

# **REAL, az MTA Könyvtárának repozitóriuma - úton a minősített repozitórium felé**

Networkshop 2012, Veszprém

**Holl András**

*MTA Könyvtára*

*MTA CsFK Konkoly Thege Miklós Csillagászati Intézet*

*MTMT*

Röviden bemutatjuk az Akadémiai Könyvtár repozitóriumát és annak gyűjteményeit.

A repozitóriumok felállítása után az első lépés a gyűjtemény gyarapításának megszervezése. Amikor a repozitórium működik, már dokumentumok találhatóak benne, sőt, számuk folyamatosan gyarapodik - nem könnyű ezeket elérni - át kell tekinteni a minőségbiztosítási kérdéseket. (Célszerű ezekre már a repozitórium tervezése során gondolni, ám a gyakorlatban a minőségbiztosítás bevezetése gyakorta későbbre marad.) A minőségbiztosítás körébe tartozik a megfelelő szolgáltatási megállapodások kialakítása, a hosszú távú működést és a dokumentumok hosszú távú elérhetőségét, felhasználhatóságát garantáló tervek elkészítése, a kidolgozott eljárások bevezetése, dokumentálása, végül a repozitórium auditálása. Ki kell alakítani az nemzetközi projektekkal való esetleges együttműködés, és a hazai tudományos információs szolgáltatásokkal (pl. MTMT) való kapcsolat feltételeit.

A REAL még éppen csak elindult ezen az úton. A repozitóriumok hazai minősítésének rendszere kialakulófélben van - MTMT felkérte a kidolgozására a HUNOR konzorciumot. Ismertetjük a repozitóriumok minősítésére, auditálására kidolgozott külföldi eljárásokat, és a készülőben lévő hazai terveket.

A repozitóriumok létrehozása után biztosítani kell gyarapításukat, és meg kell teremteni működésük, a tárolt dokumentumok hosszú távú megőrzésének és szolgáltatásának feltételeit. Cikkünkben az MTA Könyvtárának Repozitóriuma, a REAL példáját követve mutatjuk be a repozitóriumok szervezésének, auditálásának kérdéseit. Az általános szempontokat és a REAL esetét egymás mellett bemutatva az MTMT számára készülő - a repozitóriumok minősítésére létrehozott HUNOR munkacsoport elé a szerző által beterjesztett - tanulmánytervezet menetét követjük.

\*\*\*

A REAL-t az MTA Könyvtára az OTKA felkérésére, OTKA támogatással állította fel. A repozitórium feladata lett az OTKA kutatási zárójelentések közzététele és hosszú távú megőrzése, valamint lehetővé kellett biztosítani a támogatott kutatások publikációinak elhelyezésére - amennyiben a kutatóknak más Open Access közzétételi mód nem állna rendelkezésére. Egyúttal fel kellett készülnie az MTA intézményi repozitóriumának szerepére is. Szerepelt a kitűzött célok között a nemzetközi projekteken való részvétel megteremtésének lehetővé tétele is. A REAL második gyűjteménye - a REAL-d - az MTA doktori ill. kandidátusi értekezéseinek tárolására állt fel 2010-ben.

A repozitórium feladatai megkövetelték, a rendelkezésre álló pályázati támogatás lehetővé tette a dedikált, tartalékolt szerverek alkalmazását, hosszú távon bővíthető

tárolórendszer beszerzését. A REAL az EPrints (ingyenes és szabad forráskódú) szoftvert alkalmazza.

Repozitóriumok létesítése előtt számba kell venni az igényeket, feladatokat, rendelkezésre álló erőforrásokat. Az első kérdés, szükség van-e repositóriumra?

Teljes szövegű dokumentumokat fel lehet tenni az intézményi, tanszéki vagy munkacsoport honlapra, meg lehet osztani a szakmai közönséggel tudományos közösségi rendszereken keresztül, mint a Mendeley vagy az Academia.edu. A repositórium elhelyezés többet kínál: szakmai (könyvtárosi) szűrést, szabványos metaadat-megosztást, hosszú távú biztonságot. A honlapon való elhelyezés, a tudományos közösségi rendszerek használata nem igazi Open Access. A repositórium alapvető kritériuma az OAI-PMH protokoll alkalmazása. Megvalósítható egyedi fejlesztéssel is - ám kiváló ingyenes, nyílt forráskódú repositórium szoftverek állnak rendelkezésre. Repositóriumot létesíteni, az Open Archives Initiative nyomán készült szoftvert alkalmazni érdemes!

