

Az Országgyűlési Könyvtár „Digitalizálni, de...” c. konferenciáján tartott előadás cikk változata a TMT számára

REAL - Az MTA könyvtárának repozitóriuma

Holl András
MTA Könyvtára,
MTA Konkoly Thege Miklós Csillagászati Kutatóintézete

Az MTA Könyvtára széleskörű elektronikus könyvtári szolgáltatásokat nyújt - ezek között jelent meg 2009-ben egy repozitórium. Az e-könyvtári szolgáltatások között szereplő „elektronikus gyűjtemények” és a repozitórium sok szempontból a szolgáltatási spektrum átellenes oldalán találhatóak. Az előbbiek - mint például a Carbo kódex, a Körösi Csoma-, Bolyai vagy Radnóti anyag - egyedi, különösen értékes dokumentumokat tárnak a nagyközönség elé, igényes grafikai kivitelezésben, gondos feltárómunka nyomán. A repozitórium anyagai inkább a tudóstársadalomnak szólnak mint a laikus közönségnek, nagy mennyiségű anyagot tartalmaznak teljes szöveggel, inkább adatbázisként, mint kiállításként. A repozitórium egyes gyűjteményei az elektronikus katalóguson keresztül is elérhetőek lesznek majd szándékaink szerint.

A repozitórium létesítésének indítékai között szerepel az elektronikus könyvtári szolgáltatások, és ezen belül a teljes szövegű anyagok bővítésének szándéka, a hagyományos gyűjtemények - mint például a doktori dolgozatok - elektronikus változata megteremtésének igénye. Az OTKA 2009-ben olyan szabályozást fogadott el¹, miszerint minden általa támogatott kutatóprogram tudományos eredményeit a Nyílt Hozzáférés (Open Access, OA) szabályai szerint hozzáférhetővé kell tenni. Ugyan nem szabta meg, ezt milyen módon szükséges megtenni - valamely repozitóriumban való elhelyezéssel, OA folyóiratban, vagy hagyományos folyóiratban az OA lehetőség megvásárlásával - ám biztosítani kívánt egy olyan repozitóriumot, melyben a cikkek más lehetőség híján mindenképpen elhelyezhetőek. Ezért felkérte a Könyvtárat egy repozitórium felállítására, és a szolgáltatás kiépítését anyagilag is támogatta. Ugyanezt a repozitóriumot választotta az OTKA az általa támogatott kutatóprogramok pályázati zárójelentéseinek nyilvánosságra hozatalára és megőrzésére.

A létesítendő szolgáltatásnak az Open Archives Initiative kritériumainak megfelelően kellett felépülnie, alapvető elvárás volt a nemzetközi rendszerekhez való kapcsolódás képessége. A rendelkezésre álló támogatás mértéke, a Könyvtár adottságai alapján szabad szoftver alkalmazása mellett döntöttünk. A számos lehetőség közül megvizsgált EPrints, DSpace, CDS Invenio és a Fedora Commons közül egyszerűsége miatt az elsőt választottuk. EPrints -alapú repozitóriumok már működtek is az országban, - igaz, a szoftver korábbi verziójával - így támaszkodhattunk hazai segítségre is. Saját tapasztalatainkat, az új változat magyar fordítását mi is igyekszünk az érdeklődők rendelkezésére bocsájtani, ennek megfelelő fórumot teremt a HUNOR egyesület és az NIIF különböző rendezvényei: az IPSZILON szeminárium és a NETWORKSHOP.

Dedikált szerver számítógépet szereztünk be a szolgáltatáshoz, sőt, még egy tartalékgépet is beállítottunk, valamint kiépítettünk egy más feladatokra is felhasználható, bővíthető, nagy megbízhatóságú, SAN hálózaton keresztül elérhető tárolórendszert. A szerver Linux operációs rendszer

1 http://www.otka.hu/index.php?akt_menu=106&hir_reszlet=133

alatt működik. A repozitórium építése 2008-ban indult, a szolgáltatás 2009 nyarán indult meg.

