



## 600 000 könyv online: az Osztrák Nemzeti Könyvtár és a Google közös digitalizálási projektje

Az *Osztrák Nemzeti Könyvtár (ONK)*, együttműködésben a Google-lal, digitalizálja és online elérhetővé teszi történelmi könyvgyűjteménye teljességét. A könyvtár, amelynek történelme a XIV. századig nyúlik vissza, Ausztria egyik legfontosabb egyetemi könyvtáraként is funkcionál. Történelmi könyvgyűjteménye a világ egyik legnagyobbja, lévén, hogy a könyvtár a Habsburg Birodalom bírósági könyvtára volt. A XVI. században az egyik legkorábbi ilyen rendelkezés már ingyenes kötelepéldányt írt elő a gyűjtemény számára, jelentősen növelve ezzel az állomány nagyságát. A gyűjteményben ma kb. 44 000 XVI. századi könyv található, s részét képezi *Savoyai Jenő* 15 000 kötetes könyvtára is. Azon túl, hogy az egyik legnagyobb német nyelvű gyűjtemény, az ONK jelentős állománnyal rendelkezik a legkülönbözőbb kelet- és közép-európai nyelveken is.

Az ONK az elmúlt években több ambiciózus digitalizálási projektet hajtott már végre, s folyamatosan növelte digitális szolgáltatásai sorát. Analóg reprodukciós szolgáltatásukat felváltotta az olvasói kérésre történő digitalizálás; megnyitották történelmi lapgyűjteményük digitális adatbázisát (ANNO – jelenleg kb. 6,5 millió digitális oldalt tartalmaz), a jogi dokumentumokat tartalmazó ALEX gyűjteményt (kb. 3 millió digitális oldal), illetve a könyvtár kép- és grafikai archívumának digitális állományát (Bildarchiv Austria). Együttműködésben az *Osztrák Médiaközponttal (Österreichische Mediathek)* és az *Osztrák Tudományos Akadémia Hangarchívumával* több évig tartó munkával digitalizálták a könyvtár analóg hanggyűjteményét. Jogszabályi kötelezettségének eleget téve az ONK gyűjti az online publikációkat, s archiválja az osztrák weboldalakat is.

A folyamatosan gyarapodó digitális gyűjtemények és szolgáltatások következtében szükségessé vált egy, a digitális dokumentumok hosszú távú megőrzésére vonatkozó stratégia kidolgozása is. E

munka részeként 2008-ban külön szervezeti egységet állítottak fel a digitális megőrzéssel kapcsolatos feladatok megoldására; illetve több, ilyen témájú uniós projektben is aktívan részt vettek.

A Google-lal való együttműködés egy minden eddiginél nagyobb jelentőségű lépést jelent az ONK hosszú távú stratégiai terveinek megvalósításában. Az *Austrian Books Online* projekt keretében kb. 600 000, szerzői jog által nem védett, megjelenési évüket tekintve a XVI. század eleje és a XIX. század második fele között kiadott kötet kerül digitalizálásra, összesen mintegy 200 millió oldallal növelve a könyvtár digitális állományát. A digitalizált dokumentumok ingyenesen lesznek elérhetők a *Google Books* oldalain, illetve az ONK Digitális Könyvtárában.

### Google Books

A GoogleBooks szolgáltatásnak 2004 óta két fő forrása van: az ún. „partnerprogram” keretében a kiadók ajánlják fel könyveiket a Google számára digitalizálásra és online hozzáférhetővé tételre. A „könyvtárprogram” keretében a Google jelenleg mintegy 40 (köztük 13 európai) könyvtár állományát digitalizálja (többek között Harvard, Michigan, Stanford és Oxford egyetemi könyvtárai, illetve a *New York-i Közkönyvtár* is része ennek a körnek). Az ONK 2010 márciusában kötött szerződést a világcéggel, röviddel az után, hogy az *Olasz Kulturális Minisztérium* bejelentette az együttműködést a Google és a római, illetve a firenzei *Olasz Nemzeti Könyvtárak* között. Hasonló közös projektben vesz részt a Google-lal a *Holland Nemzeti Könyvtár* 2010 nyaratól; a *Cseh Nemzeti Könyvtár* 2011 tavaszától és a *British Library* 2011 júniusától. A német nyelvterület könyvtárai közül az ONK mellett a *Bajor Állami Könyvtár* is aláírta a Google együttműködési megállapodását, amelynek részeként

2007 óta már több mint 500 000 könyvet digitalizáltak.

