

Repozitóriumok - publikációk és a többi...

Networkshop, 2013, Sopron

Holl András

MTA KIK, MTA CsFK CsI

A repozitóriumok létrehozása után a legnagyobb feladat a tartalommal való feltöltésük. Előadásunkban megvizsgáljuk, mit lehet ennek érdekében tenni, mit tettünk/teszünk a REAL-nál (különös tekintettel az MTA OA határozatára)?

A repozitóriumokban elosztott tartalom keveset ér megfelelő globális vagy nemzeti keresési lehetőségek, aggregáció nélkül. Megvizsgáljuk az MTMT, mint nemzeti tematikus keresőfelület lehetséges szerepét.

A tudományos adatok elérhetővé tétele és hosszú távú megőrzése a publikációkénál sokszorta nagyobb kihívást jelent. Felvázoljuk, milyen megoldások léteznek a világban, és megvizsgáljuk, mi lehet a repozitóriumok szerepe?

A repozitóriumokra gyakorta jellemző a kihasználatlanság. Bizonyos tudományterületeknél megszokott a repozitóriumi elhelyezés - mint például a preprintek elhelyezése a fizikai tudományok területén az arXiv-ban. Általánosságban viszont az önkéntes elhelyezési arány 20% körüli (Gargouri et al., 2012). Az elhelyezési arányokat jelentősen növeli, ha kötelezzik a kutatókat a repozitóriumi elhelyezésre (az ilyen szabályozásra az angol „mandate” kifejezést használja a szakirodalom).

Az egyik legnagyobb tudományfinanszírozó szervezet, a National Institutes of Health az Egyesült Államokban 2005-ben repozitóriumi elhelyezésre kérte a támogatást elnyerő kutatókat, melynek eredményeként a támogatott cikkek 19%-a került a PubMed Central-ba (megjelölték a kívánt repozitóriumot). 2008-ban a kérést követelménnyé erősítették, és az elhelyezési arány hamarosan 49%-ra emelkedett (Poynder, 2012). 2012-re az elhelyezési arány 75%-ot ért el, és a NIH bejelentette, hogy 2013-tól szankcionálni fogja az előírást nem teljesítő kutatókat - a támogatás hátralévő részét visszatárhathatja (Matthews, 2013). Kíváncsian várjuk a szigorítás eredményét!

Az Open Access publikálási kötelezettségek láthatóan fontosak, valamint fontos az ellenőrizhetőség (a NIH OA előírása esetében a kijelölt repozitórium biztosítja ezt), és fontos a visszacsatolás (a kutató legyen tisztában azzal, hogy az Open Access-t számára elrendelő szervezet figyeli a rendelkezés teljesítését, és szankcionálja annak elmaradását). Mindezeket túl fontosnak tartjuk a támogatást is - a kutató ne egyedül kerüljön szembe a kiadókkal, ne maradjon magára az OA megvalósítása során.

A legerősebb motivációt az jelenti, ha a repozitóriumi elhelyezés előfeltétele valamiféle fokozat

megszerzésének. Az MTA doktora fokozat megszerzése során a folyamat része a dolgozat elhelyezése a REAL-d-ben - ezt nem is a fokozatra pályázó végzi, hanem a Doktori Tanács Titkársága.

A Magyar Tudományos Akadémia elnöke 27/2102-es határozatával 2013-tól elrendelte a Nyílt Hozzáférés biztosítását a tudományos publikációkhoz, minden MTA támogatással született műre (<http://real.mtak.hu/eprints/mandate.html>). A határozat bizonyos esetekben megengedi a zárt körű elérhetőséget - ebben az esetben a REAL-t, az MTA KIK repozitóriumát jelöli meg az elhelyezésre. Ezekről az esetektől eltekintve az Open Access mind a kiadóknál, mind a REAL-ban, illetve az MTA támogatással működő külső kutatócsoportoknál a befogadó egyetem repozitóriumában biztosítható, de akár nemzetközileg elfogadott tudományterületi repozitóriumokban is lehetséges. A határozat ellenőrzésének lehetőségét az MTMT-ben elhelyezendő, a szabadon elérhető tartalomra mutató link biztosítja.

Széleskörű támogatási rendszer kiépítésével kívánjuk elősegíteni a rendelkezés betartását. Az egyes kutatóközpontokban, intézetekben és kutatócsoportokban REAL-adminisztrátorokat jelöltek meg, akiknek az MTA KIK oktatást tartott. Azt javasoltuk, hogy a REAL adminisztrátorok lehetőleg könyvtárosok, és egyúttal MTMT adminisztrátorok is legyenek. Feladatuk lehet a repozitóriumi elhelyezések elvégzése a kutatók helyett, vagy segítségnyújtás a kutatóknak abban, hogy maguk töltsék fel a cikkeiket.

