

AZ MTA KÖNYVTÁR ÉS INFORMÁCIÓS KÖZPONT GYŰJTEMÉNYE A WORLDCAT¹ NEMZETKÖZI KATALÓGUSBAN

Gyuricza Andrea
 Gyűjteményszervezési Osztály
 MTA Könyvtár és Információs Központ
 gyuricza.andrea@konyvtar.mta.hu
 ORCID: 0000-0003-3960-5499

Haász Antal
 Szakinformatikai Osztály
 MTA Könyvtár és Információs Központ
 haasz.antal@konyvtar.mta.hu



DOI: 10.31915/NWS.2018.7

The collection of the Library and Information Centre of Hungarian Academy of Science in the international WorldCat catalogue The Library and Information Centre of HAS (MTA KIK) joined to the international WorldCat catalogue in February 2017. Currently MTA KIK is the only active WorldCat member from Hungary, although a lot of Hungarian libraries are planning to join. MTA KIK submitted more than 770.000 records (about 76% of its online catalogue) to the WorldCat international catalogue in November 2017, after almost 1 year preparatory phase. This study is about the contracting with OCLC and we describe the workflow of the preparation and the record-export.

Keywords: MARC21, WorldCat, common catalog, service improvement

Bevezetés

Könyvtárunk katalógizálási gyakorlatában 2016 januárjában történt meg a váltás HUNMARC-ról a MARC21 szabványhasználatára. Az átállásközvetlen előzményének egyrészt az Országos Széchényi Könyvtár MARC21 kommunikációs formátum melletti elköteleződése tekinthető², másrészt a könyvtár vezetőségének azon határozott törekvése, hogy e nemzetközi szabvány használatára alapozva intézményünk hatékonyabban tudjon bekapcsolódni a globális könyvtári szolgáltatások világába. Ez utóbbi tekintetében mérföldkőnek számít a WorldCat nemzetközi katalógushoz 2017 februárjában történt csatlakozásunk. A csatlakozási feltételek teljesítése – amelyet röviden úgy lehetne összefoglalni, hogy minden beadandó rekordnak teljes mértékben meg kellett felelnie a MARC21 szabvány követelményeinek – hosszas előkészítési szakaszt igényelt, amely során először felmértük katalógusunk nem MARC21-kompatibilis elemeit, majd a feltárt hibákat módszeresen, különböző online és kötegelt műveletekkel javítottuk. A javítások után kerülhetett sor az átadandó rekordok meghatározott szempontú leválogatására, majd az export-csomag kívánalmaknak megfelelő összeállítására és elküldésére. A feladat gyors és hatékony teljesítése a Szakinformatikai Osztály és a Gyűjteményszervezési Osztály munkatársainak szoros együttműködése révén valósulhatott meg.

1 The World's Largest Library Catalog, hozzáférés: 2018.06.30., <https://www.worldcat.org>

2 Az OSZK átáll a MARC21 használatára, hozzáférés: 2018.06.30., <http://www.oszk.hu/hirek/az-oszk-marc21>

1. Szerződéskötés és kapcsolattartás az OCLC³-vel

A könyvtárvezetősége 2016 augusztusában kezdte meg a tájékoztató WorldCat-csatlakozás feltételeiről. Az OCLC-vel való kapcsolatfelvétel és árajánlatkérés a T-Systems Magyarország Zrt. munkatársának közvetítésével indult el. Ezzel párhuzamosan igyekeztünk minél több tapasztalatot összegyűjteni a csatlakozás gyakorlati vonatkozásairól, ezért a Szegedi Tudományegyetem Klebelsberg Könyvtárának munkatársaival személyesen is konzultáltunk a WorldCat betöltéssel kapcsolatos tapasztalataikról. Végül 2017 februárjában került sor a csatlakozási szerződés aláírására. Tudomásunk szerint jelenleg egyetlen magyar könyvtárként van érvényes csatlakozási szerződésünk az OCLC-vel a „világkatalógusban” való részvételre. A szerződés értelmében a beadott rekordjaink integrálásra kerülnek a WorldCat rendszerével. A rekordok átadásának módjaként a batch-módú betöltést választottuk. A kezdő betöltés mellett évente háromszori frissítésre van lehetőségünk. Az előfizetés a teljes betöltés kezdetétől 1 éves időtartamra szól. A betöltési folyamat során egy számunkra kijelölt OCLC-adatspecialista segítette a munkánkat. A vele folytatott kommunikáció angol nyelven, elektronikus levelezés útján történt és történik azóta is.