A kiadókkal és könyvtárakkal folytatott közös munkák eredményeképpen a GoogleBooks oldalain (<http://books.google.com>) ma több mint 15 millió digitalizált könyv kereshető, amelyek közül mintegy 3 millió szabad felhasználású. (Az USA-ban folytatott gyakorlatától eltérően a Google Európában csak a szerzői jog által nem védett munkákat digitalizál.)

2010 decemberében a világcég „Google eBookstore” néven egy új felületet hozott létre, amelyről ma kb. 3 millió szabad felhasználású könyv tölthető le ingyenesen. (Az Egyesült Államokban több mint 200 000, a kiadókkal való együttműködésből származó könyv érhető el a Google-nál – ezek azonban csak meghatározott összegekért.)

Az ONK digitalizált könyvei a cég e-könyvszolgáltatásának ingyenesen elérhető gyűjteményét gazdagítják majd.

### **Az együttműködés mérföldkövei**

Egy akkora projekt, mint az *Austrian Books Online* – Ausztria eddigi legnagyobb PPP (public-private partnership) projektje – nehezen lenne kivitelezhető az ONK számára egy, a Google-hoz hasonló világcég részvétele nélkül, akár a feldolgozandó kötetek számát, akár a digitalizálás költségeit nézzük.

Míg a Google finanszírozza a teljes szövegű digitalizálást, a könyvek szállítását és biztosítását, a könyvtárat terhelik a könyvek kiválasztásának, előkészítésének és polcra való visszaosztásának költségei. A könyvtár költségein készülnek a metaadat-frissítések, folyik a rekordok minőségének ellenőrzése, az adatok tárolása és a digitális objektumok elérhetővé tétele is.

A költségek mellett az időfaktor a projekt másik kulcstényezője: a legutóbbi időkig a könyvtári digitalizálási projektek keretében évente jellemzően 5000–10 000 kötetet tudtak feldolgozni a könyvtárak. Ezt alapul véve egy, az *Austrian Books Online* projekthez hasonló kezdeményezés évtizedekig tartana...

Az együttműködés mindkét fél számára előnyös: a könyvtár a Google tevékenysége révén a világ bármely táján élő érdeklődők számára elérhetővé

tudja tenni gyűjteményét, míg a világcég jelentős mennyiségű és értéket képviselő állományhoz jut hozzá a közös munka révén, ez pedig jelentős mértékben segíti fő célja elérésében, nevezetesen, hogy kereshetővé és megtalálhatóvá tegye a világ összes könyvét.

A hozzáférhetővé tételen túl a digitalizálás fontos szerepet játszik a könyvek megőrzése, a könyvtári állományok védelme területén is, hiszen a digitális példányok létrejöttét követően az eredeti dokumentumok használata ritkábbá válik. A digitalizálás ugyanakkor a katasztrófahelyzetekre való felkészülés, a kulturális örökség meg- és átmentésének egyik eszköze is.

Az ONK döntése a PPP megoldás alkalmazásáról jól illeszkedik az *Európai Bizottság* kezdeményezéséhez is, melyet „*The New Renaissance*” címmel publikáltak 2011 januárjában, s amely nagy vitát generált az érdeklődők körében. A jelentés kiemeli, hogy míg a tagállamok elkötelezték magukat könyvtári, levéltári és múzeumi állományaik *Europeana*-n való elérhetővé tétele mellett, lényegében alig van közpénz az ehhez alapvetően szükséges digitalizálási projektek elindítására. A projektek pénzügyi hátterének megteremtéséhez jó megoldást jelenthet a köz- és magánszféra összefogása, vagyis a PPP együttműködések.