Helyi jelentőségű szolgáltatás, várhatóan csekély terhelés esetén fel lehet állítani egy repositóriumot beruházás nélkül, egy más célra használt szerverre telepítve, a meglévő eszközök felhasználásával. A dedikált szerver megteremtése, a tartalékolás későbbi feladat lehet, a felhasználás növekedésének megfelelően.

A frissen felállított repositórium tesztelésére - a szabványos OAI-PMH kimenet működésének, megfelelő konfigurálásának ellenőrzésére az Open Archives Initiative is kínál eszközt. Fontos lépés az új repositórium bejegyzése a nemzetközi jegyzékekbe, az OpenDOAR-ba és a ROAR-ba. Az ellenőrzött, bejegyzett repositórium működésre kész!

\*\*\*

A felállított repositóriumot dokumentumokkal kell benépesíteni. Ha nem építjük ki a gyarapítási mechanizmusokat, a repositórium üres héj marad csupán. Ezen mechanizmusok közé tartoznak az elhelyezési kötelezettségek (angolul a *mandate* kifejezést alkalmazzák), az ellenőrzés és a jutalmazás illetve számonkérés. Nem kevésbé fontos a támogatás: az információs anyagok biztosítása és az intézményes segítségnyújtás. A repositórium elhelyezés szoftverekkel való segítése, a megfelelő automatizmusok, csatornák kiépítése nyújtja a legjobb eredményt.

A REAL alapgyűjteményének (OTKA gyűjtemény) benépesítéséről az OTKA Iroda által egy-két évente átadott kutatási zárójelentések gondoskodnak. A kutatások során létrejött publikációkból (cikkek, konferenci cikkek, könyvek, könyvfejezetek) kevés van ma még. Ezen a téren az első lökést az OTKA által a támogatott kutatóknak kiküldött levél adta meg ez év elején. A jövőben szükség lesz kidolgozni a publikációk OA elérési címének az OTKA zárójelentésekben való szerepeltetését, továbbá ki kell építeni a tájékoztatás, segítségnyújtás és ellenőrzés mechanizmusait.

A REAL-d folyamatos gyarapítását a folyamatban lévő akadémiai doktori eljárások anyagainak feltöltése biztosítja, amit az MTA Doktori Tanácsának Titkársága végez. Régi dolgozatok elhelyezésére is van mód: a Könyvtár szívesen fogad elektronikus formában rendelkezésre álló értekezéseket, illetve digitalizálási kéréseket és engedélyeket.

A REAL jelenleg több, mint 2800 tételt, a REAL-d majdnem 400 tételt tartalmaz.

\*\*\*

A repozitórium működhet egyedi szolgáltatásként, ám szóba jöhet az is, hogy repozitóriumok hálózatával működjön együtt egy országos vagy éppenséggel nemzetközi program keretében. A legegyszerűbb esetben csak a hálózatba kell "bejelentkezni" - ilyen a PhD dolgozatok repozitóriumait összefogó DART-Europe.

Más esetekben szükség lehet a tárolt metaadatok bővítésére, a "dobozból kicsomagolt" repozitórium-szoftver hangolására. Néhány elterjedtebb szabad repozitórium szoftver (DSpace, EPrints) számára foltozócsomagok alkalmazhatók az OpenAIRE kompatibilitás megteremtésére. Az OpenAIRE mellett felsorolandó itt az előzmény, a DRIVER, illetve az Europeana is - ezeknek is megvannak a repozitóriumokkal szemben támasztott kompatibilitási feltételei.

