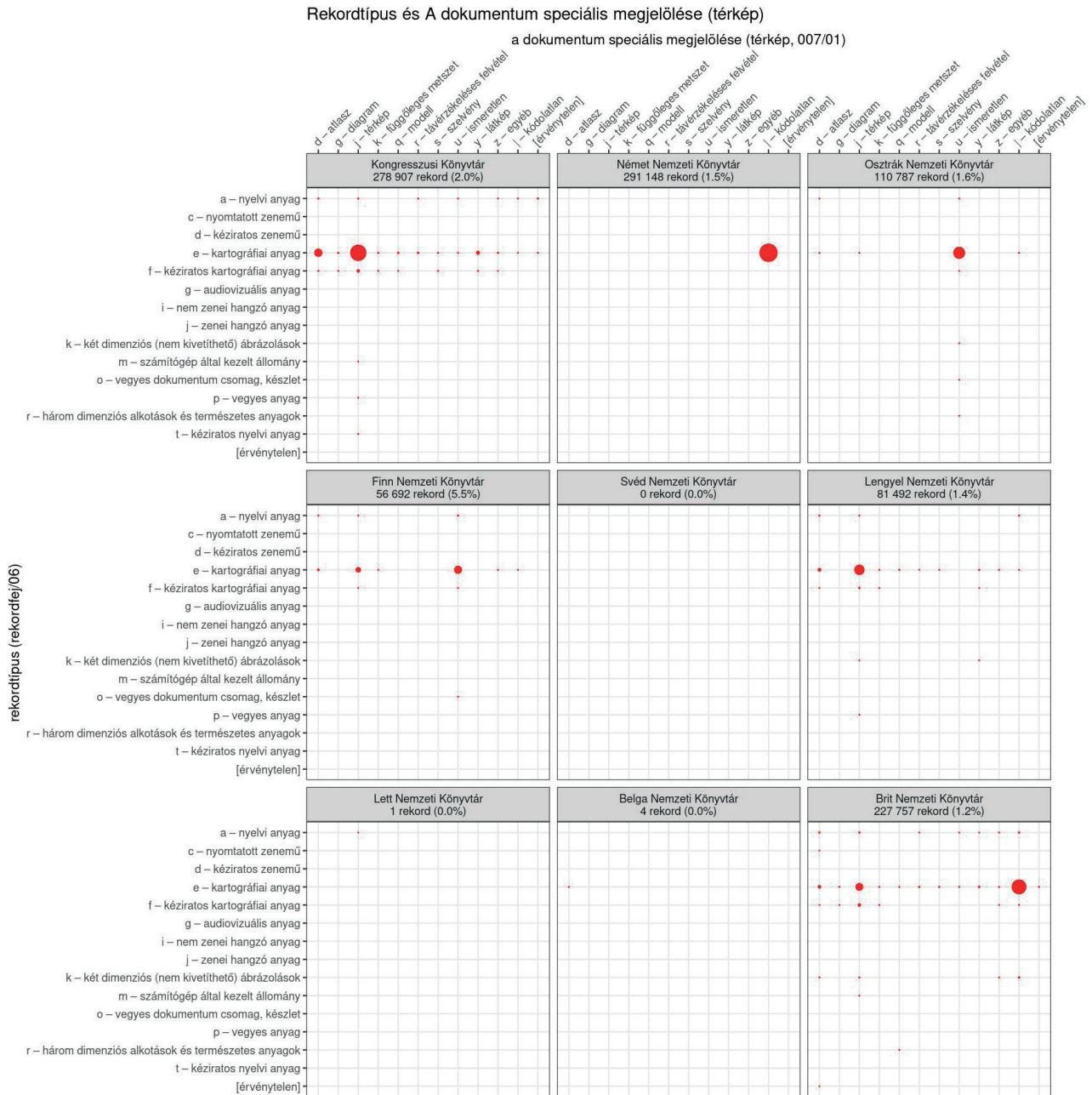
Dokumentum speciális megjelölése (007/01 t)	LC 13817649 = 100 %		DNB 19026644 = 100 %	
a – normál nyomtatás	8261	0.06 %	10218462	5.37 %
b – nagybetűs nyomtatás	31	~0.00 %	–	–
c – Braille-írás (vakírás)	–	–	–	–
d – szabadlapos kötésben	2	–	–	–
u – ismeretlen	550	~0.00 %	10165227	0.534 %
z – egyéb	44	~0.00 %	–	–
– kódolatlan	60660	0.44 %	–	–
[érvénytelen, hibás jelzet]	14	~0.00 %	1	–

Rekordtípus és A dokumentum speciális megjelölése (szöveg)

a dokumentum speciális megjelölése (szöveg, 007/01)



7. diagram Rekordtípus és a dokumentum speciális megjelölése szöveges dokumentumok esetén



8. diagram Rekordtípus és a dokumentum speciális megjelölése térképek esetén

Az angolban (BL) azonban a nem kötelező minősítések (00X) megadása kevésbé részletes, mint a másik kettőben.

A német (DNB) és az osztrák (ÖNB) nemzeti könyvtárban valószínűleg szabályzat rögzíti, hogy a dokumentum speciális megjelölése esetén alapértelmezés a z=egyéb, illetve az u=ismeretlen. Az összes többi könyvtárban a táblázatban, illetve diagramokban látható adatok az egészhez képest elenyészőek, és véletlenszerűen keletkezettek. Vajon ezekben szabályzat írja ezt elő?

