

Kómár Éva, Bánki Zsolt et al.
Fehér Könyv. Módszertani útmutató a közgyűjteményi kulturális örökség digitalizálásához és közzétételéhez , 258 p. (2019)
kiadta az Emberi Erőforrások Minisztériuma
[könyvfejezet]

Repozitóriumok – a digitalizált közgyűjteményi dokumentumok archiválására és szolgáltatására alkalmas rendszerek

Holl András
MTA KIK

A repozitóriumok az Open Archives Initiative¹ nyomán létrejött digitális gyűjtemények (és az ezeket gondozó szervezeti egységek, valamint az általuk használt szoftverek). A tudomány és a felsőoktatás területén elterjedtek, de megtalálhatóak nemzeti könyvtárakban² és múzeumokban³ is.

A repozitóriumok alapvető ismérvei: i.) az OAI-PMH protokoll alapján való aggregálhatóság; ii.) teljes dokumentumok tárolása; iii.) a felhasználói feltöltés lehetősége; iv.) a kurátori (könyvtárosi ellenőrzés; v.) a szabad hozzáférés biztosítása a dokumentumokhoz; vi.) a hosszú távú fenntartásra, archiválásra való törekvés; vii.) a SWORD protokoll támogatása. Előfordul, hogy az ismérvek egyike vagy másika nem teljesül – az OAI-PMH protokoll támogatása nélkül viszont egy digitális könyvtár nem tekinthető repozitóriumnak. Egyes repozitórium gyűjteményeknél a felhasználói feltöltés nem megengedett, és a repozitóriumokban tárolt egyes tételek lehetnek hiányosak (csak metaadatok, digitális objektum nélkül), vagy korlátozottan elérhetőek.

Az első repozitóriumok többnyire szöveges dokumentumokat tartalmaztak, de bármi egyéb média (digitális kép, hanganyag, videó) elhelyezésével is találkozhatunk. Újabban szerepet kaptak a digitális oktatási anyagok és kutatási adatok archiválásában és hozzáférhetővé tételében is. Eleinte a repozitóriumok az egyes intézmények vagy tudományterületek a hagyományos könyv- és folyóiratkiadásban megjelent közleményeinek másodlagos lerakatai voltak (illetve a tudományterületi preprint-repozitóriumok még a meg nem jelent kéziratokat tárolták és szolgáltatták – itt a másodlagos elhelyezés időben megelőzte az elsődlegest). A felsőoktatásban általánosan alkalmazzák ezeket az adattárakat a hallgatók diákköri és szakdolgozatainak, illetve a doktori iskolákban megvédett PhD dolgozatok archiválására és hozzáférhetővé tételére is – a diákköri és szakdolgozatok esetében a dokumentumok többnyire csak az intézményen belül érhetőek el. A digitalizálási projektek esetében is gyakorta ezt a formát választották a dokumentumok közzétételére. Többnyire a szolgáltatási minőségű kópiákat helyezik el repozitóriumokban, a nagyobb felbontású, állományvédelmi kópiák (amennyiben készülnek ilyenek) többnyire nem nyilvánosak, és terjedelmük miatt sem szokták a repozitóriumokban elhelyezni ezeket. De semmilyen elvi akadálya nincs a nagy felbontású kópiák repozitóriumban való elhelyezésének sem – ugyanannak a dokumentumnak több különböző felbontású digitális változata is elhelyezhető egy tételben, eltérő elérhetőségi beállítások mellett.

Az OAI nyomán számos nyílt forráskódú, szabadon hozzáférhető repozitórium szoftvert fejlesztettek⁴ – ilyenek például a Magyarországon is elterjedt DSpace és EPrints, ám vannak egyedi szoftvert használó repozitóriumok is, és előfordul, hogy eredetileg nem repozitóriumnak készült

1 <https://www.openarchives.org/>

2 Ausztrália <https://www.nla.gov.au/content/national-library-of-australia-digital-object-repository>

3 Arizona Sonora Desert Museum <http://www.desertmuseumdigitallibrary.org/public/index.php>

szoftverben biztosítják az OAI-PMH-nak való megfelelést. Több hazai példa van a szabad forráskódú alapra épülő, de hazai informatikai cég által módosított szoftverek alkalmazására is. A technikai hátterek palettáján további – talán a hazai körülmények között kevésbé reális – lehetőség a külső szervezet által üzemeltetett repozitóriumi szolgáltatás igénybevétele.

A repozitóriumokat üzemeltető közösség nemzetközi szervezete a COAR⁵. Kulcsszerepet töltenek be az aggregátorok: ilyen a Bielefeldi Egyetem (Németország) által üzemeltetett BASE⁶, vagy az Egyesült Királyságban a JISC által létrehozott Core⁷. Tematikus európai aggregátor a DART-Europe, ami PhD dolgozatokat tesz kereshetővé. Fontos szerepet töltenek be a láthatóság növelésében a repozitóriumokat tartalmazó jegyzékek, mint az OpenDOAR⁸.

