

## **KÉZIRAT – Beküldve (AgriaMedia 2017 konferenciakiadvány)**

**Holl András**

MTA Könyvtár és Információs Központ

*holl.andras@konyvtar.mta.hu*

### **ÚJRAGONDOLT REPOZITÓRIUMOK**

#### **Bevezetés**

A repozitóriumok kettős feladata a dokumentumok megőrzése és közreadása. Az elmúlt két évtizedben számos sikert értek el mindkét téren. Mégis feltehetjük a kérdést: lehetne jobban csinálni? Hogyan tudnánk az elhelyezett dokumentumok láthatóságát növelni? Mi módon növelhetnénk a szerzők feltöltési aktivitását? Hogyan illeszkedhetnének ezek a szolgáltatások jobban a kutatók munkafolyamataiba? Milyen új technológiák segíthetnek mindezekben? Ezeket a kérdéseket sok helyen felteszik világszerte. A Confederation of Open Access Repositories (COAR) munkacsoportot állított fel az új generációs repozitóriumok megtervezésére. Amellett, hogy kíváncsian várjuk a repozitóriumi közösség megoldásait, mindannyian, akik repozitóriumot üzemeltetünk, el kell gondolkozzunk a saját válaszainkon.

#### **Repozitóriumok**

A repozitóriumok teljes szövegű tartalmakat megőrző és szabad olvashatósággal bemutató digitális könyvtári gyűjtemények, melyek az OAI-PMH protokollnak megfelelően arathatóak, aggregálhatóak. Általában lehetővé teszik a dokumentumok szerzőinek publikációik feltöltését – ugyanakkor a feltöltött dokumentumok könyvtárosi ellenőrzésen mennek keresztül. Az egyes tételek – bár az Open Access hozzáférés elvének széles körű alkalmazása követelmény, egyes dokumentumok esetében a hozzáférés korlátozása, hosszabb-rövidebb hozzáférési embargók alkalmazása gyakran előfordul.

Az első repozitórium – az arXiv – tematikus repozitórium, ami a fizikai és rokon tudományok területén ad közre cikkeket megjelenés előtti állapotban. Szintén ismert és elismert tematikus repozitórium a PubMed Central az orvosi és egyéb élettudományok területén. E két archívum együttesen mára közel 6 millió cikket tartalmaz. Az adott tudományterületen használt bibliográfiai információs rendszerekkel együtt – ilyen a nagyenergiájú fizika területén az INSPIRE, a csillagászatban a NASA/SAO ADS és az orvostudományban PubMed – mára a kutatási infrastruktúra nélkülözhetetlen, alapvető részévé váltak.

Az intézményi repozitóriumok – a nagy tematikus repozitóriumoktól eltérően – többnyire szabad szoftvereket használnak, és az adott intézmény publikációs, oktatási, dokumentációs produktumát tárolják és adják közre. A közreadás, nyilvános elérhetőség lehet az adott intézmény területére, kutatóira, hallgatóira korlátozott, és lehet valódi Open Access, minden potenciális olvasó számára hozzáférhető. A belső nyilvánosság megteremtése is előnyös, ugyanakkor a teljes nyilvánosság nyilvánvalóan még előnyösebb, noha esetenként csak nehézségek árán valósítható meg a kiadói jogok miatt.

A ROAR és az OpenDOAR jegyzékek több ezer intézményi repozitóriumot listáznak, a legnagyobbak milliós tételszámmal büszkélkedhetnek. A repozitóriumok területi alapon is aggregálhatóak – a latin-amerikai repozitóriumokat tömörítő LaReferencia másfél millió dokumentumot tesz kereshetővé egy felületen. A Bielefeldi Egyetem közös keresője – a BASE – százmilliónyi dokumentumot aggregál ötezer forrásból. PhD dolgozatokat összesít a DART-Europe, pillanatnyilag tizenhat hazai egyetem dolgozatai érhetőek el ezen a szolgáltatáson keresztül. A hazai repozitóriumok közös szervezete, a HUNOR, 11 hazai intézményi repozitóriumot sorol fel a honlapján, az MTA KIK és az MTA SZTAKI közös repozitórium keresője pedig 15 hazai intézmény 26 repozitórium gyűjteményéből több, mint 210 ezer dokumentum adatai között kínál keresési lehetőséget. Az MTMT Repozitóriumminősítő Szakbizottsága minősítési rendszert működtet az intézményi repozitóriumok minőségbiztosításának támogatásául.

## **Imposztorok**

Nem minden elektronikus könyvtár repozitórium – még akkor sem, ha az alkalmazott szoftver elterjedt repozitórium szoftver. Gyakran előfordul, hogy az üzemeltetők a könnyebb utat választják, és repozitóriumukban túlsúlyba kerülnek a teljes szöveg nélküli, csak bibliográfiai adatokat tartalmazó, vagy pedig nem nyilvánosan olvasható tételek.