Mivel a 65X-es tárgyszó mezőkben például a vakírást, sőt akár annak speciális fajtáit is osztályozhatják, összetett kereséssel a nyelvi anyagnak megfelelő tárgyszóval, és a vakírás vagy a Braille-írás tárgyszóval keresve az ilyen adat mindezek ellenére megtalálható – csak éppen ezt a fajta keresést legfeljebb a nagyon gyakorlott referenzkönyvtáros képes megvalósítani a mai online katalógusokban, lévén, hogy a felhasználók többsége összetett keresésre csak nagyon ritkán fanyalodik.

A dokumentum speciális megjelölése (térkép) és Illusztráltság (könyvek)

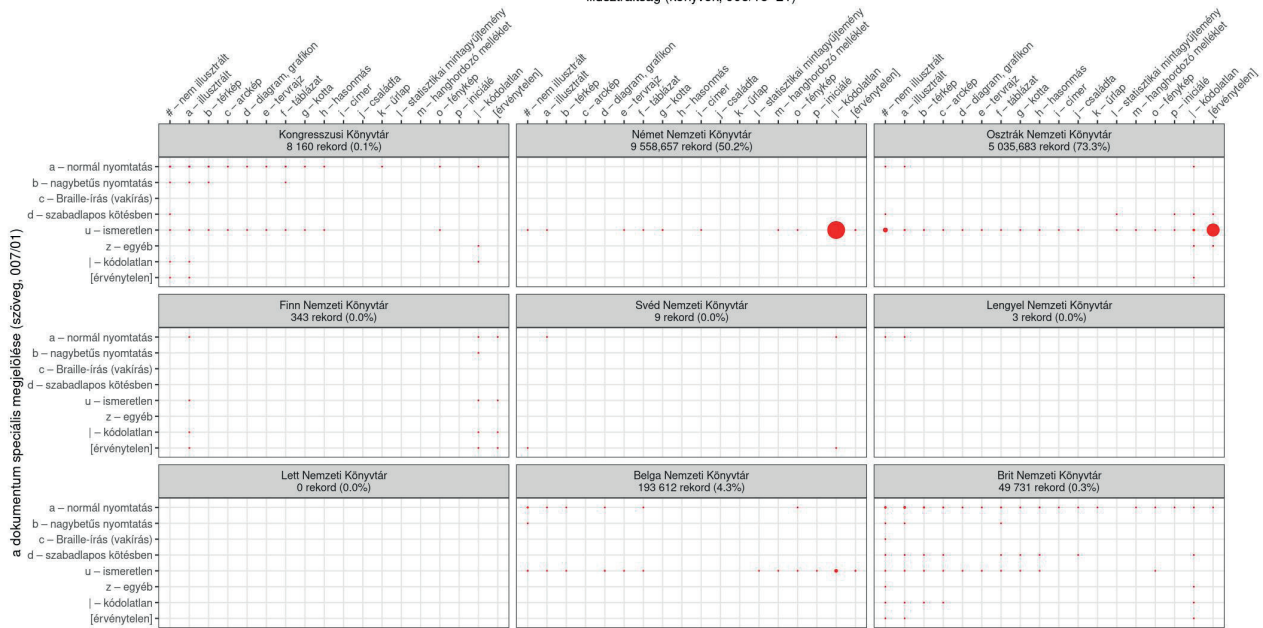
illusztráltság (könyvek, 008/18–21)



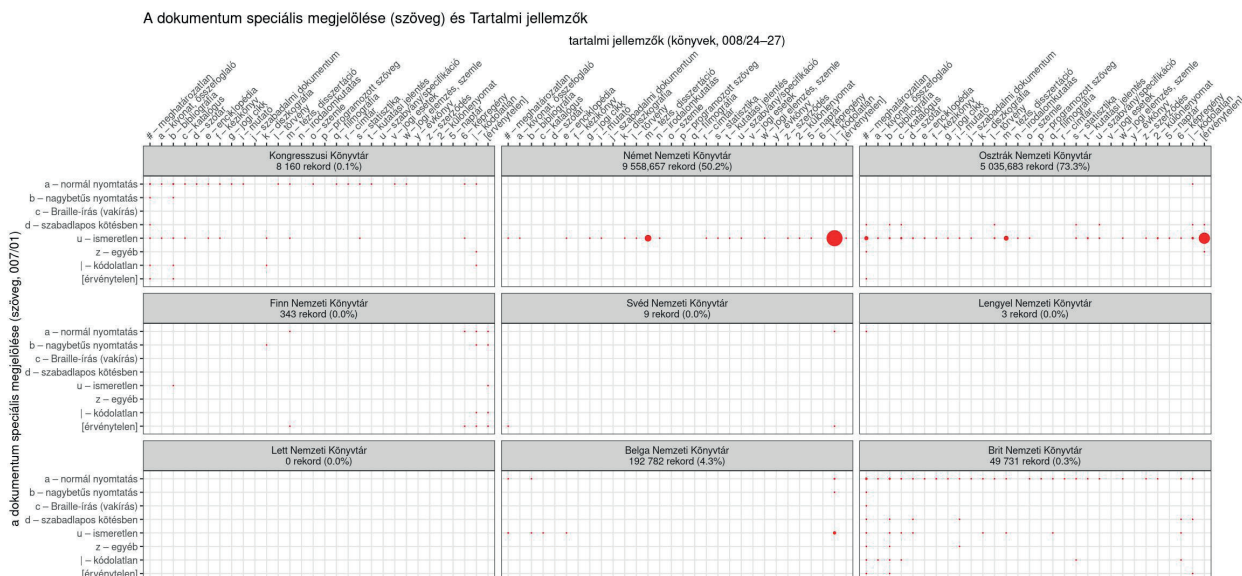
9. diagram Könyveknek minősülő térképek és illusztráltak

