

Az országos régi könyves adatbázis fejlesztéséről

Hegy Ádám

Országos Széchényi Könyvtár – SZTE BTK Könyvtártudományi Tanszék

2003 őszén, az Országos Széchényi Könyvtárban régi könyves “MOKKA”-ként indult projekt alapvetően két célt tűzött ki maga elé.¹ Az egyik, hogy az 1800 előtt készült Magyarországon található nyomtatványok lelőhelyjegyzékeként és egyedi adatainak nyilvántartásaként működjön. A másik szempont az, hogy felgyorsítsa a régi nyomtatványok feldolgozását azzal, hogy interneten keresztül elérhető katalógizáló felületet nyújt olyan gyűjtemények számára, ahol eddig semmilyen feltáró munka nem folyt. Természetesen mindezekhez társul még sok más elvárás is, mint például az, hogy a Kárpát-medence régi könyves közös katalógusaként működjön a projekt, valamint sok könyvtörténeti adattal (pl.: possessor, ex libris, kötés, bejegyzések) is bővíthető legyen. Gyakorlatilag ez egy olyan szolgáltatás lesz, amely közös katalógusként és katalógizáló felületként egyszerre fog működni, illetve könyvtörténeti kutatásokra lesz alkalmas. A következőkben azt tekintjük át, meddig jutottunk az eddigi fejlesztés során és a közeljövőben milyen fejlesztések várhatóak.

2003. decemberében egy országos felmérés keretében tájékoztunk a hazai gyűjtemények aktuális állapotáról. Ebben igyekeztünk megtudni a muzeális állományok nagyságát, a feltáró katalógusok jellemzőit és a számítógépes infrastruktúrát. Magyarországon ma összesen 234 olyan gyűjtemény van, ahol található valamilyen 1850 előtti állomány. Ezek közül kb. 20 könyvtárban összpontosul az országban található régi nyomtatványok 90%-a. A felmérést ezért csak a nagyobb intézmények körében végeztük el. Így összesen 118 helyre küldtük ki a kérdőívet és erre 49-en válaszoltak. Ebből a következő kép rajzolódott ki: a magyar könyvtárak összességében kb. 1 millió régi nyomtatvány található, amelyekből valamilyen eszközzel (cédula, nyomtatott katalógus, számítógép) kb. 80 % feltárt. Digitális formában viszont csak 5%-a. A számítógépes állományok közül 88%-a található valamilyen integrált könyvtári szoftverben, míg 12%-a egyéb adatszerkezetben (Folio, Word) van. Számunkra ez azt jelenti, hogy sok állományt szükséges lesz konvertálni, másrészt sok könyv tényleges leírása még várat magára. Lényeges viszont az tudni, hogy az OSZK-ban néhány éven belül igen jelentős mennyiségű HUNMARC rekord fog létrejönni a régi katalógusok konverziója során.

¹ Lásd erről: HEGYI Ádám: *Az országos régi könyves katalógus kialakításának lehetőségei = Egyházi könyvkultúra Magyarországon és Európában a 15. század végétől a 18. század közepéig*, konferencia, Sárospatak, 2003. október 1-3.,

Elég vegyes a kép a számítógépes infrastruktúrát illetően. Valamilyen internet kapcsolattal a vizsgált könyvtárak fele rendelkezik, legelterjedtebb az ADSL vonal. Közös katalógizáló rendszerben (MOKKA, THECA, VOCAL) 14 vesznek részt. A könyvtárak 82%-ában van legalább egy PC, amit könyvtári munkára használnak. Ezek karbantartásához, fejlesztéséhez 21 esetben van az intézményeknek saját rendszergazdájuk, programozójuk esetleg könyvtári informatikusuk. Igen kevés azok száma (27 gyűjtemény), ahol van olyan szakember, aki ért a régi nyomtatványokkal kapcsolatos teendőkhöz.

A körülmények tisztázása után a projekt működését a lehető legjobban igyekszünk a feltételek javítására kidolgozni. Ezzel párhuzamosan folyik egy részletes megvalósíthatósági tanulmány elkészítése és egy tesztadatbázis építése.