Vannak a dokumentumtárak között azonban magukat repozitóriumnak feltüntető imposztorok is. Ilyen az Academia.edu, a ResearchGate és a Mendeley. Nem feltétlenül repozitóriumként definiálják magukat, és valóban sok olyan tulajdonságuk van, amelyekkel a repozitóriumok nem rendelkeznek. Elsőnek a sokkal erősebb közösségi média jelleget kell említenünk. Mindazonáltal a felhasználókban gyakorta azt a benyomást keltik, mintha repozitóriumok lennének. Legtöbbet használt funkciójuk a tartalommegosztás. Ám nem biztosítanak valódi nyílt hozzáférést – regisztrációkötelesek. Mivel kockázati tőkére épülő üzleti vállalkozások – vagy mostanra már multinacionális cégek - állnak mögöttük, a felhasználók nem lehetnek biztosak abban, az általuk feltöltött tartalmak a jövőben milyen feltételekkel lesznek olvashatóak. Annál is bizonytalanabb ezen tartalmak helyzete, mert a szolgáltatások nem szűrik kiadói jogosultságok szempontjából a feltöltéseket.

Mindhárom említett szolgáltatásra jellemző az agresszív felhasználó-toborzás – ha a módszerek különböznek is. Kettő közülük: az Academia.edu és a ResearchGate

etikátlan, a felhasználót megtevesztő módszereket is alkalmaz. Az imposztorok konkurenciája jelentős fenyegetés az intézményi repozitóriumokra nézve.

## Helyzetértékelés

Bár az intézményi repozitóriumok nagy része néhány általánosan elterjedt szabad szoftverplatformot használ (mint a DSpace, EPrints, Digital Commons), a működési gyakorlat jelentősen különböző lehet a repozitóriumok között. Mint láttuk, a repozitóriumok az OAI-PMH protokoll alapján aggregálhatóak, de ennek a technológiának is látszanak a korlátai. Míg a nagy tematikus repozitóriumok jól integrálódtak a tudomány szövetébe, az intézményi repozitóriumokban található publikációk láthatóságát lehetne növelni. Ehhez mind a felhasznált technológia fejlesztése, mind az üzemeltetési gyakorlat egységesítése szükséges. Továbbá elengedhetetlen, hogy az üzemeltető intézmények integrálják repozitóriumukat szolgáltatásaik struktúrájába, és erősödjön a kapcsolat az épülő globális tudományos kommunikációs infrastruktúrákkal is.

Sok egyetem, kutatóintézet könyvtára működtet úgy repozitóriumot, hogy az nem illeszkedik intézményi Open Access stratégiába (mivel esetleg ilyen nem is létezik, vagy gyenge, csupán ajánlás szintű). Ahol a publikációk repozitóriumi elhelyezése az értékelés alapjául szolgál – mint a Liege-i Egyetem – ott a repozitóriumok feltöltöttsége is megfelelő. Még jobb a helyzet, ha a repozitóriumok integrálódnak a kutatást, oktatást segítő rendszerekbe – de ez inkább csak kivételesen teljesül jelenleg.

## Next Generation Repositories

A repozitóriumok nemzetközi szervezete, a Confederation of Open Access Repositories munkacsoportot állított fel a repozitóriumok jövőjének vizsgálatára. 2017 februárjában született meg az első nyilvános vitaanyag<sup>1</sup>.

Két szála fűzte fel tevékenységét a vitaanyagot létrehozó, húsz szakemberből álló munkacsoport. Egyfelől a repozitóriumok működését kívánták jobban szabványosítani: jelenleg csupán az OAI-PMH jelenti az általánosan elfogadott technológiai alapot. Másrészt erősíteni kívánták a kapcsolatot a tudományos kutatás alapvető infrastruktúráival, elsősorban az azonosítás területén (legyen szó akár a szervezetek, akár a pályázatok, a kutatók vagy publikációk egyértelmű azonosításáról). Fontos cél volt az értéknövelt szolgáltatások kialakítása – a repozitóriumoknak meg kell haladniuk a rektár szerepet. A tudományos kommunikáció terén való előrelépés általános feltételének tekintették a közösségi jellegű, elosztott irányítási struktúrák, nyílt rendszerek, egységes

---

<sup>1</sup> Next Generation Repositories. Vitaanyag. <https://www.coar-repositories.org/files/COAR-Next-Generation-Repositories-February-7-2017.pdf>

gyakorlatok kialakítását, és a folyamatok felgyorsítását – az rendszerek valós idejű működésének megvalósítását.