Az ONK és a Google megállapodása figyelembe vette a jelentés és az Európai Bizottság digitalizálási projektekre vonatkozó ajánlásait. A két fél a legfontosabb pontokban már a tárgyalások elején megállapodott:

- Kizárólag a szerzői jog által nem védett dokumentumokat digitalizálnak.
- A Google-lal való együttműködés nem kizárólagos; a könyvtár szabadon dönthet úgy, hogy ugyanazon állományrészeket más partnerrel is digitalizáltatja.
- A könyvtár minden digitalizált objektumot megkap egy példányban, s azokat nem kereskedelmi céllal elérhetővé is teheti felhasználói számára.
- Mindkét fél köteles a digitális objektumokat ingyenesen online elérhetővé tenni a projekt időtartamán túl is.
- Az ONK a digitális objektumokat más platformokon keresztül is hozzáférhetővé teheti (l. pl. *Europeana*).
- A könyvtár teljes autonómiát élvez abban, mely könyveket választja ki a projekt keretében történő digitalizálásra.

- A logisztikát és a digitalizálási eljárást folyamatosan nyomon követi és ellenőrzi a könyvtár állományvédelemért felelős részlege.
- A könyvtár azonnal felmondhatja az együttműködést, ha úgy érzi, az nem az elvárásainak megfelelően folyik.

A PPP együttműködés kulcspontja a projekt folyamatos és jól látható kommunikációja is. 2010 júniusában, a közös munka indulását jelző sajtótájékoztatóra a könyvtár összeállított egy dokumentumot az együttműködésre vonatkozó legfontosabb, leggyakrabban feltett kérdésekből és az arra adott válaszokból. Ezt, az angol és német nyelven kiadott dokumentumot rendszeresen frissítik. A klasszikus kommunikációs eszközök mellett számos új megoldást (pl. Twitter) is bevetettek a hatékony kommunikáció érdekében, melynek hatására a projektet folyamatosan nagy média- és állampolgári érdeklődés kíséri, s eddig javarészt pozitív véleményeket, értékeléseket kapott.

### **A projekt előkészítése**

A könyvtár általános témájú történeti gyűjteménye mellett digitalizálásra kerülnek a *Térképtár*, a *Ritka Könyvek és Kéziratok Tára*, a *Zenei Gyűjtemény*, valamint a *Színháztörténeti Múzeum Könyvtárának* szerzői jog által nem védett könyvei is. Része a projektnek a *Fidei Commiss Library* könyvállománya is, amely a *Habsburg-Lotharingiai Ház* egykori magánkönyvtára. Míg a gyűjtemény többi részéhez már léteznek metarekordok, ez utóbbi állományt most katalogizálják teljességében először.

A projekt hét munkacsomagra tagolódik, amelyek a teljes munkafolyamatot lefedik: könyvek mozgatása; metaadatok és katalógusok; állapotmegőrzés és -visszaállítás; adatok, rekordok és minőségellenőrzés; online hozzáférés; IT infrastruktúra; projektmenedzsment. A mai napig mintegy 70 könyvtáros dolgozott a feladaton, akik közül kb. 20 főt kizárólag a projektben foglalkoztatnak (könyvek mozgatása, könyvek előkészítése, állapotvisszaállítás, rekordok és minőségellenőrzés, szoftverfejlesztés és projektmenedzsment).

A projekt előkészítő munkálatai 2010 júniusa és decembere között zajlottak; végrehajtása lényegében a könyvtár valamennyi szervezeti egységét érinti, hiszen összhangba kellett hozni az intézményben futó más feladatokkal és projektekkel. Az előkészítő munka része volt a szükséges emberi erőforrások tervezése és a sikeres végrehajtáshoz

nélkülözhetetlen szervezeti változások előzetes lebonyolítása is.

Ez utóbbi elfogadását, elfogadtatását intenzív belső kommunikációs munka segítette, amely elsősorban a projekt céljaira, hasznosságára és nélkülözhetetlen voltára helyezte a hangsúlyt. A szervezeti egységeknek újra kellett gondolniuk és szervezniük munkafolyamataikat ahhoz, hogy biztosítani tudják a projekthez szükséges erőforrásokat. Ennek következtében több, korábban elkezdett belső projekt, feladatmegoldás háttérbe szorult, vagy akár le is állt.