A hazai felsőoktatási és kutatási intézmények számára alapvető fontosságú a kutatási eredmények OA közzététele (ha másért nem, azért, mert az OTKA megköveteli), és szerepeltetése a nemzeti kutatási bibliográfiai adatbázisban, az MTMT-ben. Fontos szempont a kétszeres adminisztráció megszüntetése, a repozitóriumok és az MTMT adatbázis összekapcsolása. Ez kétféle irányból történhet: a repozitórium felöltés MTMT-n keresztül való lehetőségének megteremtésével, illetve a repozitóriumok MTMT által való rendszeres szüretelésével. Az előbbi esetben a repozitóriumoknak a SWORD protokollt kell alkalmazniuk, ami a repozitórium-szoftverek legújabb verzióiban már benne is lehet, vagy foltozócsomag segítségével telepíthető. Amennyiben a repozitóriumokban tárolt teljes szövegű publikációk adatait kívánjuk szcientometriai adatbázis építéséhez felhasználni, kulcsfontosságú a publikációk metaadatainak kiegészíthetősége és a duplumok kiszűrése (ezek a publikációk egyértelmű azonosításával oldhatóak meg), és a szerzők azonosítása. Az MTMT számára ezért fontos, hogy a repozitóriumok tárolják és szolgáltatassák a publikációk natív egyedi azonosítóit (pl. DOI vagy ISBN) és a szerzők egyedi azonosítóit (hamarosan lehetőség lesz az ORCID alkalmazására, addig is az MTMT-azonosítót vagy a ResearcherID-t). A tudományfinanszírozó szervezetek számára pedig a projektazonosítók használata lényeges. További, az MTMT szempontjából lényeges kérdés a tudományos közlemények megfelelő elválasztása a repozitóriumban esetlegesen tárolt egyéb anyagoktól.

A REAL egyedi publikáció-azonosító és OTKA projektazonosító mező t tartalmaz, a szerző azonosító mező bevezetését tervezzük. A DOI szerepeltetésére ösztönzést ad az EPrints DOI import funkciója - a DOI megadásával a többi metaadat begépelése megspórolható, azt a rendszer a CrossRef adatbázisából importálja. A SWORD protokoll implementálását toldalék (plugin) telepítésével megoldottuk (Holl et al., Workshop 2009). Az egyedi azonosítók megjelenítését a PMH kimeneten, illetve a Qualified Dublin Core szerinti leképezésüket a létrehozandó MTMT szabványok szerint meg kell oldani.

\*\*\*

Ahhoz, hogy a repozitórium hosszú távon biztosítani tudja küldetését, a dokumentumok megőrzését és szolgáltatását, már a tervezés fázisában célszerű néhány szempontot végiggondolni. Persze, elképzelhető repozitórium indítás más stratégiával is: előbb installálunk egy ingyenes szoftvert, kísérletezünk vele, beindítunk egy kísérleti szolgáltatást, és ha beválik, akkor kezdjük csak kialakítani a hosszú távon megalapozott szolgáltatás feltételeit.

Milyen kérdéseket kell megfontolnunk? Vannak triviális szempontok, amelyekre a válasz egyes helyzetekben adódik, mégis fontosnak tartjuk megemlíteni ezeket is.

A repozitóriumnak be kell épülnie a működtető szervezetbe, meg kell jelennie az alapító okiratban, küldetésnyilatkozatban vagy statútumban. Be kell tagozódnia az intézmény szervezetébe, felelős vezetővel, munkatársakkal. (Ezek a követelményeket - a későbbiekben említettekhez hasonlóan - a kitzűzött feladatok, célok függvényében kell kielégíteni. Kezdetben, vagy csekély követelmények esetén elegendő a repozitóriumhoz kapcsolódó feladatokat más feladatokat ellátó munkatársak munkaköri leírásába beilleszteni.) Számolni kell azzal, hogy a kezdetben egy meglévő szerverre ültetett repozitórium-szolgáltatás a későbbiekben tárolóhely-bővítést, saját szervert igényelhet, a kezdeti támogatásból beszerzett eszközök elöregednek, és előbb-utóbb cserére lesz szükség.

Ki kell alakítani, milyen mutatók figyelemmel követésére (pl. dokumentumok száma, felhasznált tárolóhely nagysága, regisztrált felhasználók száma, letöltési statisztikák) követésére, naplózására és jelentésére van szükség.