A dokumentum speciális megjelölése (szöveg) és Illusztráltság (könyvek)

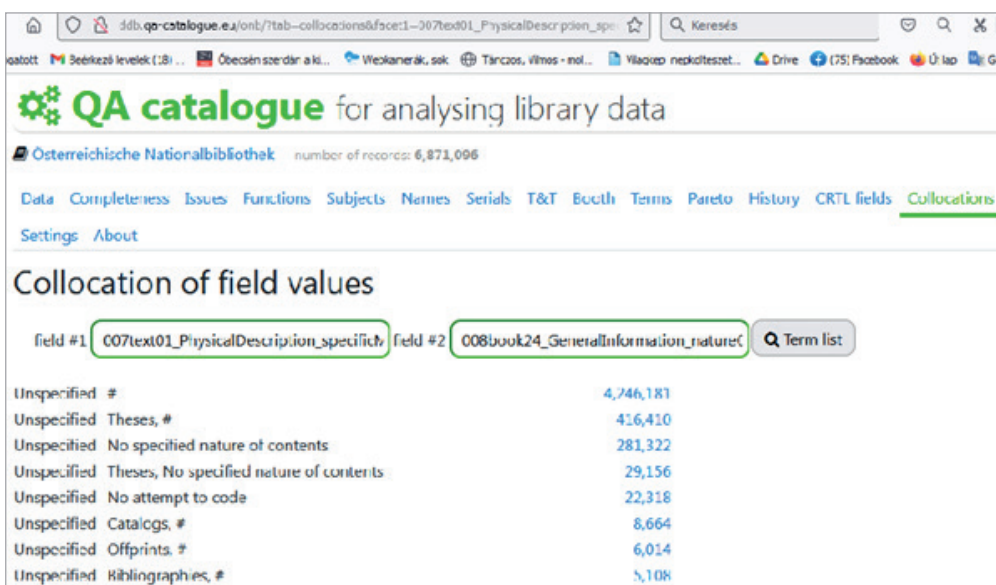
illusztráltság (könyvek, 008/18–21)



10. diagram Szöveges könyvek illusztráltsága



11. diagram Szöveges dokumentumok tartalmi jellemzői



2. ábra Az ismeretlenként minősített speciális dokumentum megjelölések (007/00=u) és a könyvek kettős kereszttel (#) minősített, tehát érvénytelen tartalmi jellemzőinek (008/24) metszete az Osztrák Nemzeti Könyvtárban¹²

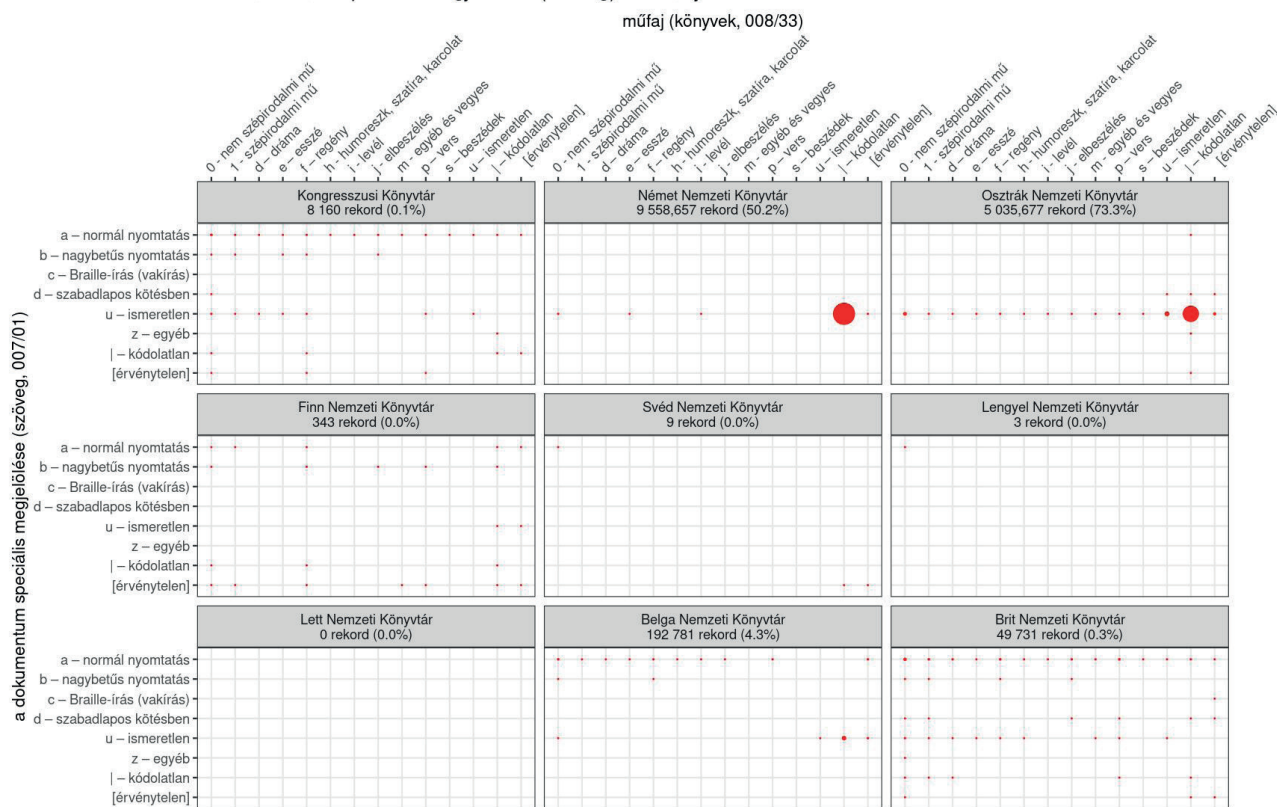
A 007 és 008 mezők kitöltetlenségének alapvetően az az oka, hogy a dokumentumtípusok speciális osztályozása – ahogy erre előzőleg már utaltunk – formai tárgyszavakkal/deszkríptorokkal (655) sokkal rugalmasabban megoldható. Ráadásul a fizikai jellemzők is könnyen megadhatók szövegesen is az 588 mezőben (Megjegyzés a terjedelemtől/fizikai jellemzőkről).