Előadókörúton több kutatóközpontba/intézetbe is eljutottunk az OA határozattal összefüggő kérdések kifejtése végett, és a REAL weblapján, valamint e-mailben is szolgáltatunk információkat a felhasználóknak és adminisztrátoroknak. Két elektronikus levelezési címet is létesítettünk segítségnyújtás céljából: a real_help a repozitóriumunkba való feltöltésekkel, az openaccess a szerzői jogokkal, általános, a Nyílt Hozzáféréssel kapcsolatos kérdésekkel kapcsolatban áll a kutatók rendelkezésére.

A kezdeti időszakban a linkek elhelyezését átvállaljuk a REAL felhasználóitól, mindaddig, amíg az MTMT-oldalról kezdeményezhető SWORD feltöltés lehetősége meg nem valósul.

Az OAI-PMH protokollt az aggregálást szem előtt tartva hozták létre - a Z39.50 hibáiból okulva. Számos kereső, aggregátor szolgáltatás jött létre - de kevés lett sikeres közülük. A Michigan Egyetemen fejlesztett, jelenleg az Online Computer Library Center-ben működő OAIster (<https://www.oclc.org/oaister.en.html>) nem a tudományos kutatásra szabott eszköz. A kutató többet remélhet - és kaphat - a Bielefeld Academic Search Engine-től (<http://www.base-search.net/>). A legismertebb a Google szolgáltatása, a Scholar. Hazai aggregátor is létezik: az NDA@SZTAKI (<http://nda.sztaki.hu/kereso/>), különböző hazai, OAI-PMH-képes szolgáltatásokat fog össze, köztük egyes tudományos tartalmúakat is - mint például a REAL.

Az aggregátorok fontos erénye a szüretelt repozitóriumok nagy száma, ám mindehhez szelektivitásra, a keresés megfelelő szűkíthetőségére is szükség van. Véleményünk szerint az MTMT töltheti be a hazai tudományos aggregátor és kereső szerepét. Ehhez szükség van a tartalmi, tárgyszavas osztályozás bevezetésére, a keresési lehetőségek bővítésére és finomítására (szerepelnie kell a választható feltételek között a teljes szövegre, ill. szabadon elérhető szövegre mutató link meglétének). A tudományos tartalmak elhelyezése az egyetemek, kutatóintézetek és közgyűjtemények repozitóriumában történhet, és az MTMT portál-szolgáltatása biztosíthatja ezekhez a tartalmakhoz a kaput.

Az EU és az USA a publikációkhoz való Nyílt Hozzáféréseken túl immár az adatok nyílt hozzáféréseinek és hosszú távú megőrzésének elérését tűzte ki célként¹. Az adatok elhelyezése sokkal komplexebb feladatot jelent, mint a publikációké - elegendő az esetleges óriási mennyiséget, vagy azt a tényt megemlíteni, hogy speciális szoftverekre lehet szükség a megjelenítéshez, értelmezéshez. Mindamellett csekély mennyiségű, egyszerű adathalmazok kezelésére is szükség van - itt a klasszikus repozitóriumoknak is lehet szerepe.

A „Big Science” adatai speciális tárolást igényelnek, és a rendelkezésre álló erőforrások ezek kialakítását lehetővé is teszik. Az is jellemző, hogy a kutatási projektek tervezése során gondolnak az adatok közzétételére is, sőt, ez a követelmények közé tartozik. Példa a CERN / LHC esetében a DASPOS².

Kisebbségi projektek adatainak tárolására az intézményi adatrepozitóriumok lehetnek alkalmasak. Példaként említhető a Harvard Egyetemen fejlesztett Dataverse³. A Dataverse az adatok hivatkozhatóságát a szokásos DOI-n kívül az Universal Numerical Fingerprint segítségével biztosítja, megteremtve a verziókövetés lehetőségét (Altman és King, 2007). Csillagászati adatok tárolására fejlesztették, a Virtuális Obszervatórium szabványai szerint a Saada-t⁴ Strasbourgan.

Egyedi - egy-egy publikációhoz tartozó, vagy éppen negatív eredményt adó, esetleg be nem fejezett kutatáshoz tartozó, ezért soha le nem közölt - adatállományok tárolására is születtek kezdeményezések. Megemlíthető a figshare⁵, ahol 1GB-nyi tárhely áll rendelkezésre ingyenesen minden regisztrált kutatónak. Az elhelyezett állományokat tallózva prezentációkat, táblázatokat és - nem meglepően - ábrákat találunk. Egy másik adatrepozitórium a Dryad⁶. Itt hangsúlyozottan a referált cikkekhez kapcsolódó adatállományok elhelyezése a cél.

