

A könyvtár új gyűjteménye: a digitális repozitórium

The New Collection of the Library: The Digital Repository

Holl András

informatikai főigazgató-helyettes, Magyar Tudományos Akadémia Könyvtár és Információs Központ, Budapest
holl.andras@konyvtar.mta.hu

Absztrakt

Bemutatjuk a 200 éves Akadémiai Könyvtár modern digitális gyűjteményét, a REAL repozitóriumot. Felvázoljuk a szolgáltatás történetét, és az egyes gyűjteményi egységek ismertetése mellett foglalkozunk a repozitóriumok jelentőségével a kutatók és könyvtárak szempontjából.

Abstract

We present the REAL repository, the modern digital collection of the 200-year-old Library of the Academy and we outline the history of the service. In addition to describing the individual collection units, we discuss the significance of repositories from the perspective of researchers and libraries.

Kulcsszavak: digitális repozitórium, REAL, nyílt hozzáférés

Keywords: digital repository, REAL, open access

A repozitórium egyszerre a könyvtár digitális megjelenési formája és a tudományos kommunikáció új csatornája. Egyszerre könyvtári szolgáltatás és része a kutatói eszköztárnak. A könyvtári digitalizáció fontos komponense, ugyanakkor a hagyományos könyvtári működésmódtól némileg idegen szervezet, ahol nem az „olvasó”, hanem az olvasójeggyel nem is rendelkező „kutató” a partner. Olyan dokumentumtár, amelybe nem feltétlenül a már megjelent művek, hanem esetenként a megjelenés előtt álló kéziratok vagy könyvterjesztésbe sohasem kerülő dokumentációk kerülnek, a kutatási adatokról nem is beszélve. Ugyanakkor minőségbiztosított dokumentumtár: a preferált közleményállapot a megjelent kiadói pdf-fájl vagy legalábbis a közlésre elfogadott kézirat, ezen túl a repozitóriumba kerülő tételek könyvtárosi ellenőrzésen mennek át. További fontos követelmény a nyílt hozzáférés biztosítása: ugyan lehetőség van korlátozásra vagy a hozzáférés megnyitásának késleltetésére, ez azonban csak kivételes esetben alkalmazandó.

A REAL (Repository of the Academy's Library), az Akadémiai Könyvtár – és ezáltal a Magyar Tudományos Akadémia – repozitóriuma 2008-ban jött létre az OTKA (Országos Tudományos Kutatási Alapprogramok) felkérésére. A szolgáltatás a következő évben indult, és rövidesen a hazai repozitóriumi hálózat meghatározó elemévé vált. Tízezer nyi regisztrált felhasználó, 300 000-nél több tétel, 3,5 milliós éves letöltés,

valamint hazai szinten a tárolt dokumentumokra mutató linkek száma alapján kimagasló láthatóság jellemzik.

A repozitóriumi tételek leírásában alig több mint 60 000 utalás szerepel a könyvtári katalógusra, ami azt jelenti, hogy a könyvtár papíralapú dokumentumainak csak néhány százalékaról van digitális másolat a repozitóriumban, és az ott tárolt tételek ötöde származik a könyvtári állomány digitalizálásából. (Az állománydigitalizálás korábban kiszervezve, kampányszerűen történt, mára a saját digitalizáló műhely biztosít kisebb volumenű, ám folyamatos gyarapodást.) A REAL digitális dokumentumai többségükben nem találhatók meg a könyvtár raktárainak polcain.

A REAL alapozza meg az Akadémiai Könyvtár részvételét egyes hazai és nemzetközi projektekben, szervezetekben, esetenként pályázati bevételeket is biztosítva. Az *Akadémikusok* adatbázis a REAL-ra alapozva mutatja be az Akadémia tagjaihoz kapcsolódó dokumentumokat („Akadémikusok. A Magyar Tudományos Akadémia tagjai”).