Ugyanakkor a 300-as szöveges mező \$a almezőjében ugyancsak megadható a fizikai egységek száma, a fizikai jellemzők. Az 515-ös szöveges mezőben pedig az elektronikus dokumentumok típusa.

Mindez nem változtat azon, hogy egy, a jelenleginél sokkal jobban kidolgozott egyszerű tipológiai és egyéb jellemzők osztályozási rendszere a 007-és 008-as vezérlőmezőkben kifejezetten hasznos lehetne – amennyiben a katalógusok keresőit ezek könyvű, felhasználóbarát alkalmazására felkészítenék.

12 Többször megkíséreltünk levélben erre az Osztrák Nemzeti Könyvtárban (ÖNB) rákérdezni, de nem válaszoltak

A dokumentum speciális megjelölése (szöveg) és Műfaj



12. diagram Szöveges könyvek műfaji jellemzői

8 Rekordtípus és dokumentum speciális megjelölése térképek esetén (Rf/06 vs. 007/01 a)

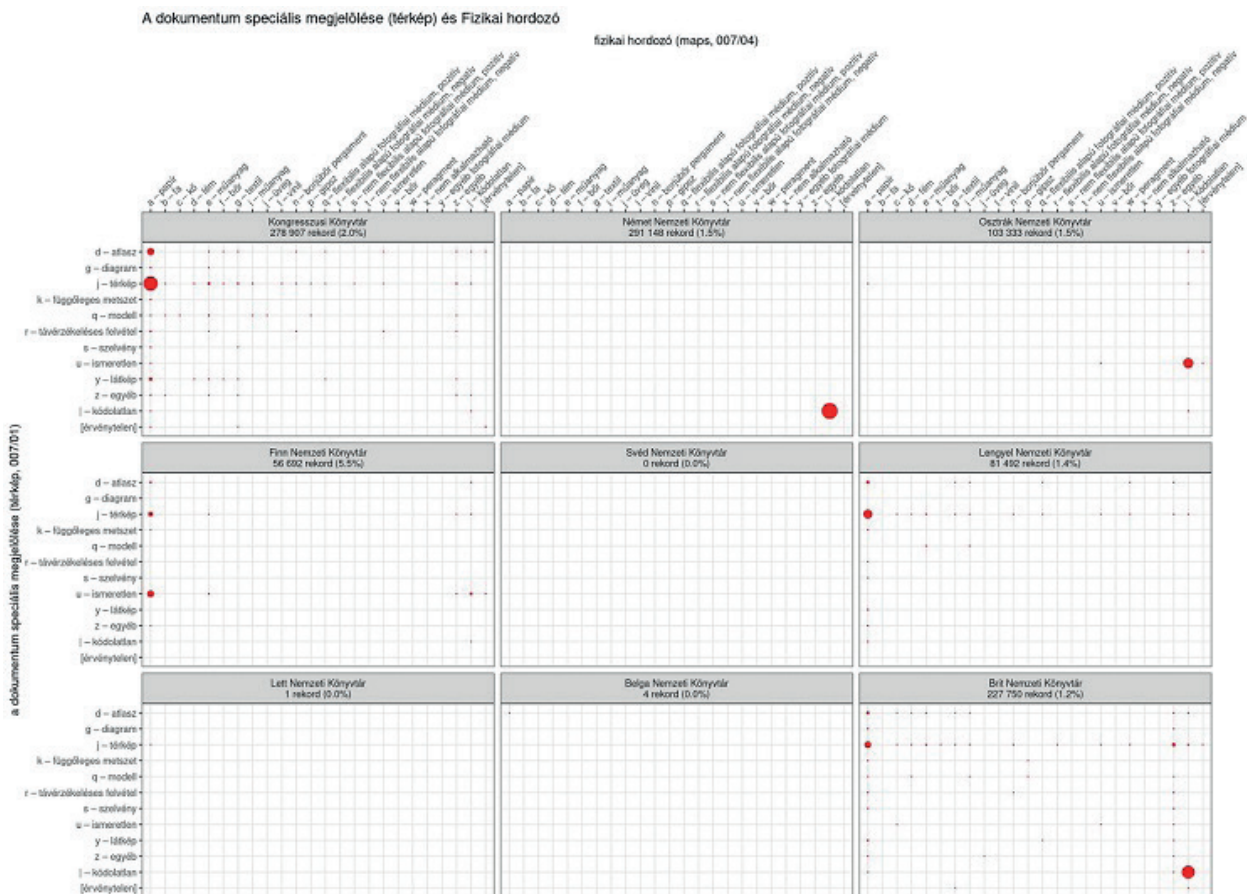
A térképek esetében még kevésbé használják a 007 vezérlő mező 01 pozícióját osztályozásra. A két német nyelvű könyvtárban (DNB, ÖNB) érdemben itt még alapértelmezett besorolást se alkalmaznak, mint a könyvek esetén. A kongresszusi (LC) könyvtárban ugyan látható a kartográfiai rekordok sűrűsödése atlaszok és térképek esetén (007/01=a, ill. =j), de az előforduló kevés adat legfeljebb véletlenszerű, nem lehet belőlük következtetéseket levonni. Itt még az se feltételezhető, hogy a nemzeti könyvtárban alapértelmezéseket adnának meg a kitöltési szabályzatok.