Vajon a publikációs repozitóriumok alkalmasak adatok elhelyezésére? Nyilvánvalóan igen, az EPrints szoftvert alkalmazó repozitóriumok között található például a southamptoni eCrystals⁷. Többféle korlátja lehet azonban az ilyen irányú felhasználásnak. A rendelkezésre álló szabad repozitóriumi szoftverek nem petabájtos adatmennyiségek kezelésére készültek. A publikációs repozitóriumok többnyire könyvtárakban működnek - probléma lehet az, hogy nincs meg a megfelelő szakértelem a tudományos adatok kezeléséhez. A speciális tudományos adatok megfelelő leírását - metaadatokkal való ellátását - csak maguk a kutatók végezhetik, egy könyvtáros ezt még ellenőrizni sem tudja (a megoldást minden bizonnyal többszintű metaadat-sémák jelenthetik, ahol egyes szintek már ellenőrizhetőek a szakterületen kívülállóknak is; Zuiderwijk et al, 2012). Probléma az adatok megjelenítése is, valamint az a tény, hogy a speciális adatformátumok előbb-utóbb szükségessé váló migrálása sem könnyen elvégezhető egy könyvtárban.

1 EU: Towards better access to scientific information, http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf

USA: Holdren Memorandum,

http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

2 DASPOS: <https://daspos.crc.nd.edu/>

3 Dataverse: <http://thedata.org/>

4 Saada: <http://amwdb.u-strasbg.fr/saada/spip.php?article32>

5 figshare: <http://figshare.com/>

6 Dryad Digital Repository: <http://datadryad.org>

7 eCrystals: <http://ecrystals.chem.soton.ac.uk/>

Milyen feltételekkel lehetne adatokat elhelyezni a hagyományos repozitóriumokban? Már említettük a viszonylag kis mennyiséget. További szempont lehet a szabványos, szélesen elterjedt formátum - ahol nincs szükség speciális szoftverre a megjelenítéshez, nem jelent aránytalanul nagy problémát az esetleg szükségessé váló migráció. Fontos az értelmezhetőség: olyan adatállományok elhelyezése oldható meg könnyen, ahol a megjelenített adatok a felhasználó számára értelmezhetőek is, ehhez sincs programra szükség. Az adatok minőség-ellenőrzését sem végezhetik el a repozitórium munkatársai - olyan megoldást lehet elképzelni, hogy a cikkekhez kapcsolódó adatállományok a cikkel együtt kerülnek ellenőrzésre a *peer review* során.

A REAL esetében foglalkozunk azzal a kérdéssel, hogy folyóiratok kiadóival és szerkesztőségeivel vegyük fel a kapcsolatot. A repozitóriumban elhelyezhetőek lennének a folyóiratban terjedelmi okokból le nem közölhető ábrák és táblázatok. Elképzeléseink szerint a szakmai bírálókat kiterjedne ezekre az anyagokra is, és elérhetőségük fel lenne tüntetve a cikkben is. A három oldalú (szerző-kiadó-repozitórium) együttműködés mindegyik fél, és a tudományos közönség számára is hasznos lehet.

Irodalom

Altman, M., King, G. (2007) A Proposed Standard for the Scholarly Citation of Quantitative Data. *D-Lib* 13(3/4). <http://dlib.org/dlib/march07/altman/03altman.html>

Gargouri, Y., Lariviere, V., Gingras, Y., Brody, T., Carr, L. and Harnad, S. (2012) Testing the Finch Hypothesis on Green OA Mandate Ineffectiveness. In, Open Access Week 2012 <http://eprints.soton.ac.uk/344687/>

Matthews, S. (2013) NIH will withhold grant money to enforce public-access policy, *Nature Medicine*, 19, 3, doi:10.1038/nm0113-3

Poynder, R. (2012) Open Access Mandates: Ensuring Compliance. in: Open and Shut? (blog), <http://poynder.blogspot.hu/2012/05/open-access-mandates-ensuring.html>

Zuiderwijk, A., Jeffery, K., Janssen, M. (2012) *The Potential of Metadata for Linked Open Data and its Value for Users and Publishers*, *JeDEM*, 4, 2, <http://www.jedem.org/article/view/138>

Repositories - publications and more...

Filling the repositories with publications is a challenge considerably harder than creating these databases. We will review the case of REAL, the Repository of the Academy's Library, with particular emphasis on the Academy's OA mandate.

To expose and make searchable the contents of repositories global and nation-wide aggregating services are needed. We will look into the possibility of using the MTMT, the Hungarian National Scientific Bibliography database as a thematic portal and search engine for the contents of repositories at the MTA and Universities in Hungary.

Providing permanent access to scientific data is a challenge. We review the existing solutions, and discuss the possible role of repositories.