A nyolc repozitóriumi gyűjtemény

REAL: a kutatási gyűjtemény

Ebbe a gyűjteménybe regisztrálhat minden magyar kutató, de regisztráció nélkül is feltölthet az MTMT-ből. Idekerültek az induláskor az OTKA-projektbeszámolók, ezt a gyűjteményt használhatták az open access közzétételi kötelezettségük teljesítésére az MTA kutatóhálózatában dolgozók, szintén ide kell feltölteniük az MTA pályázatai hasznélvezőinek támogatott kutatásaik publikációit. A már megjelent közleményeket vagy közlésre elfogadott kéziratokat várjuk, hogy a kutatások eredményei nyilvánosak lehessenek (előbb-utóbb). A REAL nyilvános, digitális könyvtár, nem pályázati beszámolási rendszer. (Persze indokolt esetben a feltöltött közlemények lehetnek meghatározott vagy határozatlan időre zárolva.) A könyvtárosok is töltenek fel ebbe a gyűjteménybe digitalizált folyóiratokból kivágott cikkeket – többnyire a néhai akadémikusok munkáit.

Ez a gyűjtemény teremti meg a biztonságos hátterét a könyvtári folyóiratplatform működtetésének, és a DOI Iroda által biztosított egyediazonosító-szolgáltatásnak. (Ugyanakkor ezek a szolgáltatások hozzájárulnak a REAL-gyűjtemények folyamatos gyarapításához.)

REAL-J: a folyóiratszámok vagy -kötetek gyűjteménye

Idekerültek a 2010-es évek közepén a könyvtár digitalizálási programja keretében elkészült folyóirat-kötetek: mind az Akadémia folyóiratai, mind olyan régi folyóiratok, amelyek a könyvtár gyűjteményében megtalálhatóak voltak. Az MTA Könyv- és Folyóirat-kiadó Bizottsága által támogatott folyóiratok digitális kópiái is bekerülnek ebbe a gyűjteménybe, amennyiben rendelkezésre állnak füzetenként. Azok a folyóiratok, amelyek füzeteket már nem képeznek, a REAL-alapgyűjteménybe archiválhatnak,

ide csak mutatórekordok kerülnek, elvezetve egy adott évfolyam különálló cikkeihez. Elhelyezünk ide másolatokat a könyvtár Open Journal Systems rendszerében megjelenő folyóiratok füzetéből is.

Általában nyitva áll a gyűjtemény a hazai kiadású tudományos folyóiratok előtt, segítve biztonságos archiválásukat és nagyobb láthatóságot biztosítva cikkeiknek. Számos tudományos folyóirat jelzi honlapján a REAL archiválást, és vannak intézmények, amelyek megfelelő saját tárhely hiányában a REAL-J-ben „tartják” folyóiratukat.

REAL-EOD: a könyvgyűjtemény

Ez a gyűjtemény az „eBook on Demand” projekt támogatására jött létre, de idekerülnek a könyvtári digitalizálási programokban elkészült könyvek is, mint az Akadémiai Kiadó 8000 kötetes anyaga. Ezek többsége az Akadémia 200 éves évfordulójáig zárolt volt. Teljesen feltárt – tanulmányonként leírt – kötetek is megtalálhatóak ebben a gyűjteményben az Akadémia alapításától kezdve.

REAL-R: régi könyvek

Viszonylag kis gyűjtemény az Akadémia alapítása előtt kiadott nyomtatott könyveknek.

REAL-MS: kéziratos anyagok

A könyvtár különgyűjteményeinek, a Kézirattárnak és a Keleti Gyűjteménynek a kéziratos anyagait tesszük közzé ebben a gyűjteményben. Idekerültek a digitalizált gyászjelentések, Arany János hivatali levelezése és különféle levéltári anyagok külső intézményekből is, mint a Neumann János Számítógép-tudományi Társaság Informatikatörténeti Fóruma vagy a Nemzeti Információs Infrastruktúra Fejlesztési Program.

A gyűjteményben megtalálhatóak kéziratos kódexek (mint például a *Rohonci kódex*) és olyan különleges kincsek, mint a *Bori notesz*. A Keleti Gyűjtemény arab, héber, mongol, perzsa, tibeti és török anyaga is e gyűjteményben található, részben szabadon olvashatóan, részben zárolva.

REAL-D: disszertációk

2010-től kötelezően idekerülnek az MTA doktori dolgozatok, de a korábbi tudomány doktora és a kandidátusi disszertációk közül is digitalizálásra került néhány. Amennyiben disszertációra másolási kérés érkezik, azt már digitalizálással teljesítjük. A kurrens dolgozatokat az MTA Doktori Tanács Titkárságának munkatársai töltik fel, további pályázati anyagok (például a bírálatok és az azokra adott válaszok) kíséretében.