Szükség van vészforgatókönyvre különböző elképzelhető problémák - pl. szerver meghibásodás - esetére. Az adatbiztonság megteremtéséhez, az adatmentésekhez is tervet kell készíteni, eljárásokat kialakítani a tárolt adatok bithűségének ellenőrzésére. Évtizedes távlatokra előretekintve fontos az adatformátumok kezelésének tervezése - előre el kell dönteni, milyen formátumú dokumentumokat fogad be a repozitórium, hogyan kísérik figyelemmel az elfogadott formátumok avulását, elképzeléseket kell kialakítani az esetleg szükségessé váló adatmigrációról. Nem szabad úgy repozitóriumot létesíteni, hogy ezeket a kérdéseket át ne gondolnánk, és amit lényegesnek ítélünk, le ne írunk. Elképzelhető külön-külön katasztrófaterv, adatbiztonsági dokumentáció, migrációs stratégia, stb. létrehozása, de ez esetben is a feladat diktálja, milyen eszközöket kell alkalmaznunk. Adott esetben elegendő egy-egy bekezdést szánni a fontos kérdéseknek egy általános dokumentációban. A felsorolt szempontok csak példák, a lajstrom korántsem teljes.

A REAL számára a következő dokumentumok készültek el az induláskor és azóta:

*Koncepció és rendszerspecifikáció; Gyűjtő kör, adatformátum, metaadat és adatpolitikai dokumentáció; Telepítési és üzemeltetési útmutató (szoftver- és hardver-üzemeltetés); Üzemeltetési dokumentáció (szervezeti feltételek); Szolgáltatási megállapodás; Szerzői jogi információk; GYIK; Feltöltési információk; Open Access segédlet.* Ezek egy része nyilvános, a REAL oldalain elérhető. A dokumentációk többsége rövid, esetenként csupán néhány mondatos vagy néhány bekezdéses.

Célszerű nek láttuk az elfogadott dokumentumformátumok körét a PDF-re korlátozni, és a PDF/A formátumot ajánlani - ez a döntés a későbbiekben jelentős erőfeszítéseket takaríthat meg a dokumentum-migráció terén, és drasztikusan leegyszerűsíti az adatformátum-politikát.

Számos eljárás készült a repozitóriumok minőségbiztosítására és auditálására világszerte (Sierman, 2009). Népszerű ön-átvilágító segédeszköz a Drambora Toolkit. Az utóbbi évek egységesítési törekvései nyomán egy háromlépcsős rendszer látszik kibontakozni: az *European Framework for Audit and Certification of Digital Repositories*. Meglévő minősítési rendszerek kombinálásával született, az első lépcsőt a DANS Data Seal of Approval minősítése jelenti, a másodikat az ISO

szabvány szerinti önátvilágítás, a harmadikat pedig egy ugyancsak ISO szabványnak megfelelő külső minősítő szervezet által való auditálás.

A formális minőségbiztosítási eljárások alkalmazása munkaigényes. Az eljárások szempontjainak tanulmányozása viszont még akkor is hasznos, ha az auditálást (egyelőre) nem kívánjuk elvégezni.

A REAL esetében ez évben tervezzük a minősítési eljárások lépéseinek áttekintését. A követelmények egy része a repozitóriumon túl más digitális gyűjtemények üzemeltetésénél is alkalmazható lehet.

## **Irodalom**

Holl A., Karácsony Gy., Harangi G., Cserbák M., 2009, 'Teljes szövegű publikációk cseréje repozitóriumok között - a SWORD protokoll tesztelése az MTA Könyvtárában és a Debreceni Egyetem Könyvtárában', Networkshop előadás  
<http://www.konkoly.hu/staff/holl/sword/sword.pdf>

Sierman, B., 2009, 'Trust, but verify', APA 2009 konferencia előadás  
[http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/02/07\\_BarbaraSierman\\_KB\\_241109.pdf](http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/02/07_BarbaraSierman_KB_241109.pdf)

## **URL-ek**

REAL

<http://real.mtak.hu>

REAL-d

<http://real-d.mtak.hu>

OAI Data Provider Validation and Registration

<http://www.openarchives.org/Register/ValidateSite>

OAI Repository Explorer

<http://re.cs.uct.ac.za/>

DART-Europe

<http://www.dart-europe.eu/>

Europeana Technical Requirements for Data Providers

<http://pro.europeana.eu/technical-requirements>

OpenAIRE kompatibilitás

<http://www.openaire.eu/hu/support/toolkits/repository-managers>

Drambora Toolkit

<http://www.repositoryaudit.eu/>

European Framework for Audit and Certification of Digital Repositories

<http://www.trusteddigitalrepository.eu/>