A pozíciók gyenge kitöltöttsége összefügg azzal hogy a 300-as Terjedelem/Fizikai jellemzők mező \$a és \$b almezői, továbbá a 0t 588-as Megjegyzés a terjedelemtől/fizikai jellemzőkről megjegyzés mezőkben ugyancsak minden térképtípust meg lehet adni szövegesen (lásd a 4. táblázatot).

9 Könyveknek minősülő térképek és illusztráltságuk

A rendkívül kevés adat, illetve az adatok teljes hiánya ugyanarra enged következtetni, mint az előző, 8. diagramcsoport esetén. Minden jel szerint a kartográfiai anyagok illusztráltságának finomabb osztályozását a 008 mező 18–21. pozíciójában nem tekintik fontosnak. Ez talán azért van, mert – hasonlóan a térképek speciális osztályozásához az előbbi diagramcsoport tanulságához – a térképek speciális illusztráltságát részletesen le lehet írni szövegesen az 588-as Megjegyzés a terjedelemtől/fizikai jellemzőkről mezőben (lásd a 4. táblázatot), a 65X tárgyszó mezőkben pedig pontos, speciális, természetes nyelvű tárgyszavakkal/deszriptorokkal is osztályozni lehet.

A Kongresszusi Könyvtár (LC) és brit (BL) nemzeti könyvtárban (némileg a lengyelben is, BNPL) a kevés feldolgozott rekord ugyanakkor arra utal, hogy a térképeknek nagyon is van illusztráltsága.



13. diagram Térképajták fizikai hordozói

10 Szöveges könyvek illusztráltsága (007/01 t vs. 008/18–21)

Figyelemre méltó, hogy a szöveges dokumentumok (007/01=t) illusztráltságának (008/218–21=a) vagy semmi jelentőséget nem tulajdonítanak, vagy a 008-as mezőben nem kezdenek vele semmit. A német (DNB) és osztrák (ÖNB) nemzeti könyvtárakban különös módon legfeljebb annyit tesznek, hogy a rekordok felében-háromnegyedében a 007/01 szöveget u=ismeretlennek, a 008/18–21 pozíciókon pedig az illusztráltságot kódolatlanak vagy érvénytelennek minősítik.

Több nemzeti könyvtárban (LC, DNB, ÖNB, KBR, BL) elenyésző számú esetben tapasztalható volt, hogy igenis lenne itt osztályozni való, de aztán mintha gyorsan föladták volna.

Mind az előző 9. mind erre a diagramcsoportra érvényes, hogy a MARC21 az illusztrációk osztályozásának dolgában különösen redundáns. A 300-as szöveges mező \$b almezőjében (Egyéb fizikai jellemzők) ugyancsak megadhatók az illusztrációk. Ez önmagában nem baj, ha a MARC21-en szabályoznák egymáshoz viszonyítva a két mező tartal-

mának használatát. Ez egyébként az előző diagramcsoportokban említett redundanciák esetében sem történik meg.

11 Szöveges dokumentumok tartalmi jellemzői (007/01 t vs. 008/24–27)

A szöveges dokumentumok (007/0= t) tartalmi jellemzőinek (008/24–27) megadására is ugyanaz jellemző, mint az előző, 10. diagramcsoportra. Alig van a könyvtárakban feldolgozás, a németben csak csupa kódolatlan (008/24–27=|), az osztrákban meg csupa ismeretlen.

Itt is lehetséges, hogy a tartalmi jellemzők szerinti dokumentumosztályozást a 655 formai tárgyszavak mezőjében oldják meg, és ezért nem foglalkoznak a 008/14–27. pozíciók kitöltésével. Sajnálatos, hogy e téren hiányos a szakmai nyilvánosság, a beszámolók, elemzések, nem utolsó sorban pedig a helyi szabályzatok nyilvános hozzáférhetősége, szakmai megtárgyalása.

A kutatási jelentéseket egyébként részletesen osztályozni lehet az 513-as mező \$a almezőjében (lásd a 4. táblázatot). A szöveges mezők használá-

tára vonatkozóan már a korábbi diagramcsoportokban jeleztük, hogy kereshetőségük egyáltalán nem megoldott.

Megint csak figyelemre méltó az osztrák (ÖNB) nemzeti könyvtár gyakorlata, ahol a dokumentum speciális megjelölésére az ismeretlen minősítést (007/00=u), a könyvek tartalmi jellemzőjeként viszont kettős keresztet (#) adnak meg, ami nem megengedett.