Az előkészítő szakasz részeként a könyvtár munkatársai konzultációt folytattak más, a GoogleBooks által érintett európai és amerikai könyvtárak képviselőivel is: tapasztalataik nélkülözhetetlenek voltak az osztrák könyvtárosok számára.

Nagy figyelmet szenteltek a munkafolyamatok megszervezésének is, különös tekintettel a könyvek eljuttatására eredeti helyükről a digitalizálás helyszínére és vissza, illetve a dokumentumok digitalizálásra való előkészítésére. Költségvetési és hatékonysági szempontokat figyelembe véve az a döntés született, hogy a dokumentumokat szó szerint polcra polcra viszik digitalizálni. Kivételt ez alól csak az extra méretű vagy a rendkívül rossz állapotú könyvek, illetve a különleges formátumú (pl. kihajtható térképeket tartalmazó) dokumentumok esetében tettek.

### **A projekt végrehajtása**

Egy sikeres tesztszállítást követően 2010 végén elindult a dokumentumok nagy tömegű elszállítása a Google-hoz, s 2011 tavaszán a digitalizálás is elkezdődött. (Az első digitális objektumok már elérhetők a GoogleBooks-ban.)

A könyvek digitalizálásra való előkészítése egy speciálisan erre kialakított helyen folyt a könyvtárban. A feldolgozó könyvtárosok minden egyes kötetet kézbe vettek, ellenőriztek és vonalkóddal láttak el. Ez a vonalkód teszi lehetővé egyrészt a dokumentumok nyomon követhetőségét a szállítás során és a digitalizálási folyamatban, másrészt viszont ennek a segítségével kapcsolják össze a digitális objektumot az elektronikus katalógusban lévő metarekorddal. Bizonyos mértékű frissítésekre, javításokra minden könyv metarekordjában szükség volt, a vonalkódokat össze kellett kötni a metarekorddal, a több kötetes műveknél pedig

minden egyes kötethez külön metarekordot kellett készíteni. Végül, a digitalizálásra küldés előtt, minden egyes könyvet „ki kellett jelentkeztetni” az integrált könyvtári rendszerben.

Mindezeknek az alapvető, nélkülözhetetlen előkészítő feladatoknak az elvégzésére a feldolgozó könyvtárosoknak kötetenként nyolc perc állt rendelkezésükre. Azokat a bonyolultabb katalogizálási feladatokat, amelyekre ez az idő nem volt elég, más könyvtárosok végezték el. (Amikor pl. egy kötetben több mű volt összekötve, minden egyes művet azonosítva metarekordot kellett róluk készíteni, s hozzá kellett mindegyiket kötni a vonalkódhoz.)

A Fidei Commiss Library 100 000 kötetének előkészítése egy további különleges feladatot igényel: először négy könyvtárosnak katalogizálnia kell őket. Ez, azon túl, hogy alapvető követelménye a dokumentumok digitalizálásának, további haszonnal is jár: most először válnak kereshetővé az elektronikus katalógusban!

Az előkészítő feladatok részeként minden egyes könyvnek elvégezték az állapotfelmérését is, s szükség esetén meg is tisztították a köteteket, illetve meg is javították a kötéseket.

Maga a szkennelés a Google németországi digitalizáló műhelyében folyik. A kötetek odaszállítása előre megtervezett rend és előírások, szabályok szerint folyik. E szabályokat az is meghatározza, hogy a digitalizálásra kerülő kötetek java része védett, nemzeti kincs, így külföldre szállításukat az *Osztrák Szövetségi Örökségvédelmi Hivatal*nak kell engedélyeznie.

A visszaszállítást követően minden egyes könyvet kézbe vesznek és alaposan megvizsgálják a könyvtárosok, „visszaveszik” az integrált könyvtári rendszerbe is, majd visszakerülnek a polcra. A fenti folyamat átlagosan három hónapot vesz igénybe, a felhasználók ennyi ideig nem jutnak hozzá az egyes dokumentumokhoz.