2. A katalógus elemzése és a rekordok javítása

Az elemzések, illetve az arra épülő javítások megkezdése előtt döntés született arról, hogy könyvtári rendszerünkben (ALEPH)⁴ egy új, a MARC21 keretein belül maradó, de a meglévőnél differenciáltabb dokumentumtipológiát⁵ vezetünk be, amelynek alapja a rekordok formátuma lesz. Az új dokumentumtipológia alkalmazása – több más előnye mellett – lehetőséget adott a rekordállományunk kisebb egységekre való bontására, ezáltal jelentősen megkönnyítette az elemző és konverziós munkákat.

A rekordok elemzését - könyvtári rendszerünk szerviz-funkcióinak használatával - adatbázisból történő legyűjtésekkel és mezőkitöltésekkel végeztük. A legyűjtés eredményeit aztán – az elemzés meghatározott szempontjai szerint – különböző szerkesztőprogramokban⁶ csoportosítottuk és rendeztük. Az elemzések során a következő hibák feltárására törekedtünk: érvénytelen mezők/almezők, hiányzó tartalmak (például kötelező mezők/almezők, nyelvkódok, indikátorok), szabálytalanul ismétlődő mezők/almezők, hibás adatok (különösképpen LDR és 008 értékek). Ezek mellett az OCLC munkatársaitól is segítséget kértünk: egy kisebb próbacsomag elemzésére kértük őket. Visszajelzésük alapján akartuk ellenőrizni, hogy a tervezett javítások pontosan illeszkednek-e az előírásaikhoz.

A próbacsomag összeállításánál a fő szempontunk az volt, hogy minden gyűjteményből, rekordtípusból és dokumentumtípusból kerüljön rekord a halmazba. Az összeállításakor az adatbázis felmérése már megtörtént és az adattisztítás már jó ideje folyamatban volt, ekkor már tudomásunk volt a javítandó problémák többségéről. A csomag 1300 rekordból állt, melyeket az adott állapotukban adtunk be, nem javítottunk rajtuk egyesével. A nagyobb rekordhalmazt érintő hibák, korábbi feldolgozási szabályoknak

3 Online Computer Library Center, hozzáférés: 2018.06.30., <https://www.oclc.org>

4 A könyvtárban jelenleg az Aleph 21.1.9 verziója van használatban.

5 MTA KIK Katalógizálási szabályzat: dokumentumtipológia ALEPH-ben, hozzáférés: 2018.06.30., <https://doi.org/10.14755/MTAKIK.MARC21.2017>

6 Elsősorban a Microsoft Excel és Notepad++ programokat, illetve szerver-oldalon a 'grep' alkalmazást használtuk erre a célra.

megfelelő vagy külső katalogizálás során létrejött rekordok is átadásra kerültek, hogy lássuk, milyen hibatípusba kerülnek a betöltési folyamat során. A csomagba kerültek törzsgyűjteményi, keleti gyűjteményi, régi könyves és elektronikus könyvrekordok, folyóiratok, részcímes folyóiratfüzetek, évkönyvek, kéziratok, levelezések, ösnyomtatványok, disszertációk, mikrofilmek, számítógépes fájlok, audiovizuális illetve kartográfiai anyagok és részdokumentumok. Megpróbáltuk tartani a dokumentum-típusok arányát a teljes gyűjteményhez viszonyítva.

A leválogatott rekordokat jelöltük és korlátoztuk a hozzáférésüket az ellenőrzési folyamat végéig. El kívántuk kerülni, hogy az ellenőrzésre átadott rekordokban módosítások, javítások legyenek, amíg nem elemezzük az OCLC hibaüzeneteket. Készítettünk egy WCAT felhasználónevet, mellyel cseréltük a rekordokban az OWN (tulajdonos) mező PUBLIC tartalmát WCAT tartalomra. Így továbbra is meg lehetett tekinteni a leírásokat, de csak a WCAT felhasználónévvel lehetett módosítást végrehajtani azokban.