12 Szöveges könyvek műfaji jellemzői (007/01 t vs. 008/33)

A szöveges dokumentumok (007/01=t) műfajának megadására is az előbbi 11. *diagramcsoport* jellemzői érvényesek. Alig van a könyvtárakban feldolgozás, a németben (DNB) és osztrákban (ÖNB) megcsupa kódolatlan (008/33=|).

Itt is fölvetődik, hogy a műfajok szerinti dokumentumosztályozást valószínűleg a 655 formai tárgyszavak mezőjében oldják meg, és ezért nem foglalkoznak az egyébként sem kötelező 008/33. pozíciók kitöltésével. Kétségtelen ugyanis, hogy a dokumentumok formai osztályozására nagyságrendileg sokkal, de sokkal több tárgyszóra van szükség, semmint, hogy azokat a vezérlőmezők pozícióiba egy karakteres jelzetekkel bele lehetne préselni.

Megnézve a német (DNB) nemzeti könyvtár „kódolatlan” minősítésű rekordjait, olyanok vannak benne, mint pl. életrajz, feladatgyűjtemény, gépkönyv stb., és tényleg, se a 008/33, se a 008/25–27 nem rendelkezik ezekhez megfelelő tartalmi jellemzőkkel, illetve műfajokkal. Ugyanakkor olyanok is szerepelnek a kódolatlanok között, mint például regények, esszék, melyeknek van a 008/33-ban típusuk, az „f=regény”, „e=esszé”. Az életrajzok pedig az „1=szépirodalom” típusba sorolhatók.

Minden jel szerint a könyvtárak nem akarnak se az illusztrációk, se a tartalmi jellemzők, se a műfajok besorolásaival a vezérlő mezőkben (00X) vesződni. Feltehetően azért, mert a szűkös, néhány pozícióhoz kötött, egy karakteres osztályozási választék használata sok fejtöréssel jár, és a vonatkozó MARC 00X mezők értékei se nem elégségek, se nem alkotnak koherens műfaji-tartalmi rendszert.

Mindez a 008 könyvek esetére vonatkozik, de valószínűleg ugyanez a helyzet az összehasonlításba nem bevont időszaki kiadványok és zenei anyagok esetén is, ahol szintén vannak ilyen tartalmi jellemzők és műfajok.

13 A térképfajták (007/01) fizikai hordozójuk (007/04)

A 13. *diagramcsoport* tipikusan példázza a 007-es mező nem kötelező voltának a következményeit.

A mező 04. pozícióján a mindenkori dokumentumtípus (esetünkben a térképek) hordozóit kellene osztályozni. Ez vitathatatlanul fontos keresési szempont lehet. Ennek ellenére a kitöltöttsége minimális. Ez már önmagában fölveti, hogy akkor egyáltalán miért van a viszonylag elenyésző esetekben mégis kitöltve a pozíció?

Ugyanakkor a kongresszusi (LC) könyvtár esetében még a viszonylag kevés (178907 rekord, 2%) esetében is jól látható már, hogy nagyon sok már hordozótípus is előfordul, nemcsak a papír. Teljes kitöltöttség esetén a diagramok feltehetően nagyon látványos képet mutatnának.

Kérdés, mennyire osztályozzák a forrásokat a hordozó szempontjából más mezőkben (főleg a 65X mezőkben)?

14 Az ugyancsak dokumentumtípusokhoz használható más mezők

Redundáns mezők a 00X mező pozícióinak tartalmával. (4. táblázat)

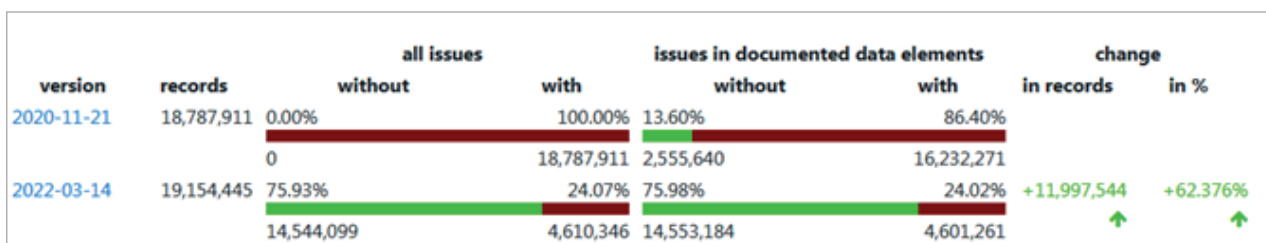
Összegzés

A rekordfej és a vezérlőmezők (00X) értelmezése, alkalmazása minden jel szerint további pontosítást, fejlesztést igényel. Jelenleg rendkívül széttartó a gyakorlat. Hogy pontosan milyen, az a hatalmas adatmennyiség miatt csak szűrőpróbaszerűen ellenőrizhető. Az egyes nemzeti könyvtárak szakemberei azonban, ismerve a saját gyakorlatukat, mindezek alapján jobban tudják értelmezni mindazt, amire ebben a tanulmányban talán sikerült jelzésszerűen utalni. És ezen a nyomon haladva lehetőség nyílhat a katalógusaik javítására, és a MARC21 fejlesztésére is.