REAL-PhD

Amennyiben egy egyetem nem tudja megoldani a PhD-dolgozatok digitális archiválását és nyilvános szolgáltatását, ezeket az „árva” dolgozatokat elhelyezheti hozzánk. A gyűjtemény létrejöttétől a mai napig az Andrassy Gyula Budapesti Német Nyelvű Egyetem, az Evangélikus Hittudományi Egyetem, a Kaposvári Egyetem, a Károli Gáspár Református Egyetem, a Liszt Ferenc Zeneművészeti Egyetem, az Országos Rabbiképző – Zsidó Egyetem, a Pannon Egyetem, a Pázmány Péter Katolikus Egyetem, valamint a Magyar Testnevelési és Sporttudományi Egyetem archivált ebbe a gyűjteménybe. Amennyiben valamelyik intézménynek időközben saját repozitóriuma lett, a dolgozatok egy másolata oda is bekerült, de a REAL-ban is megmaradt. A gyűjteményben egy-egy szórványdolgozat szerepel más (esetenként külföldi) egyetemekről is, a szerzőjük szándéka szerint. A megállapodás alapján, szisztematikusan idearchiváló egyetemek esetében a feltöltést nem a szerző, hanem a doktori iskola munkatársa végzi.

REAL-I: képek (fényképfelvételek, könyvillusztrációk, festmények, térképek, képeslapok, kisnyomtatványok)

Legfiatalabb egységünk a képgyűjtemény. Itt megtalálhatóak például Stein Aurél digitalizált fényképfelvételei és a Szántai Lajos-féle térképgyűjtemény anyaga is. A könyvtár kiállításainak képanyagát is elhelyezzük ebben a gyűjteményben.

A kutatói eszköztár része

Nem a könyvtárak, a könyvtári szoftvereket gyártó cégek találták ki és indították el a repozitóriumokat, hanem kutatók: Paul Ginsparg, Herbert van de Sompel és Stevan Harnad (Hernád István). De a rendszerek hamarosan a tudományos szakkönyvtárakban kötöttek ki: az arXiv például a Cornell Egyetemen. Bár az integrált könyvtári rendszerekben, könyvtári platformokban is megjelentek a digitális anyagok kezelésére szolgáló funkciók, a repozitóriumok funkcionalitását ezek jelenleg nem képesek maradéktalanul megvalósítani, nyitottságukkal nem tudnak versenyezni. Az egyik repozitóriumlajstromban, a ROAR-ban (Registry of Open Access Repositories) található repozitóriumok szoftverei között 2530 Dspace-t és 766 EPrintset találunk (ezek a leggyakrabban használt szoftverek Magyarországon), viszont az ExLibris szoftverei közül csak 9 DigiToolt és 2 Esplorót. A repozitóriumok különböznek a hagyományos könyvtári szolgáltatásoktól: más felhasználói kört szolgálnak, más gyarapítási stratégiák, más leírási szabványok alkalmazásával, valamint sokkal nagyobb nyitottsággal. (A technikai különbségek része a hagyományos könyvtáraknál alkalmazott Z39.50 helyett a gyorsabb OAI-PMH-protokoll, a könyvtári MARC-metaadat-séma helyett a könnyedebb Dublin Core alkalmazása. Mindezt a teljes szövegek digitális

elérhetősége teszi lehetővé. A mesterséges intelligencia [MI] alkalmazása talán csökkenti majd a tartalmat leíró metaadatok szerepét, de ma a repozitóriumi modellben szereplő egyszerűbb leírások előnye még érezhetőek.)

Szerep a Gold OA korában

Egyszerre siker és kudarc a kiadói (gold) nyílt hozzáférés elterjedése. A tudományos publikációk hozzáférhetősége jelentősen javult, a publikálás költségei viszont drasztikusan emelkedtek. A repozitóriumi (zöld) út használata nélkül az árakat nem sikerülhet kordában tartani (jelenleg a szerzői jogok visszatartása – *right retention* – tűnik megoldásnak). Ám a szerzőnél maradó jogok semmit sem érnek, ha nincsenek repozitóriumok, ahova a cikkeket elhelyezhetik. (A másik megoldás a gyémánt open access – a független, közlési díjat nem alkalmazó folyóiratok használatának elterjesztése. Az óriás kiadvállalatok ellenében azonban mindkét megoldás párhuzamos alkalmazására, minden lehetőség megragadására szükség lesz.)