Tucatnyi példával mutatja be a vitaanyag az elvárásokat:

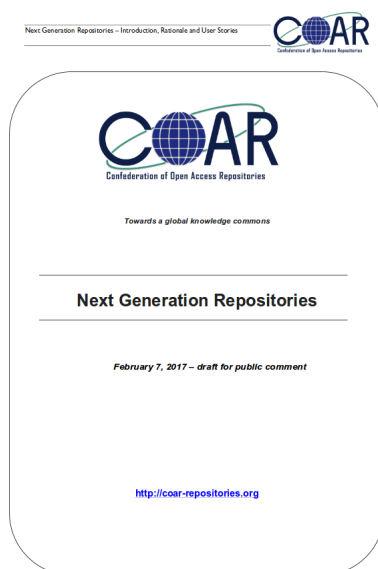
1. Tudományos információk metaadatainak felfedezhetősége
2. Tudományos információk azonosítójának felfedezhetősége
3. Felhasználhatósági jogok felfedezhetősége

A fenti követelmények leginkább a robotok számára fontosak.

4. A felhasználó felismerése
5. Annotálás, kommentálás és lektorálás lehetőségének megteremtése

Ez a két követelmény a közösségi média-szerű felhasználás erősítését célozza.

6. Automata ajánló rendszerek létrehozása
7. Információátadás közösségi rendszerek felé
8. Erőforrások szinkronizálása
9. Adatbányászat
10. Kutatói munkafolyamatok támogatása
11. Felhasználási statisztikák
12. Hosszú távú megőrzés



1. ábra A COAR munkaanyagának címlapja

Ugyancsak foglalkozott a repozitóriumok jövőjével, és ezen belül a munkacsoport törekvéseivel a szervezet éves gyűlése Velencében<sup>2</sup> az OAI10 konferencia Genfben, ez év júniusában<sup>3</sup>.

## **Technológiai fejlődés**

Herbert Van de Sompel, az OAI-PMH egyik fejlesztője új protokollokon, technológiákon dolgozik. Az említett genfi konferencián, illetve a COAR 2016-os bécsi gyűlésén beszélt a lehetséges új technológiákról. Véleménye szerint a metaadatokon alapuló aggregáció felett eljárt az idő (Van de Sompel és Nelson, 2015). Véleményünk szerint nem kell elsietni a metaadatok temetését, mindazonáltal érdemes az új technológiákra odafigyelni.

Figyelemre méltó a permanens azonosítók (DOI, ORCID) terén történő infrastrukturális fejlődés. A DataCite, a CrossRef, az ORCID közös projektjei új, közösségi alapú infrastruktúrát teremtenek a tudományos kommunikáció számára. Az egyik közös projekt – a THOR – nemrégiben tartott rendezvényt az MTA KIK-ben<sup>4</sup>.

Az utóbbi idők egyik legnagyobb hatású fejleménye az ImpactStory által létrehozott oaDOI adatbázis, és a ráépülő Unpaywall. A repozitóriumok láthatóságát átütően javíthatja amennyiben a tárolt tartalmak az eredeti DOI-juk alapján az oaDOI adatbázisban láthatóvá válnak – ebben szerepet játszhat a BASE. Az MTA KIK EISZ irodája által fejlesztett COMPASS (Dér, 2017) is szerepet kaphat repozitóriumi tartalmak elérésében az oaDOI adatbázis felhasználásával.

## **Hazai lehetőségek, példák, REAL**

Mint korábban bemutatuk, Magyarországon is kiépült egy széleskörű repozitóriumi rendszer. (Bár vannak még intézmények, akik nem üzemeltetnek repozitóriumot, vannak olyan egyetemek, amelyek saját repozitórium híján sem használják PhD dolgozataik nyilvánossá tételére az MTA KIK által nyújtott központi repozitóriumi szolgáltatást, és a repozitóriummal rendelkezőknek sem mindegyike kapcsolódott be a hazai hálózatba.) Mindazonáltal itthon is megérett az idő a továbblépésre. Magasabb szintre kell emelnünk a repozitóriumi rendszer használatát, nem szabad megelégedni azzal, hogy létezik a szolgáltatás és vannak benne dokumentumok. Ezen az úton elindult már néhány intézmény. A Debreceni Egyetemen az iDEa Tudóstér integrálja a repozitóriumot az intézményi információs rendszerbe (Pergéné Szabó, 2016). Szövegbányászattal kísérleteznek a Szegedi Egyetemen (Nagy, 2016).

---

<sup>2</sup> COAR 2017-es közgyűlés: <https://www.coar-repositories.org/community/coar-annual-meeting-2017/>

<sup>3</sup> Lásd az OAI10 programot: <https://indico.cern.ch/event/405949/timetable/#all.detailed>

<sup>4</sup> THOR Bootcamp Budapest: <http://openaccess.mtak.hu/index.php/esemenyek/6-thor-bootcamp>

Az MTA KIK REAL repozitóriumának régóta törekszünk láthatóbbá tenni, minél jobban bekapcsolni az MTA és az ország tudományos életébe. Új fejleményként jelennek meg a repozitóriumban tárolt anyagokra támaszkodó tematikus webes gyűjtemények (Sallai-Tóth, 2017).