Az OTKA gyűjtemény

Angol betűszó lett az új repozitórium neve: REAL (Repository of the Academy's Library). A megnyitás idején körülbelül kétezer dokumentumot tartalmazott - az utóbbi években lezárult OTKA projektek zárójelentéseit. Ezeket a jelentéseket az OTKA iroda helyezte el - és helyezi majd el a jövőben is - a REAL-ba. A tervek szerint ezek feltöltése automatikus lesz majd - az OTKA Elektronikus Pályázati Rendszeréből a SWORD protokoll segítségével kerülhetnek majd át a repozitóriumba. Tudományos folyóiratcikkek még csak nagyon kis számban találhatóak a REAL OTKA-gyűjteményében: az OTKA OA szabályozásának életbe lépte után elfogadott projektek még nemigen eredményeznek cikket.

A szakcikket majd a szerzőknek maguknak kell feltölteniük. Ezt a REAL webes felületén tehetik meg: először azonosítót kell készíteniük maguknak (hasonlóan ahhoz, ahogyan egy levelezőrendszerbe vagy közösségi oldalra regisztrálnának). Azonosító birtokában a fájlok feltöltése, a metaadatok megadása nagyjából tíz perces munkát jelent. Az EPrints szoftver itt is kínál egérutat: a DOI import funkcióval csak ezt az egyedi azonosítót kell megadni, a többi metaadatot a szoftver lekérdezi a CrossRef adatbázisából. Természetesen a könyvtár munkatársai szerzői feltöltés esetén is ellenőrzik, hogy a tétel a repozitórium gyűjtőkörébe tartozik, megfelelőek a metaadatai, és segítségen nyújtanak, probléma vagy kérdés esetén.

A REAL megtalálható az MTAK weboldaláról, az OTKA OA anyagai is hivatkoznak a címére². Az EPrints szoftver keresési lehetőségeket is kínál, azonban a REAL-ban tárolt tételeket az OAI-PMH protokoll alkalmazása miatt valószínűleg más keresők, aggregátorok segítségével fogják majd megtalálni az érdeklődők (jelenleg is kereshető például az nda@sztaki segítségével). A REAL háttértárat biztosít más projekteknek - mint például az OTKA-nak vagy az MTMT³-nek. Természetesen a repozitóriumok nemzetközi lajstromaiba (mint pl. a ROAR, OpenDOAR) is bejegyzésre került.

A disszertáció-gyűjtemény

Az OTKA-gyűjtemény után (ez, mint láthattuk, az OTKA projektek zárójelentéseit és tudományos eredményeit, többnyire folyóiratokban megjelenő cikket tartalmaz) a következő kollekció az akadémiai doktorok disszertációit tartalmazó gyűjtemény nyílik majd meg. (Ebben kerülnek elhelyezésre egyes kandidátusi dolgozatok is.) Ennek a gyűjteménykörnek az ötletét már 2003-ban felvetette Domsa Károlyné (Bánhegyi, 2003). REAL-d lett a doktori gyűjtemény neve, mely külön URL-en érhető el⁴.

A gyűjtemény feltöltése ez esetben is több csatornán keresztül történik. A nyomtatott, gépelt dolgozatok teljes digitalizálására nem gondolhatunk - ez a jelenleg rendelkezésre álló lehetőségekkel nem valósítható meg, de valószínűleg igény se lenne rá. A gyakorta gyenge minőségű gépelés, papír, kötés miatt digitalizálásuk nagy kihívást jelent, különösen úgy, hogy az eredeti könyvet is meg kell őrizni, felválni nem lehet. Ezért az igény szerinti digitalizálás stratégiáját választottuk: azokat az értekezéseket

² <http://real.mtak.hu>

³ Magyar Tudományos Művek Tára - <http://www.mtmt.hu/>

⁴ <http://real-d.mtak.hu>

szkenneljük, amelyeket olvasóink kerestek és keresnek. A jövőben fénymásolási igényeket már nem teljesítünk a hagyományos módon: az igényelt művet szkenneljük és kinyomtatjuk - és egyúttal PDF formában a repozitóriumban is elhelyezzük. A szerző maga is kérheti a doktori vagy kandidátusi dolgozatának digitalizálását - és be nem kötött, vagy felvágható példány beküldésével segítheti is azt! Ily módon évente nagyjából száz dolgozat digitalizálása történhet meg. A munkához olcsó, nem automatizált könyvszkennert használunk.