A próbacsomag rekordjait végül MARC formátumban és MARC-8 karakterkonverzióval adtuk át 2017. június 20-án.⁷ Az export-csomag összeállításának menetét később, a teljes export vonatkozásában fogjuk részletezni.

Az OCLC pozitív visszajelzései alapján megállapítható volt, hogy megfelelően jelöltük ki a javítások irányait, így hát folytattuk a megkezdett munkát és megszüntettük a korlátozást a jelölt rekordokon is.

Az elemzések eredményei határozták meg, hogy milyen javítási módszert kellett alkalmaznunk. A kisebb rekordcsoportot érintő és komplikáltabb javítási mechanizmust igénylő problémákat kézi erőbefektetéssel, online javították az illetékes feldolgozó munkatársak. Az online javítások mellett – a rekordok nagy száma miatt – főként kötegelte módosításokat végeztünk, amelyeket elsősorban az ALEPH szervizprogramjaival végeztük, de ezt kiegészítették a linuxos és windows-os környezetben végzett szerkesztési és programozási műveletek. Alapvetően az alábbi módszereket használtuk a javítások során:

- az Aleph javító szervizeivel végzett közvetlen adatmódosítás az adatbázisban
- az adatok kitöltése után módosítás editorokban, majd visszatöltés szervizzel
- az adatok kitöltése után módosítás ALEPH fix-programokkal⁸, majd visszatöltés szervizzel
- az adatok kitöltése után módosítás programozással (PHP), majd visszatöltés szervizzel.

Mind az elemzések, mind a javítások tekintetében jelentősen kibővítette a lehetőségeinket az, hogy egy programozási ismeretekkel rendelkező kolléganőnk is bekapcsolódott az adattisztítási munkálatokba. Az általa írt programgyűjtemény egy belső használatú weboldalon folyamatosan elérhető volt a javítások során.

7 Bár a kiértékelést nem akadályozta, az OCLC kérése az volt, hogy a később küldendő export-csomagokat már UTF-8 kódolással adjuk át.

8 A fix-programok olyan speciális szkriptek az Aleph-ben, amelyek meghatározott utasításkészlet felhasználásával egy 9 oszlopos paraméter-táblában hozhatóak létre.

NETWORKSHOP 2018

Minden javítási feladatot részletesen dokumentáltunk, szükség esetén többször újraterveztünk és próbarekordokon teszteltünk.

A fent ismertetett javítási módszerekkel több mint 1 millió rekordot módosítottunk, egy rekordot többször is. Összességében az 1 millió rekordunk 6,7 millió alkalommal módosult a javítások során közel egy év alatt.

3. Rekordok beadása

A tervezett javítások végeztével az ekkor közel 1.022.800 db rekordot számláló állományunkból kerültek leválogatásra az alábbi előkészített halmazok:

- törzsgyűjteményi anyag majdnem egésze
- régi könyvek, ősnymtatványok egésze
- Keleti Gyűjtemény könyveinek egésze
- Keleti Gyűjtemény kéziratok [Goldziher levelezés, ORI kéziratok (kivétel: perzsa, héber)]

Miután meghatároztuk, hogy rekordállományunk mely részei kerülnek átadásra, megkezdődhetett az export-csomag összeállítása. Az alábbiakban képernyőképekkel illusztrálva mutatjuk be az export-fájl létrehozásának menetét. Példáink az általunk használt ALEPH-rendszerből valók, de úgy gondoljuk, hogy azok – a könyvtári szoftverek működésbeli hasonlósága miatt – más könyvtári integrált rendszerek vonatkozásában is jól értelmezhetők.

Elsőként OPAC-műveletek elvégzésével egy halmazba integráltuk az átadandó rekordokat. Az alaphalmaz az „all_wcat” elnevezést kapta. Minden további procedúrát ezen az alaphalmazon végeztünk el.