Először vegyük sorba, milyen szervezeti, együttműködési lépések történtek. A tesztelés időszakára az Országos Széchényi Könyvtár, a Debreceni Egyetem Egyetemi és Nemzeti Könyvtár, az ELTE Egyetemi Könyvtár, a Dunamelléki Református Egyházkerület Ráday Gyűjteménye és a Szegedi Tudományegyetem Egyetemi Könyvtár kialakított egy laza együttműködést, amelyben mintarekordokat adtak át az adatbázis számára. Ezt követően a különböző adatbázisokban készült rekordok szerkezetileg közös formára való átalakítása történt meg, mert rekordszerkezet helyett kötetlen folyószövegben gondolkodunk. Ehhez egy univerzális adattárolási szabványt, az XML-t választottuk. Erre azért volt szükség, mert ez egy olyan tárolási forma, amely sokkal könnyebben formálható, mint bármilyen más integrált könyvtári rendszer.

A fejlesztés jelenlegi szintjén ez a közös katalógus megvalósította azt, hogy eltérő rendszerekből kapott állományokat egyszerre kereshetővé tett. Interneten keresztül ez az OPAC a <http://www.eruditio.hu/lectio> címen érhető el. Mivel a munka tervezési fázisban van ez a felület egyelőre Mozilla internetes böngészőn keresztül használható teljes mértékben, mert csak ez a böngésző tudja a korszerű internetes XHTML és CSS szabványokat.. A közös katalógus kísérleti állományát az OSZK és az FSZEK antikva-katalógusa adja, amelyekhez virtuálisan kapcsolódik a szegedi, a debreceni könyvtár anyaga. Ez azt mutatja, hogy a rendszer már most képes arra, hogy különböző könyvtári adatbázisokban szabványosan (Z39.50) keressen. Továbbá az adatbázisból USMARC, XML, és szöveges formában is lehet hamarosan letölteni.

A fentebb felsoroltakon kívül a közös katalógus elsődleges állományát alkotják majd a 2004. december 31-ig az RMNY, RMK kötetek elektronikus változatai. Ezt követően

kerülnek be az adatbázisba a leendő résztvevők számítógépes rekordjai, miközben párhuzamosan folyik majd a feldolgozatlan könyvek rögzítése.

A későbbi igények érdekében az országos régi könyves katalógus adatszerkezete össze van kapcsolva az Eruditio projekt² adatbázisával. Ez azt jelenti, hogy technikailag megoldott nagyon bonyolult és mély keresések végzése is, ugyanis ha egy könyv szerzőjére vagyunk kíváncsiak, akkor nemcsak azt tudjuk meg, hány mű és hol található tőle Magyarországon, hanem azt is, hogy 1500-1700 között hol és ki olvasta azt.

A 2003-as szeptemberi kezdés óta a technikai eredményeken kívül szervezeti fejlesztések is történtek. Legfontosabb ezek közül, hogy elkészült a közös katalógizálási szabályzat első változata. Igen megkönnyíti munkánkat ebben, hogy 2004-ben az OSZK-ban ezzel párhuzamosan készül a régi nyomtatványok leírási szabványa.

A fentebb megnevezett intézmények együttműködésére alapozva elkezdjük kidolgozni egy jogi együttműködés kereteit. Ez lehetővé teszi azt, hogy minden olyan gyűjtemény, amely szeretne csatlakozni egyértelmű feltételekkel kapcsolódhasson a közös katalógushoz.

Ebben az évben terveink szerint további fejlesztések várhatóak a keresőfelületen, a katalógizálás szabályokban, és az együttműködés formájában. Ezeken kívül elkezdjük a katalógizáló felület megtervezését, az állomány fel-és letöltés megvalósítását. Nézzük meg ezeket részletesebben.

Az OPAC 2004. május 31-re végleges formát fog öltetni. Lehetőség lesz benne külön az országos régi könyves katalógus anyagában, illetve az Eruditio projekt állományával közösen keresni. Mindezt egyszerű és összetett lekérdezést lehetővé tevő úrlapon keresztül, ahol használhatóak Boole-operátorok és POSIX regurális kifejezések is. Ez a keresési felület nem maximálja a találatok számát, képes arra, hogy az összes találatot megjelenítse.

2004 végéig a közös katalógizálási szabályzatnak végleges formában kell tartalmaznia a bibliográfiai leírás szabályait. Valószínűleg hosszabb folyamat lesz az egyedi jellemzők (possessor, ex libris, kötés, stb) közös leírási szabályainak kidolgozása hasonlóan a személynevek, földrajzi nevek, testületi nevek egységes alakjaihoz. Ez utóbbiakhoz szerencsére sok hazai adatbázis már felhasználható, mint pl.: a *Clavis Typographorum Hungariae*,³ *Régi magyarországi szerzők*.⁴

A megvalósíthatósági tanulmány elkészültével az is el fog dőlni, milyen keretek között érdemes működtetni ezt a projektet. Az egyértelműen látszik, hogy szükséges a résztvevő

² *Bibliotheca Eruditionis*, Szeged, <http://www.eruditio.hu>, (2004. 02. 18.)