A repozitóriumok megbízható dokumentumszolgáltatók. A megbízhatóság magában foglalja a tárolt (és közreadott) dokumentumok valamiféle azonosságát a folyóiratokban vagy könyvként megjelent publikációkkal vagy eredeti műtárgyakkal, a disszertációk azonosságát a fokozatszerzés alapját képezővel, a metaadatok megfelelőségét, a felhasználásra vonatkozó jogosultságok kezelését. Nem elhanyagolható szempont a hosszú távú megőrzés és szolgáltatás biztosítása: a repozitóriumok feladata mind az elérhetőség, mind az olvashatóság biztosítása. Ezek érdekében meg kell oldaniuk az URL-ek fenntartását, vagy URN-ek alkalmazását, és a digitális dokumentumok formátumának szükség szerinti konvertálását is, amikor az eredeti formátum elavul. A megbízhatóság biztosítására széles körben elfogadott minőségbiztosítási eljárások szolgálnak⁹.

A minőségbiztosítási eljárások nem csupán az alkalmazott technikát és az üzemeltetési eljárásokat vizsgálják, hanem az intézményi hátteret is. A repozitóriumok nem egyszerű weboldalak, amelyeket az esetleges pályázati fenntartási periódus leteltével sorsukra lehet hagyni, hanem a digitális vagyont kezelő intézmények – közgyűjtemények – szerves részei, a bennük tárolt dokumentumok az intézmény által kezelt, nyilvántartott állomány darabjai kell legyenek, még ha „csak” az állományba tartozó fizikai objektum digitális másolatai is, még ha több helyen, több egyenértékű másolat létezik is.

Magyarországon a repozitóriumok üzemeltetői a HUNOR¹⁰ szervezet keretében segítik egymás munkáját. A hazai repozitóriumok minőségbiztosítását az MTMT keretein belül működő Repozitóriumminősítő Bizottság¹¹ támogatja. A bizottság nyilvános önértékelési kérdőíve és minősítési protokollja azoknak a repozitóriumoknak is útmutatást adhat, amelyek (még) nem akarják minősíteni magukat. Hazai repozitóriumi aggregátor is működik: az MTA KIK és az MTA SZTAKI közös vállalkozásában üzemeltetett OAkereső¹². Repozitóriumnak tekinthetjük az OSZK és a MEK Egyesület által működtetett EPA-t és MEK-et is.

A repozitóriumok használata elterjedt, mégis sok kritika érte őket az utóbbi években a tárolt dokumentumok nem megfelelő láthatósága, a szegényes felhasználói szolgáltatások tekintetében. A repozitóriumi közösség felvette a kesztyűt, és foglalkozik a fejlesztéssel. A láthatóság növelésének eszközei között található a DOI azonosítók alkalmazását. Nem feltétlenül arról van szó, hogy

4 Nyílt forráskódú repozitórium szoftverek: http://oad.simmons.edu/oadwiki/Free_and_open-source_repository_software

5 Confederation of Open Access Repositories, <https://www.coar-repositories.org/>

6 Bielefeld Academic Search Engine, <https://api.base-search.net/>

7 <https://www.jisc.ac.uk/core>

8 <http://v2.sherpa.ac.uk/opensoar/>

9 <http://www.trusteddigitalrepository.eu/Trusted%20Digital%20Repository.html>

10 HUNGarian Open Repositories <https://openscience.hu/hu/hunor>

11 <https://www.mtmt.hu/repozitoriumminosito-szakbizottsag>

12 <http://oaikereso.sztaki.hu/kereso/index.php>

minden repozitóriumi tételt DOI azonosítóval lássanak el – ez csak az elsődlegesen repozitóriumokban elhelyezett digitális objektumok esetében fontos. A másodlagos elhelyezéseknél az eredeti megjelenéshez kapcsolódó azonosítót szükséges a repozitóriumban rögzíteni. A DOI-val ellátott (és csak előfizetéssel elérhető) tudományos cikkek szabadon hozzáférhető repozitóriumi változatai (a közlésre elfogadott kéziratok) az oaDOI/Unpaywall¹³ adatbázison segítségével válhatnak láthatóvá. Egyre nagyobb szerepet töltenek be a repozitóriumi tartalmak láthatóságának növelésében a könyvtárak olvasóik számára biztosított egyablakos keresői (discovery services). Ezek többnyire keresnek a könyvtári katalógus, és a könyvtár által előfizetett digitális tartalmak mellett a helyi repozitóriumban is. Ezen túl a nagyobb keresőrendszer-szolgáltatók igyekeznek jelentősebb repozitóriumokat is beemelni a minden könyvtár számára rendelkezésre álló adatforrások közé. Sok szempontból jobb megoldás, ha a repozitóriumi tételek magában a közgyűjtemény katalógusában is megjelennek. Amennyiben repozitóriumi feltöltést a SWORD protokoll segítségével a katalógus felől is lehet kezdeményezni, a katalógusban szereplő metaadatok átkerülhetnek a repozitóriumba, a katalógusba pedig hivatkozás kerül a digitális objektumra.