### **Digitális objektumok és digitális könyvtár**

Ami a projekt IT-részét jelenti: naponta mintegy 95 000 digitális objektumot kell letölteni a Google szervereiről és automatikusan feldolgozni. Ennek megoldására az ONK kifejlesztette az *ADOCO (Austrian Books Online Download and Control)*

nevű alkalmazást, amely az adatletöltést, -ellenőrzést és -feldolgozást végzi. Az adatokat a Google egy gépi interfészen keresztül biztosítja; a kép- és OCR-fájlok ugyanolyan minőségűek, mint amelyeket a GoogleBooks felhasználói számára nyújtanak.

A projekt egyik különösen fontos feladata a minőség-ellenőrzés, melyet az ONK automatizált munkafolyamatok és reprezentatív, véletlenszerű egyedi ellenőrzések kombinációjával végez. Míg a korábbi digitalizálási projekteknél az ellenőrzési folyamatok javarészt manuálisan folytak, ebben az esetben nem az egyedi, hanem a típushibák automatikus kiszűrése a cél. A szoftver által fellelt hibákat aztán második körben manuálisan is ellenőrzik. Ezzel a szisztémával olyan prototípus-listát tudnak generálni, amely más, digitális megőrzést és tömeges digitalizálást célzó európai projekteknél is használható lesz.

A digitális objektumok tárolásához és hosszú távú megőrzéséhez nem volt elégséges az ONK háttérkapacitása. A döntés előkészítéséhez készült megvalósíthatósági tanulmány alapján úgy döntöttek, hogy egy külső tárhely igénybevétele helyett a könyvtár tárolókapacitását kell bővíteni.

A digitális objektumok azonosíthatósága és hivatkozhatósága érdekében bevezették a *URN (Uniform Resource Name)* nevű, *NBN (National Bibliography Number)* alapú rendszert, mellyel minden egyes címet egyedi azonosítóval látnak el.

2012 közepén az ONK digitális könyvtárában lépésről lépésre elkezd elérhetővé tenni a dokumentumokat. Először a felhasználók a könyvtár „Quick Search” szolgáltatásán keresztül férhetnek majd hozzá a digitális dokumentumokhoz egy speciális szoftver segítségével, amely a dokumentumok lapozhatóságát és nagyíthatóságát is lehetővé teszi. A felületről egy-egy oldal, vagy oldalak sokasága is letölthető lesz PDF-ben. A második ütemben (2013-ban) elérhetővé válnak a teljes szövegek is, s ekkorra várható a mobil alkalmazásokon (pl. iPhone, iPad, Android) való kereshetőség is. Az állomány természetesen elérhető lesz az *Európai Digitális Könyvtár*on és a nemzeti könyvtárak *TEL (The European Library)* felületén keresztül is.

A dokumentumok használatában a teljes szövegek kereshetővé tételének van különös fontossága. A felhasználók nemcsak a metarekordokban lesznek képesek keresni, hanem a dokumentumok belse-

jében, a könyvek tartalomjegyzékében és szövegében is.

A középtávú fejlesztési tervek között szerepel a teljes szövegek plusz információkkal való gazdagítása is olyan technológiák révén például, mint a névfelismerő rendszerek, amelyek a szövegekben is képesek nevet azonosítani, majd tezaurszokba kötni az adatokat. Különböző, számítógépek által támogatott módszerek segítségével kifejezetten a társadalomtudományokra kifejlesztett keresési stratégiákat is bele lehet építeni a szolgáltatásba. A Google 2010-ben bevezetett *N-Gram Viewer*-je az egyes szavak keresésének és előfor-

dulásának gyakoriságát is képes kimutatni a GoogleBooks-ban elérhető szövegtestekben.

A projektről további részleteket a <http://www.onb.ac.at/austrianbooksonline/> címen lehet elérni.

**/KAISER, Max: Putting 600,000 Books Online: the Large-Scale Digitisation Partnership between the Austrian National Library and Google. = Liber Quarterly, 21. köt. 2. sz. 2012. <http://liber.library.uu.nl/publish/issues/2011-2/index.html?000540/>**

*(Kovácsné Koreny Ágnes)*