A REAL láthatóbbá tételére irányuló erőfeszítések egyike a Wikipédia kapcsolódások elkészítése. Elsősorban a REAL-EOD gyűjteményben található könyvek, dokumentumok kerültek bekötésre a magyar Wikipédiába, minthogy ezeknél a tartalmaknál nagyobb a valószínűsége, hogy a szócikkekben megjelennek.

A koncepciónk mindig az volt, hogy a repozitórium háttértároló, az olvasónak nem a repozitórium szoftver által biztosított felületen kell keresnie. Esetünkben ezt a koncepciót indokolta az általunk használt szoftver – az EPrints – azon tulajdonsága is, hogy az egyes gyűjtemények önálló URL-en, látszólag önálló repozitóriumként jelennek meg. A felhasználó véleményünk szerint az általánosan használt keresőn, esetenként a Wikipédián, a tematikus gyűjteményeken, vagy a könyvtári egyablakos keresőn (discovery service-en): esetünkben a Primo-n keresztül találja majd meg a tartalmakat. Vagy éppenséggel valamelyik repozitórium aggregátoron keresztül. A REAL dokumentumai – sok más hazai repozitóriumhoz hasonlóan – megjelennek a BASE-ben, a PhD dolgozatok a DART-Europe-ban. Az MTA SZTAKI-val együttműködésben üzemeltetünk egy hazai aggregátort is: az OAIkereso-t.

A távlati cél a REAL – esetleg általánosságban a hazai repozitóriumok – tartalmainak fejlett szövegbányászati eszközökkel való feltárása. A feladat szoftverfejlesztéssel való megoldása csak megfelelő pályázati támogatás elnyerésével lenne lehetséges. A másik lehetőség az, hogy megvárjuk, amíg az új generációs repozitórium technológiák megjelennek az EPrints-ben (ahogyan erre, vagy a DSpace hasonló fejlesztésére hagyatkozhatnak más hazai intézmények is).

## **A repozitóriumok szerepe a felsőoktatásban**

Miért fontosak a repozitóriumok a felsőoktatásban? Egyfelől lehetőséget adnak arra, hogy a hallgatók információkhoz jussanak. A tudományos szakcikk 15-20%-a repozitóriumokban található (Björk, 2017), ezeket a forrásokat vétek lenne mellőzni az oktatásban. Másfelől az intézményi repozitóriumok lehetnek azok a helyek, ahol a hallgatók első munkái nyilvánosságra kerülnek, PhD dolgozataik hozzáférhetővé válnak. Szerepet játszhatnak az intézményi kiadványok archiválásában és a szürke irodalom feltárásában. Nem utolsó sorban a repozitóriumok – és rajtuk keresztül a tudományos kommunikáció kutatóközösségi, non-profit infrastruktúrájának – ismeretét és használatát a jövő kutatóknak már az egyetemen el kell sajátítaniuk. Az infrastruktúra kereskedelmi alapon működő részeinek megismertetése mellett – ehhez jelentős támogatást adhatnak a multinacionális szolgáltatók – hangsúlyt kell helyezni a nyílt rendszerek megismertetésére, olyan kutatók, szakemberek képzésére, akik karrierjük során használni tudják majd a nyílt hozzáférés nyújtotta lehetőségeket. Ha ezt elhanyagoljuk, még jobban ki leszünk szolgáltatva a kereskedelmi szolgáltatásoknak.

## Irodalomjegyzék

- Björk, B-C. 2017. *Gold, green and black open access*. Learned Publishong, 30.2. DOI: [10.1002/leap.1096](https://doi.org/10.1002/leap.1096)
- Dér, Á. 2017. *A COMPASS adatbázis új verziójának fejlesztése*. Networkshop, Szeged, 2017.
- Nagy, Gy. 2016. *Tudományometriai és tartalmi elemzések szövegbányászati módszerekkel*. Networkshop, Debrecen, 2016.
- Pergéné Szabó E. 2016. *iDEA Tudóstér - Látható Tudomány a Debreceni Egyetemen*. Networkshop, Debrecen, 2016
- Sallai-Tóth L. 2017. *Irodalmi adatbázisok a Magyar Tudományos Akadémia Könyvtár és Információs Központban: repozitóriumi archiválás és open access publikálás a nyelv- és irodalomtudomány területén*. (előkészületben)
- Van de Sompel, H., Nelson, M.L. 2015. *Reminiscing About 15 Years of Interoperability Efforts*. D-Lib Magazine, 21, 11/12. DOI: [10.1045/november2015-vandesompel](https://doi.org/10.1045/november2015-vandesompel)