Jelentős a feldolgozottsági különbség a legnagyobb (LC, DNB, BL, de még az osztrák ÖNB) és a kisebb nemzeti könyvtárak (finn NFI, svéd libris, lengyel (BNPL), lett (LNB) és belga (KBR) nemzeti könyvtárak között. A legérdekesebb ebből a szempontból az OSZK-val való összehasonlítás lenne, de ennek katalógusát nem volt mód letölteni és elemezni a QA catalogue rendszerbe. Azt is érdekes volna tudni, hogy mely könyvtárakban áll rendelkezésre a rekordfej és a vezérlő mezők (00X) kitölt-

4. táblázat A rekordfej és a 00X vezérlő mezők dokumentumtípusok osztályozására való pozíciói és a tartalmilag velük összefüggő, ugyancsak dokumentumtípusok leírására használható szöveges megjegyzésmezők

A rekordfejben és a 00X vezérlőmezőkben	A szöveges megjegyzésmezőkben
Rekordfej (rekordtípus 06 c=nyomtatott zenemű; d=kézirat zenemű)	254 \$a A kotta típusának megevezése
Rekordfej (rekordtípus 06 m=számítógép által kezelt állomány; 007/00 dokumentumkategória c=elektronikus dokumentum)	256 \$a és \$b Elektronikus dokumentumok jellemzői
007/01 a dokumentum speciális megjelölése: térkép 007/01 a dokumentum speciális megjelölése: szöveg 008/18–21 könyvek: illusztráltság	300 Terjedelem, fizikai jellemzők \$a A dokumentum fizikai hordozójának a fajtája \$b Egyéb fizikai jellemzők, illusztrációk 588 Megjegyzés a terjedelemtől/fizikai jellemzőkről
008/24–27 tartalmi jellemzők: könyvek	513 Megjegyzés a kutatási jelentés típusáról és idejéről \$a A jelentés típusa.
Rekordfej (rekordtípus 06 m=számítógép által kezelt állomány; 007/00 dokumentumkategória c=elektronikus dokumentum)	516 Megjegyzés az elektronikus dokumentum típusáról vagy adatairól



3. ábra Javitási grafikon a Brit Nemzeti Könyvtárban

téséhez dokumentált szabályzat, és hol dolgoznak pusztán szóbelileg közvetített gyakorlat alapján.

Az elemzés fontosabb konkrét tapasztalatai:

- A rekordfej és a vezérlő mezők (00X) pozícióin a dokumentumok típusainak egy karakteres mesterséges nyelvű osztályozási jelzeteit rögzítik. Ezek a karakterek pozícióként elvileg meghatározott szempont szerinti tipológiát képviselnek. Az egyetlenszer alkalmazható, egyetlen pozícióhoz kötött, egykarakteres osztályozási jelzetrendszer csak nagyon általános, durva besorolást tesz lehetővé.
- Az egyes osztályok meghatározásai elégtelennek tűnnek. Ez az összehasonlításokban derül ki. Például nem látszik teljesen világosan, hogy mi értendő alárendelt részegységnek (Rf/07=d) kéziratok vagy plakátok esetén (Rf/06=d, ill. =k). A kitöltési szabályzatok ismerete erre talán kitér, de hozzáférhetetlenek.
- A rekordfej pozícióinak kitöltése kötelező, a 00X vezérlő mezők kitöltése nem kötelező. Ennek következtében válhatott lehetővé, hogy az utóbbi mezők feldolgozottsága rendkívül kicsi. Annyira, hogy nem is lehet érdemben részletesen elemezni a feldolgozásukat, és erről érvényesebb észrevételeket tenni.
- A pozíciók kitöltését feltehetően helyi szabályzatok vagy szóbeli gyakorlat határozza meg. Ez a gyakorlat explicit szakmai nyilvánosságának hiányában nagyban korlátozza az elemzésben felvetődő problémák átfogó, konszenzusos megoldását, nehezíti a nemzetközi gyakorlat egységesülését. Hiányzik a helyi szabályzatok szakmai közérdeket szolgáló nyilvánossága és megtárgyalása, hiányzanak a rekordfej és a vezérlő mezők feldolgozásáról szóló beszámolóok. Mindez azt is jelenti, hogy nem állnak rendelkezésre hozzáférhető információk a 00X adatmezőkkel kapcsolatos MARC21 fejlesztésekhez sem.
- A felvetődő problémák és megoldásuk nincs ellentmondásban azzal, hogy lehetséges egy, a jelenleginél jobban kidolgozott, átfogó tipológiai osztályozás a vezérlő mezőkben (007 és 008), nem utolsósorban annak érdekében is, hogy a katalógusok keresőrendszereit az ezekben a mezőkben való, felhasználóbarát keresésre felkészítsék.
- Sok esetben a tételfej dokumentumtípusával, illetve a 007 mező dokumentumtípusaival párhuzamosan léteznek szöveges mezők is (mint a 254, 256, 300, 513, 516, 588), melyekben

ugyancsak osztályozhatók dokumentumtípusok, és a 655-ös mezőben kötött tárgyszavakkal is. Nincs a MARC21-ben szabályozva ezeknek a mezőknek a viszonya, melyiket milyen értelemben célszerű használni a redundancia csökkentése, a feldolgozás egyértelműsége érdekében. Ráadásul: szemben a tételfej és a 00X mezők kötött osztályozási jelzeteivel szemben a felsorolt megjegyzésmezők szövegesek, és a katalógizáló rendszerek keresői nincsenek egyáltalán felkészítve arra, hogy felhasználóbarát módon bevonhatók legyenek laikusok által is a keresésbe.