Következő lépésként az alaphalmaz rekordjainak mezőit szekvenciálisan kitöltöttük. A kitöltés során az ALEPH „U39-DOC” rutinját használtuk annak érdekében, hogy a bibliográfiai adatok mellett a példányrekordok adatai is exportálásra kerüljenek.⁹

⁹ A példányadatok a 952-es mezőbe (amennyiben példányrekord nem volt a 852-es mező figyelembevételét kértük) kerültek átadásra. Megjegyzendő, hogy a 952-es mező a betöltés előkészítése során nagy segítséget jelentett az OCLC munkatársainak, de a mi projektünk esetében a lokális példányadatok közvetlenül nem kerülnek megjelenítésre a WorldCat-ben.

Rekordok letöltése géppel olvasható formában (print-03) - MTA01

* Input Fájl	all_wcat
* Output Fájl	all_wcat_kit
* Mező 1	ALL#
Mező 2	
Mező 3	
Mező 4	
Mező 5	
Mező 6	
Mező 7	
Mező 8	
* Formátum	ALEPH szekvenciális
* Konverzió Rutin	Nincs
* Kifejtő Rutin	U39-DOC
Karakter konverzió	Nincs
Törölt rekordok exportja?	Nem
Futás ideje	Ma
dátum:	óra
Könyvtár:	MTA01

1. ábra A print-03 ALEPH-szerviz űrlapja

Ezt követően a rekordok szerkezetét úgy módosítottuk, hogy kivettük belőlük az átadni nem kívánt mezőket¹⁰ és azok tartalmait. A kötegelt módosítást egy úgynevezett fix-programmal (wcat_atadas.fix) és egy konvertáló szerviz futtatásával (file-08) végeztük el.

MARC rekord fájl módosítása (file-08) - MTA01

* Input Fájl	tesztwcat_kit	
* Output Fájl	tesztwcat_kit_mod	
Fájl hibás rekordoknak	hibas_tesztwcat_kit	
* Feldolgozó folyamat	wcat_atadas.fix	
Futás ideje	Ma	
dátum:	óra	
Könyvtár:	MTA01	