Az elmúlt két évtizedben keletkezett dolgozatok esetében még lehet esély arra, hogy a fájlok még fellelhetőek legyenek. Az első lépések ezek összegyűjtésére reménykeltőek: már majd félszáz dolgozatot sikerült beszereznünk. Az elektronikus levélben vagy optikai lemezen beküldött dolgozatokat a Könyvtár munkatársai feltöltik - bár természetesen itt is van mód arra, hogy a kutató saját azonosítót készítsen és maga töltse fel értekezését. Fontos a gyors cselekvés: az idő múltával az elektronikus változatok beszerzésének lehetősége csökken.

A jövőben az újonnan elfogadott dolgozatok feltöltését a Doktori Tanács Titkársága végezheti majd - eleinte manuálisan, később elektronikus rendszerük ugyancsak a SWORD-felületen (lásd Holl et al., 2009) keresztül tölthet majd fel anyagokat a REAL-d -be.

A dolgozatok mellett ebben a kollekcióban tárolunk járulékos dokumentumokat: tézisfüzetet, bírálatokat és azokra adott válaszokat is. Ez utóbbiak nem feltétlenül olvashatóak nyilvánosan.

Alapelvek, tanulságok

A repozitórium létrehozásakor - lefektettünk néhány alapelvet. Első gyűjteményeink tudományos jellegűek, melyekben valamilyen módon ellenőrzött minőségű anyagokat kívánunk tárolni. Az első gyűjtemény esetében az OTKA Iroda által elfogadott kutatási jelentéseket, vagy folyóiratokban közölt (bár nem feltétlenül referált) cikkeket, a másodíknál elnyert tudományos fokozatokhoz tartozó értekezéseket. A Könyvtárnak nincs módja az elhelyezett anyagok minőségének megállapítására, így külső szűrőkre hagyatkozunk.

Nem lenne kívánatos csak szórványos feltöltöttségű repozitóriumok kialakítása. Az OTKA-támogatású kutatóprogramok folyóiratcikkeinek csak egy része fog a REAL-ba kerülni - ám itt is törekszünk arra, hogy nagy mennyiségű cikket tartalmazzon. A doktori dolgozatok és OTKA-jelentések esetében viszont már az a célt is kitűzhetjük, hogy - egy adott dátumtól kezdve - teljes legyen a gyűjtemény.

Az első tapasztalatok azt mutatják, nagy gondot kell fordítanunk az elhelyezett dokumentumok minőségbiztosításának arra az oldalára, amiben a könyvtárnak már kompetenciája van: a megfelelő metaadatokra, hitelességre, formátumra. Fontos elérni, hogy a szükséges metaadatok rendelkezésre álljanak, és minél kevesebb hibát tartsanak. Nagy jelentőségű az esetlegesen meglévő egyedi azonosítók (mint pl. a DOI) tárolása. A metaadatok nem csak a repozitórium mögött álló adatbázisban kell szerepeljenek, de a PDF állományok metaadat-mezőiben, és maguknak a dokumentumoknak a szövegében is - hiszen a cikkek, dolgozatok a repozitórium-kontextusból kiszakíthatók: letölthetők és kinyomtathatók, s a fontosabb metaadatoknak akkor is kísérni kell a dokumentumokat. Felmerült valamiféle integritás-ellenőrző megoldások használata is.

A hosszú távú megőrzés szempontjából fontos az elfogadott állományformátumok megválasztása. A sokféle elfogadott formátum megkönnyíti a feltöltők dolgát - ám idővel egyre nagyobb terheket ró a könyvtárra. Mi csak a PDF formátumot engedjük meg - ennek is a az archiválásra tervezett PDF/A⁵ változatát részesítjük előnyben. A szolgáltatás és a megőrzés érdekei néhol ütköznek: az utóbbi szempontjából az egyszerűség és szabványosság kulcsfontosságú.