- A vezérlőmezők pozícióin kis számban, de sok katalógusban fordulnak elő definiálatlan értékek, nagyon kis számú találattal. Ezek sokszor nem kellően ellenőrzött tömeges importból származhatnak, és az általunk alkalmazott QA catalogue segítségével könnyen felismerhetők és kiszűrhetők.
- Végül az egyik legfontosabb: a katalógusok, repozitóriumok stb. metaadatsémáiba be kell építeni az adott intézményben vagy felhasználói közösségben kialakult/kidolgozott egyedi használati szokásrendjének (alkalmazói profiljának, application profile) kezelésére alkalmas részt.

Kitekintés

Hogy a QA catalogue által feltárt katalógusokban mennyire eredményes lehet a hibajavítás, azt például a brit (BL) nemzeti könyvtár katalógusállományában elvégzett munka igazolja.

Ha a rekordban akárcsak egyetlen hiba van, akkor az hibás rekordnak számít. Két eshetőséget különböztetünk meg:

(1) az adat értéke egyértelműen eltér egy jól dokumentált előírástól. Például, ha a 245\$ind1 értéke térköz, akkor az nem felel meg a MARC21 előírásának, mivel ennek az indikátornak 0-t vagy 1-et kell tartalmaznia.

(2) lehet továbbá olyan adatelem, amely a MARC21-ben vagy annak ismert kiterjesztésében nincs definiálva, ezért nem ítéltető meg az érvényessége, mert nem ismerjük a szabályait. Például: 024\$9 esetében nem vagyunk biztosak benne, hogy probléma-e, lehet, hogy ez egy lokális almező, mert a 9-es erre van fenntartva (de lehet, hogy elírás is). 250\$p esetében szintén nem vagyunk biztosak; a MARC21-ben nem szerepel, de mivel nem 9, sokkal kisebb a valószínűsége annak, hogy ez egy lokális almező. Mindkét esetben előfordulhat, hogy a rekordot máshonnan importálják, és a lokális almező meghatározása az eredeti intézményben elérhető, de az aktuális katalógusban már kilóg a kontextusból.

A baloldali oszlop fekvődiagramjai az összes, tehát az (1) és a (2) eseteket mutatják meg. Ezek szerint az első vizsgálat idején (ez a felső sor) a rekordok 100%-ban volt adatmezőt érintő hiba. Ezek 70%-ban indikátor-hiba, közel 50%-ban pedig vezérlő mező hiba. További hibatípusok a vezérlő mezők (00X) esetében az érvénytelen adatértékek, az XXX adatmezők esetében a definiálatlan mezők, az indikátorok esetében a nem üres (kódolatlan=) indikátorhelyek. Ezen felül a rekordok 17%-ban a definiált mezőkben volt definiálatlan almező. Az alsó sorban a formális hibajavítások javításának az eredménye látható fekvődiagramban.

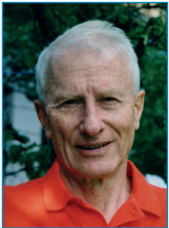
A második oszlop fekvődiagramjai csak az (1) típusú hibaesetekre vonatkoznak.

Irodalom

- [1] HUNMARC A bibliográfiai rekordok adatcsere formátuma., KSZ 4/1. 2002, március.
- [2] Király, P., Takács, Ákos, Svetlik, R., Virolainen, T., O. Heggø, D. M., Rolschewski, J., Hemme, F. (2017-) *QA catalogue, v0.5.0*, 2022. <https://www.doi.org/10.5281/zenodo.6394934>
- [3] Király, P. *Validating 126 million MARC records*, In: DATeCH2019, Proceedings of the 3rd International Conference on Digital Access to Textual Cultural Heritage Brussels, Belgium, May, 8-10, 2019, p. 161–168. <https://www.doi.org/10.1145/3322905.3322929>
- [4] Király, P. *Empirical evaluation of library catalogues*, EuropeanaTech Newsletter, 15, 2020. <https://pro.europeana.eu/page/issue-15-swib-2019#empirical-evaluation-of-library-catalogues>
- [5] MARC21 *Format for Bibliographic data*, Library of Congress. Update No. 30 (May 2020) <https://www.loc.gov/marc/bibliographic/>
- [6] Ungváry, R. *Besorolási, szabványosított, normatív vagy „autorizált”*, Tudományos és Műszaki Tájékoztatás, 66(6), p. 328–342, 2019. <https://tmt.omikk.bme.hu/tmt/article/view/12309/14064>

- [7] Ungváry, R. *MARC21 tartalmi adatmezők használata jelentősebb nagykönyvtárakban. Egy elemzés néhány tanulsága*, In: Networkshop, 2020, p. 33-53. <http://real.mtak.hu/119192/1/ungvary.pdf>
- [8] Ungváry, R. *Ismeretszervező-könyvtári rendszerek tartalmi feltárásának összehasonlító vizsgálata MARC21 környezetben*, Tudományos és Műszaki Tájékoztatás, 67(11), p. 655-680, 2020. <https://tmt.omikk.bme.hu/tmt/article/view/12776/14514>
- [9] Ungváry, R., Király, P. *Bemerkungen zu der Qualitätsbewertung von MARC-21-Datensätzen*, In: Franke-Maier, M., Kasprzik, A., Ledl, A., Schürmann, H. (eds.) *Qualität in der Inhaltsschließung*, De Gruyter Saur, Berlin, Boston, 2021, p. 177–227. <https://www.doi.org/10.1515/9783110691597-011>
-

Beérkezett: 2022. július 4.



Ungváry Rudolf
informatikus mérnök
E-mail: ungvaryr@gmail.com



Király Péter
szoftverfejlesztő, kutató
Göttingen eResearch Alliance
E-mail: peter.kiraly@gwdg.de