Mehet

Korábbi nézet

Mégsem

Súgó

Naplóba hozzáad

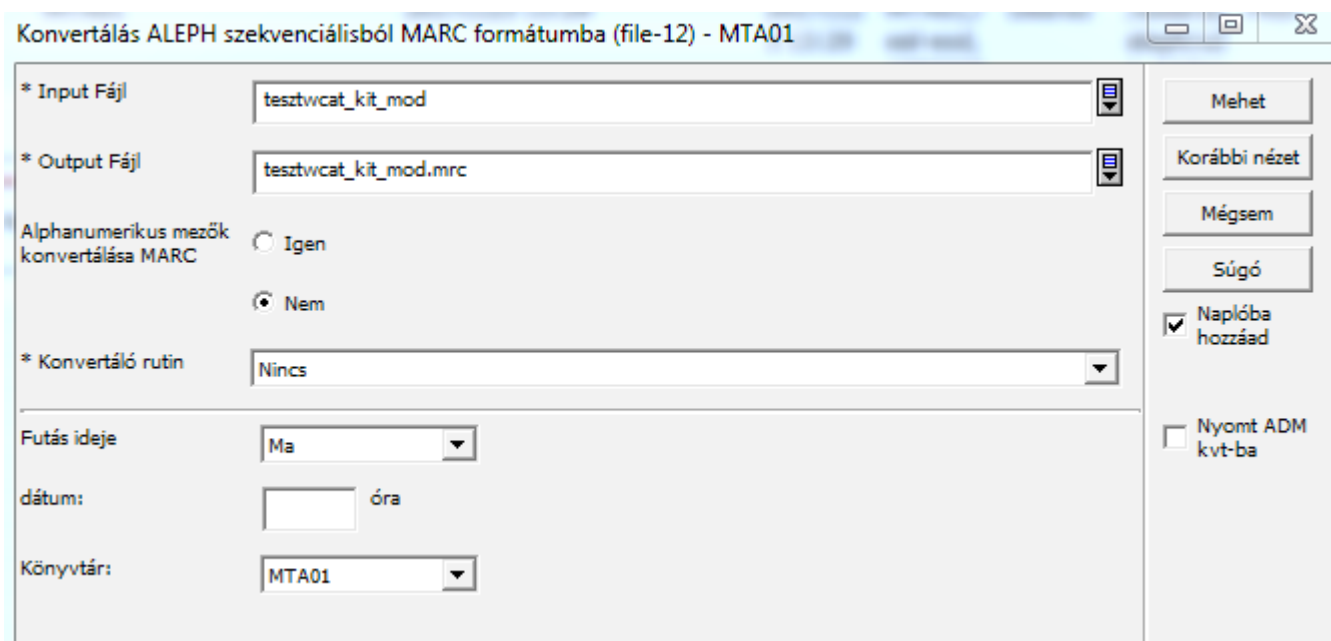
Nyomt ADM kvt-ba

2. ábra A rekordok konvertálása a file-08 szervizzel

¹⁰ Ezek elsősorban a helyi használatú mezők voltak (pl.: 59x, 9xx)

NETWORKSHOP 2018

A folyamat befejező lépéseként a rekordokat – az OCLC iránymutatását követve – szekvenciális formátumból MARC (ISO-2709) formátumba konvertáltuk úgy, hogy az alfanumerikus mezőket már ne tartalmazza az output fájl.¹¹



3. ábra Szekvenciálisból MARC formátumba konvertálás a file-12 szervizzel

Az elkészült export-fájlt az OCLC által megadott névkonvenció szerint egy megadott FTP-szerverre kellett feltöltenünk. A fájlnevnél tartalmaznia kellett a gyűjteményi azonosítónkat, a könyvtár OCLC szimbólumát és a feltöltés dátumát.

Az export-csomagot 2017. november 24-én küldtük meg az OCLC-nek. A csomag 775.919 db rekordot tartalmazott, amely megközelítőleg rekordállományunk 76 %-át jelentette.¹²

4. Kiértékelés és utómunkálatok

A betöltés kiértékelése szerint a beküldött állomány egész kis hányada (0,5 %) volt valamilyen szempontból hibás.¹³ Megkaptuk a hibás rekordok listáját és az OCLC-rendszerszámokat tartalmazó „cross reference” fájlt. A hibás rekordok listáját a könnyebb kezelhetőség érdekében áttettük Excelbe és úgy küldtük meg azt javításra a Gyűjteményszervezési Osztály munkatársainak.

11 Ilyenek például a speciális ALEPH-mezők (LKR, BAS, CAT stb.).

12 A próbacsomag 2018 januári újraküldésével további 257 db rekord került be újonnan a WorldCat adatbázisába, tehát összesen 776.176 rekordadtunk át a WorldCat-nek

13 A hibás rekordok is betöltésre kerültek a WorldCat-be, azonban esetükben megjelenítési és visszakereshetőségi problémákkal kell számolni.

AZ MTA KÖNYVTÁR ÉS INFORMÁCIÓS KÖZPONT GYŰJTEMÉNYE A WORLDCAT NEMZETKÖZI KATALÓGUSBAN

Incoming OCLC Number	Final OCLC Number	Local System Number	Exception Count	Data Sync Exception Description	Bib Validation Error
0	1014541772	000366106	1	1st \$6 in 1st 080 is too long.	Critical
0	1014684087	000840230	1	1st \$6 in 1st 100 is in the wrong format.	Critical
0	1014931386	000820917	1	1st \$6 in 1st 245 is too long.	Critical
0	1014993680	000518449	1	1st \$6 in 1st 245 is too long.	Critical
0	1014742868	000143251	1	1st \$6 in 1st 245 is too short.	Critical
0	1015062596	000833250	1	1st \$6 in 1st 245 is too short.	Critical
0	1014571466	000822376	1	1st \$6 in 1st 245 is too short.	Critical
0	1015025444	000616608	1	1st \$6 in 1st 245 is too short.	Critical
0	1014544113	000397610	1	1st \$6 in 1st 500 is too long.	Critical
0	1015092880	000514716	1	1st \$6 in 1st 500 is too long.	Critical
0	1014820636	000601882	1	1st \$6 in 1st 580 is too long.	Critical
0	1014611418	001041379	1	1st \$6 in 1st 600 is too long.	Critical
0	1014995444	000280321	1	1st \$6 in 1st 710 is too long.	Critical
0	1015130681	000945655	1	1st \$6 in 1st 830 is too short.	Critical
0	1014633193	000100817	1	1st \$6 in 1st 880 is in the wrong format.	Critical
0	1014710689	000960992	1	1st \$6 in 1st 880 is in the wrong format.	Critical