Szolgáltatások hálója

A REAL nem magában álló szolgáltatás - az OAI-PMH protokollt alkalmazó repositóriumokat eleve más szolgáltatásokkal való összekötésre tervezték. Ám nem csak a szüretelhetőségről van szó. A hazai tudományos/felsőoktatási/közgyűjteményi kör elektronikus szolgáltatásait oly módon kell hálózatba szervezni, hogy minél magasabb színvonalú, minél gazdagabb szolgáltatást nyújtsanak, legyen költséghatékony és mindez a kutatók, intézmények számára minél kisebb teherrel járjon. Repozitóriumok tárolhatják és szolgáltatathatják a dokumentumokat, bibliográfiai adatbázis - az MTMT - a szcientometriai adatokat (publikációs listákat és a publikációkra kapott idézeteket), a minősítési és bírálati eljárások során pedig a teljes szövegű dokumentumokra hivatkozó publikációs listák felhasználhatók. Aggregátorok és keresők - mint az nda@sztaki - szüretelhetik a repositóriumok metaadatait, nyújthatnak keresési lehetőségeket a felhasználóknak. A specializáció mellett redundanciák megteremtésére is szükség lehet: az NIIF például háttértárat vagy biztonsági mentési lehetőséget biztosíthat ezeknek a szolgáltatásoknak (lásd Stefán Péter cikkét ugyanennek a konferenciának az anyagában).

A szolgáltatás-háló elemeinek összekapcsolását szabványok alkalmazása segítheti, csak néhányat sorolva: a már említett OAI-PMH és SWORD, a DublinCore, az OAI-ORE⁶. Természetesen a heterogén, elosztott rendszereknél nehézségek is adódnak: nagyobb problémát jelent például a duplumszűrés (ehhez adhat segítséget az egyedi azonosítók használata).

A repositóriumok dokumentumokkal való benépesítése fontos kérdés. A REAL esetében szerencsére bizonyos dokumentumtípusoknál - OTKA zárójelentések, doktori értekezések - a folyamatos gyarapodás biztosított. A szakcikkek esetében már más a helyzet - itt még az Nyílt Hozzáférés OTKA által való megkövetelése sem jelent biztosítékot: hiszen nem minden cikk kerül majd a REAL-ba. Ezen túl fontos feladat a kutatók elhelyezési hajlandóságának növelése. Ez történhet technikai úton: mind a már említett, jelenleg is rendelkezésre álló DOI import funkció ajánlásával, mind SWORD alapú egyszerű feltöltést lehetővé tevő eszközök kifejlesztésével. De talán ezeknél is fontosabb a tájékoztatás (lásd: Holl 2010). Sok kutató nincs tisztában azzal, hogy az OA már szinte minden kiadó politikájával összeegyeztethető - nem kell publikálási szokásaikon változtatni, és az OA közzététel se jogsértéssel, se többletkiadással nem kell járjon. A REAL is ajánlja a dokumentum elhelyezése előtt a SHERPA RoMEO szolgáltatást a kiadói politika ellenőrzésére.

A REAL - az EPrints által biztosított funkcióval - nemcsak embargó periódus beállítását teszi lehetővé a kiadói érdekek védelmére, de zseniális technikai lehetőséget is nyújt a jog által biztosított lehetőségek kiaknázására: ha a feltöltő szerző megadta elektronikus levélcímét - ami természetesen a felhasználók számára nem látható - az embargós cikkekből is gombnyomással lehet a szerzőtől másolatot kérni (a

5 <http://www.digitalpreservation.gov/formats/fdd/fdd000125.shtml>

6 <http://www.openarchives.org/ore/>

szerzőnek sincs egyéb dolga min a repozitórium által küldött levélben az engedélyező hiperhivatkozásra kattintani).

Megfelelő szolgáltatás-hálózatot kiépítve a szerzőnek egy megfelelő repozitóriumban elhelyezett cikkével már nem kell többet foglalkoznia - az automatikusan bekerül publikációs listájába, tudományos fokozatszerzésnél, habilitációnál, pályázatbeadásnál vagy jelentéskészítésnél a befektetett munka megtérül.