Egyre inkább igaz, hogy a repozitóriumok megbízható, hosszú távú háttértárak, és a bennük tárolt objektumok láthatóságának növeléséről más szolgáltatások gondoskodnak – például virtuális kiállítások¹⁴. A Magyar Tudományos Művek Tára és az iDEA Tudóstér¹⁵ további példák a repozitóriumokban tárolt dokumentumokra mutató linkeket a szolgáltatásba beemelő adatbázisokra.

Repozitóriumi szoftver funkcionálisok - repozitóriumi feladatok

A repozitóriumok alábbi feladatait általában a nyílt forráskódú szoftverek is támogatják:

- 1.) Önregisztráció (korlátozható), felhasználók, adminisztrátorok felvétele.
- 2.) Feladatkörök elválasztása: regisztrált, feltöltési lehetőséggel rendelkező felhasználó; elfogadási/visszautasítási jogkörrel rendelkező adminisztrátor, főadminisztrátor.
- 3.) Munkafolyamatok, adminisztrátori felelősségi körök támogatása.
- 4.) Tételek feldolgozási folyamatának támogatása: munkaterület, ellenőrzési terület, nyilvános terület.
- 5.) Láthatósági és hozzáférési szintek támogatása: csak a felhasználónak látható (munkaterületen lévő tételek); az adminisztrátorok számára látható (ellenőrzési terület), nyilvános (szabadon hozzáférhető, csak regisztrált felhasználóknak hozzáférhető vagy zárolt dokumentum - a metaadatok mindhárom esetben nyilvánosak); embargós (valamelyik korlátozott hozzáférési szint alkalmazása adott dátumig, ami után automatikusan szabadon hozzáférhető lesz a dokumentum).
- 6.) Felhasználási licencek megadásának lehetősége, feltöltői szerzői jogi nyilatkozat beállítása.
- 7.) Különböző tételtípusok kezelése különböző űrlapok alkalmazásával.
- 8.) Adatcsomagok befogadása (más informatikai rendszerekből) a SWORD protokoll alapján.
- 9.) Tételek csoportosítása (gyűjtemények, "set"-ek kialakítása).
- 10.) Keresési funkciók támogatása: egyablakos és részletes keresés, keresés a metaadatokban és a teljes szövegben.
- 11.) Metaadatok aggregálásának támogatása az OAI-PMH protokoll alapján.
- 12.) Böngészési lehetőségek pl. szerző, évszám szerint.
- 13.) Felhasználói szolgáltatások: zárolt tételek igénylése, témafigyelés, stb.
- 14.) Statisztikák készítése.

13 <https://unpaywall.org/>

14 MaNDA DB – Szerzetesek világnapja https://mandadb.hu/cikk/1037167/Szerzetesek_vilagnapja
MTA KIK 180 éves a Szózat <http://vorosmarty.mtak.hu/> (Irodalmi adatbázis rész)

15 <https://tudoster.idea.unideb.hu/>

További, a szoftverek által nem támogatott feladatok

A nyílt forráskódú szoftverekhez általában számos közösségi bővítmény, plugin áll rendelkezésre. Amit az alapszoftver nem támogat, az is rendelkezésre állhat egy bővítőcsomagban - néha több változatban is.

Mindazonáltal az alapszoftverek nem segítenek abban, hogy az egyes tételtípusok számára rendelkezésre álló (általában módosítható) űrlapokat hogyan töltsék ki a felhasználók, hogyan ellenőrizzék az adatok megadását az adminisztrátorok. Az OAI-PMH protokoll a Dublin Core adatelemeket használja, ez a szabvány ad támpontokat, de ajánlatos ezen túlmenő útmutatókat készíteni. (Pl. a dátum elem év mezőjének kitöltéséhez az űrlap csak legfeljebb négyjegyű számokat enged bevinni. De itt is kérdés, milyen időszámítás szerint?)

Egy másik, a szoftverek által többnyire nem támogatott funkció a dokumentumok (digitális objektumok) olvashatóságának hosszú távú biztosítása. Itt ajánlatos az elfogadott adatformátumokat előre meghatározni (és szabványos formátumokat támogatni), és az adatformátum megfelelőségét a tételek ellenőrzése során vizsgálni.

Irodalom

Holl, András (2018) *Újragondolt repozitóriumok* In: Agria Média 2017. Eger, Líceum Kiadó. pp. 104-109. <http://publikacio.uni-eszterhazy.hu/2343/>