A REAL szerepe egyfelől az Akadémiához kötődő publikációk, dokumentumok egy helyre gyűjtése, láthatóságának növelése, MI-vel történő feltárása lehet. Másfelől a változó igények kiszolgálása: a távoli olvasás támogatása, a sérülékeny dokumentumok digitális kópiáinak használata. A REAL – mint a többi repozitórium – látványraktár, ahol nagy tömegű dokumentum a közönség számára is láthatóan (bár nem látványosan) tárolható. Nem dekoratív, nemigen orientálja az olvasót. Szükség lesz tematikus honlapokra, időszakos webes kiállításokra – de az ezekben szereplő dokumentumok a hosszú távú megőrzést biztosító repozitóriumból kell, hogy jöjjenek.

Repozitórium mint digitális információvagyon

A REAL tartalma – pedig a REAL működtetése, a dokumentumok digitalizálása és kezelése is pénzbe kerül – nincs benne az Akadémiai Könyvtár leltárában (más könyvtáraknál is ugyanez a helyzet). Ez a digitális információvagyon megőrzendő – a megőrzés informatikai, jogi, szervezési kérdéseit ki kell dolgozni. A digitális információk lehetnek sérülékenyebbek a papírnál – sok CD és hajlékonylemez, amely a könyvtárba került, mára már nem olvasható. A REAL-ban található anyagok hosszú távú megőrzéséről gondoskodni kell. A repozitórium a könyvtár alapszolgáltatása, folyamatosan fenntartandó, a szükséges szoftverváltoztatás esetén pedig mind az információk, mind a repozitóriumi tartalmakra mutató külső hivatkozásaik megőrzését biztosítani kell.

Repozitórium mint szövegkorpusz

A digitális szövegfeldolgozás, a digitális bölcsészet eszközparkja kibővíti a szövegek kezelésének, feltárásának lehetőségeit. A REAL-ban található magyar nyelvű, modern szövegek nagy nyelvi modellekkel való feldolgozása az Akadémia támogatásával

(az MTA Tudomány a Magyar Nyelvért Nemzeti Program „A magyar nyelv digitális támogatása a magyar tudományosság szolgálatában” alprogramja keretében) elkezdődött. A REAL-ban található tudományos szövegek a nyelvmodellfejlesztés tanító anyagai lehetnek, a REAL jövőbeli olvasói pedig remélhetőleg MI-asszisztensek támogatását is élvezhetik majd.

A REAL szövegeinek nyelvtchnológiai feldolgozása azonban nem akadálymentes. Fel kellett ismernünk, hogy a repozitóriumban tárolt szövegek gépi elemzéséhez jelentős előfeldolgozásra van szükség. Még a digitálisan született anyagok is az emberi olvasónak készültek, folyamatos, tiszta szöveget kinyerni belőlük nem könnyű. Míg az OCR (Optical Character Recognition) az egyszerű, teljes szövegben való keresést teszi lehetővé, az MI-vel való feldolgozáshoz gyakran magasabb minőség szükséges.

Mivel a magyar tudományos kibocsátás egészének nemzetközi folyóiratokban vagy más publikációs platformokon való elhelyezése nem lehetséges, sőt állítjuk, hogy nem is mindig elégséges a hazai felhasználás minden formájának támogatása szempontjából, a hazai szöveges archívumokra – így a REAL-ra – továbbra is nagy szükség lesz.

Irodalomjegyzék

- „Akadémikusok. A Magyar Tudományos Akadémia tagjai”. MTAK (letöltés 2026. márc. 14.). <https://akademikus.mtak.hu>.
- Holl András (2022). „Repozitóriumok – különleges terület a könyvtárak világában”. *Tudományos és Műszaki Tájékoztatás* 69/7, 358–365. <https://doi.org/10.3311/tmt.13180>.
- Holl András et al. (2019). „Folyóirat- és könyvdigitalizálás az MTA Könyvtárában. Lezárult az első szakasz”. *Könyvtári Figyelő* 65/3, 375–382. <http://real.mtak.hu/103280>.