³ *Clavis Typographorum Hungariae* [adatbázis], Bp. OSZK

⁴ *Régi magyarországi szerzők*, összeáll. SZABÓ Géza, PINTÉR Gábor, Bp., 1990

intézményekkel együttműködési megállapodást készíteni. Kérdés csak az, hogy a MOKKA Egyesület részeként, vagy az OSZK struktúráján belül esetleg önálló szervezeti formában.

A közös katalógus azon az elven nyugszik, hogy régi nyomtatványok esetében szükséges az egyedi példánynyilvántartás, mert mindegyik könyv jelentős történelmi értékkel bír. Egyelőre a betöltött állományok egyes rekordjai darabonként vannak tárolva, az azonos rekordokat nem vontuk össze. A későbbiekben el kell viszont dönteni azt, hogy mikor szükséges azonos bibliográfiai adatok összevonása. Felmerülhet ennek igénye akkor, amikor egy mű kiadásának magyarországi példányaira vagyunk kíváncsiak, vagy akkor, amikor valaki az adatbázisban létező rekordhoz újabb példányadatokat akar hozzáírni.

Mivel az adatbázis XML-ben tárolja az adatokat, ezeknek az egyedi példányadatoknak a kezelése könnyebben megoldható, mint a MARC-ban. Ettől függetlenül cél, hogy az adatstruktúránk összhangban legyen a MARC holding adataival. Ennek megfelelően rekordfeltöltés kétféleképpen fog megvalósulni. Az egyik módszerrel a projekt saját katalógizáló felületén lehet egy könyvet feldolgozni. A másik lehetőség az, hogy az adott könyvtár rendszeresen átadja azt a közös katalógus számára. Bonyolulttá teszi ezt az, hogy az egyszer már feltöltött rekordokat szükséges lehet módosítani. Egyszerű viszont olyan szempontból, hogy bármilyen leírási adatot a Z39.50-nek köszönhetően könnyen cserélhetnek egymás között a könyvtárak.

Önálló problémát jelent a különböző rendszerek eltérő karakterkódolása. A központi katalógus UNICODE kódolást használ, amely alkalmas arra, hogy a világ összes nyelvének karaktereit tárolja, ezért bármilyen beérkező adatbázist, rekordot erre a karakterkészletre alakítunk át. Részben ehhez kapcsolódik az, hogy az eltérő leírási szokásokból következően a nem latin betűket egyesek transliterálják, mások nem.

Ez a projekt mindenképp tartalmaz egy önálló katalógizáló felületet. Erre azért van szükség, mert sok gyűjtemény nem rendelkezik számítógépes eszközökkel könyveik feltárására. Ez a szerkesztő jelentősen kisebb költségekkel beüzemeltető és fenntartható, mint egy integrált könyvtári rendszer, mert egy átlagos teljesítményű PC, modemesnél szélesebb Internet-elérés és egy böngésző program JAVA támogatással kell hozzá. Egyszerűen arról van szó, hogy online elérésű úrlapon dolgozhatnak a résztvevő könyvtárak az adatbázisba, ráadásul nem szükséges hozzá rendszergazda, mert ennek karbantartása nem az adott intézmény feladata. Természetesen ez nem azt jelenti, hogy azok a gyűjteményeknek is ezt kell választani, ahol már van valamilyen működő könyvtári szoftver.

Az adatbázis a most épülő SZTE Egyetemi Könyvtár számítógépes infrastruktúráját fogja használni. Ez a hardver és szoftver környezet elég erős és megbízható lesz projektünk

kiszolgálására, ugyanis a teljes számítógéppark hálózaton való megosztásával működik. A tagintézmények felőli informatikai háttér megteremtése komolyabb kihívást jelent, főleg abban az esetben, ahol semmilyen számítógép nem található. Ennek megoldására közös pályázatok benyújtása jöhet szóba. Ugyanúgy fontos olyan szakemberek bevonása a munkába, akik képesek ilyen speciális nyomtatványok könyvtári feldolgozására. Ilyen irányú továbbképzés ősztől Szegeden és az OSZK-ban fog megvalósulni.

Magyarországon található 1800 előtti könyvek teljes feldolgozása még így is több évtizedet fog igénybe venni, ennek ellenére ez a projekt igyekszik ezt a folyamatot meggyorsítani.