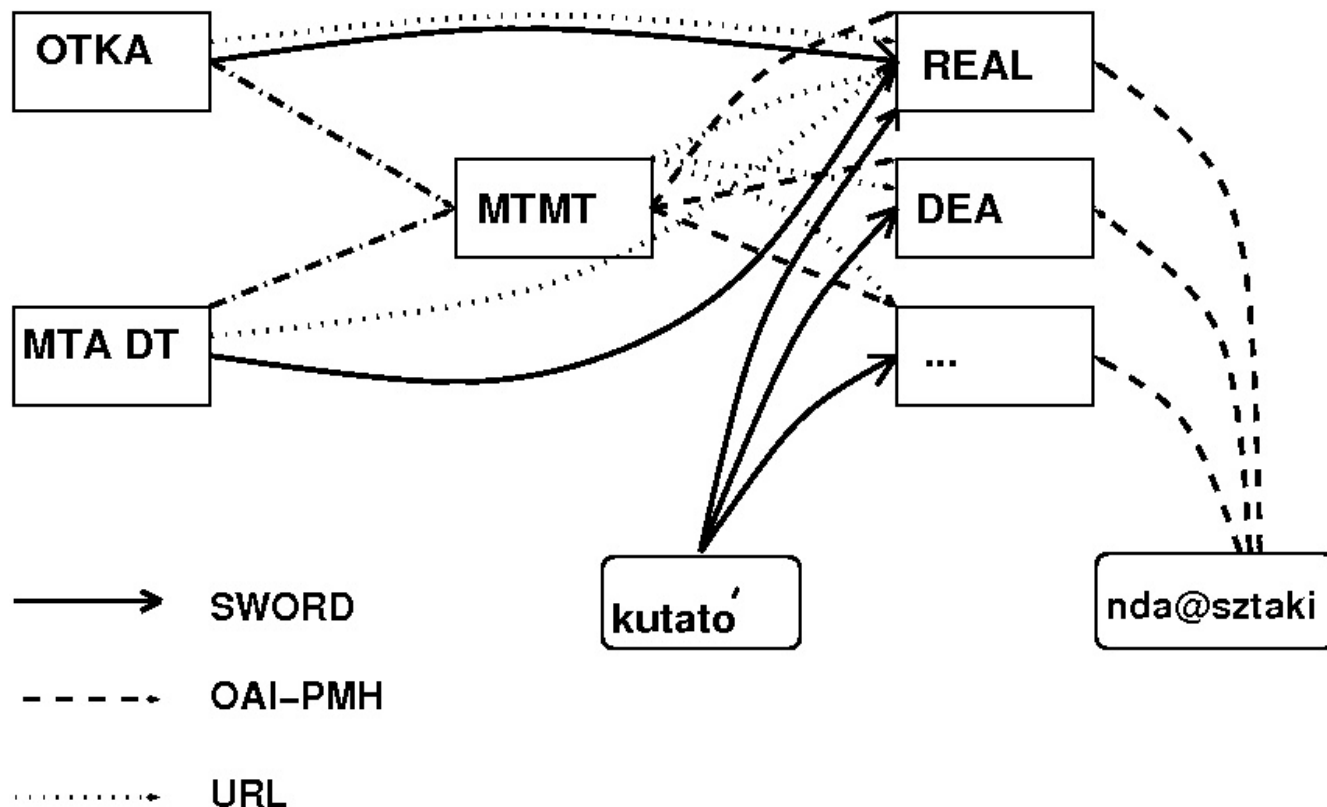
Irodalom

BÁNHEGYI ZS.: Nyílt Hozzáférés Kezdeményezés (Open Access Initiative) – Kitekintés és körkép. = Tudományos és Műszaki Tájékoztatás, 50. köt. 6-7. sz. 2003. p. 236-249.

HOLL, A., KARÁCSONY GY., HARANGI G., CSERBÁK M.: Teljes szövegű publikációk cseréje repozitóriumok között - a SWORD protokoll tesztelése az MTA Könyvtárában és a Debreceni Egyetem Könyvtárában. = NETWORKSHOP 2009.

<https://nws.niif.hu/ncd2009/docs/ehu/035.pdf>

HOLL A.: Nyílt hozzáférés a tudományos szakirodalomhoz - hazai fejlemények. = Magyar Tudomány, 171. köt. 1. sz. 2010. p. 58-61.



1. ábra Szolgáltatások hálózata