4. ábra Részlet az OCLC által küldött hibalistából

A későbbi visszakereshetőség szempontjából fontosnak tartottuk, hogy az átadás dátuma szerepeljen az érintett rekordokban. Létrehoztuk a helyi használatú WOC-mezőt, amelyet a beadás dátumával kötegetelt módosítással hozzáadtunk a rekordokhoz. Mivel a mező fontos háttér-információt hordoz a rekordokról, a feldolgozó felületről „elrejtettük” a mezőt, katalogizáláskor nem jelenik meg, tartalma nem módosítható. A WOC-mezőt ezenkívül indexeltük is, így a beadott rekordok bármikor könnyedén visszakereshetők az OPAC-ban.

A „cross reference” fájl alapján a beküldött rekordok 035-ös mezőjébe betettük a vonatkozó OCLC-rendszerszámot.

A beadást követően készült rekordjaink már teljes mértékben MARC21- és WorldCat-kompatibilisek, így a revízió áttesett, illetve a következő átadási csomagba jelölendő rekordok számára létrehoztunk egy szintén helyi WTC mezőt, melyet a feldolgozás, rekordjavítás végeztével a kollégák tesznek be a leírásokba.

A javítások nagy részét elvégeztük, de még nem mondhatjuk el, hogy teljes mértékben végeztünk. Néhány kisebb halmazt, dokumentumtípust elemeztünk, de javításukat, egységesítésüket későbbre halasztottuk. Amikor javításra kerül egy kisebb egység, ezeket is ellátjuk a WTC-mezővel. Az átadás előtt így ezek a halmazok automatikusan legyűjthetők lesznek.

NETWORKSHOP 2018

A következő beadási alkalommal az alábbi dokumentumtípusok ellenőrzését és előkészítését tervezzük:

- disszertációk,
- egy korábbi külsős feldolgozás rövidített leírásai,
- a legutóbbi retrokonverziós projekt leírásai,
- minden további kézirat leírása,
- mikrofilmrekordok,
- számítógépes fájlrekordok,
- zenei anyagrekordok,
- térképrekordok.

Ez várhatóan a teljes adatbázisunk közel 13%-a lesz.

5. Összegzés

Könyvtárunk WorldCat-hez történő csatlakozása jelentősen bővítette állományunk nemzetközi viszonylatban való láthatóságát és elérhetőségét. A csatlakozás megvalósítása több mint 1 éven keresztül erőforrásaink jelentős részét kötötte le, de a kiértékelés alapján elmondható, hogy az erőfeszítéseink elérték céljukat.

A javítási munkálatok során külön figyelmet szenteltünk a rekordlétrehozási folyamat optimalizálására és a hibalehetőségek csökkentésére, így gépi oldalról próbáljuk meggátolni a formailag téves adatbevitel lehetőségét. Folyamatosan aktualizált katalogizálási útmutató mellett a katalogizálói felületen magyar mezősűgők, példák és hibaüzenetek segítik a feldolgozó munkát. A törekvéseink eredményeképpen, az újonnan létrehozott rekordok automatikusan kerülhetnek a következő WorldCat betöltési csomagba.

A csatlakozási szerződésünk értelmében évente 3 alkalommal frissíthetjük rekordjainkat a WorldCat-ben. Az előkészítő munkálatok gördülékenysége érdekében 3 különböző típusú átadás mellett döntöttünk. Egy alkalommal az előző beadás óta létrehozott új rekordjainkat adjuk át, egy másik alkalommal az előzőleg beadott, de azóta módosított rekordjainkat küldjük, illetve a harmadik alkalommal az eltelt idő alatt törölt rekordjainkat vezetjük át a világcatalógusba.

Reményeink szerint akár már a közeljövőben megvalósítható lesz a teljes állományunk WorldCat-ben való megjelenése, majd annak folyamatos és gördülékeny aktualizálása.