

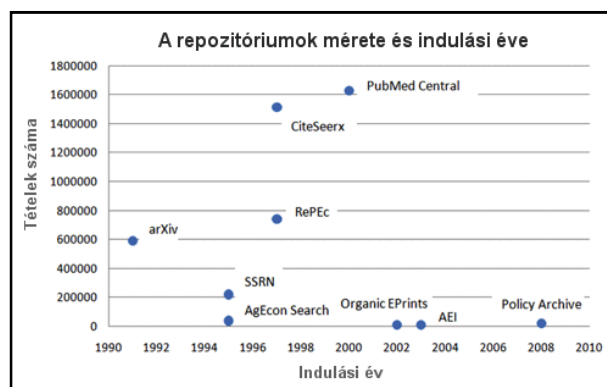


Trendek a nagyméretű tematikus repozitóriumok esetében

Miközben növekszik a különféle szakterületek publikációit és tudományos eredményeit gyűjtő repozitóriumok száma, viszonylag kevés praktikus információ olvasható még ezek fejlődéséről és menedzseléséről. Bár a digitális gyűjteményekhez általában vannak már követendő „jó gyakorlatok” és iránymutatások – például az összekapcsolhatóság, a metaadatok és a használhatóság terén –, de kifejezetten a tematikus repozitóriumok esetében még kevésbé vannak feltárva a közös jellemzők, pedig ezek ismeretében az ilyen gyűjtőhelyek kifejlesztésével megbízott projektirányítók és könyvtárosok megalapozottabb döntéseket hozhatnának. A szakirodalomban általában a *PubMed Central (PMC)*, az *arXiv* és a *RePEc* szolgáltatásokat emlegetik a tematikus repozitóriumok mintapéldájaként, de ezt a három nagy vállalkozást más és más műszaki és üzleti modell jellemzi, eltérő a beadási/gyűjtési politikájuk is, és különböző szerepet játszanak abban a szakmai körben, amelyet szolgálnak. Annak érdekében, hogy átfogóbb képet lehessen kialakítani a digitális archívumoknak erről a típusáról, és hogy bizonyos általános trendek is kirajzolódjanak, jelen tanulmány szerzői 2010 tavaszán a tíz legnagyobb tematikus repozitóriumról gyűjtöttek be információkat, kilencféle jellemzőt vizsgálva.

Az elemzés alá vont szolgáltatások kiválasztása az *Open DOAR*, a *ROAR* és a *Ranking Web of World Repositories* nyilvántartások adatainak összevetése után történt. A „top tíz”-be került szolgáltatások angol nyelvű, jellemzően preprint, illetve postprint publikációkat gyűjtő, egy vagy több szakterületre specializálódott nemzeti vagy nemzetközi archívumok. Az egyes gyűjtemények neve, alapítási éve és mérete az 1. ábrán látható. A repozitóriumok fele 2000 előtt indult; a legrégebbi az 1991-es születésű, multidiszciplináris jellegű *arXiv*, a legújabb pedig a közpolitikai témájú publikációkat gyűjtő *Policy Archive*, ami 2008-ban jelent meg. A legnagyobb tételszámmal a 2000-ben alapított *PubMed Central* büszkélkedhet, mely az élettudományokra specializálódott. A PMC gyors felfutására az a magyará-

zat, hogy a *National Institutes of Health* kötelezővé tette minden olyan publikáció beadását, amely a NIH által finanszírozott kutatás keretében készült. A legjobb tízben az informatika mellett a természet- és a társadalomtudományok dominálnak, a humán területek és a művészetek szinte teljesen hiányoznak (az SSRN kivételével, amely irodalomtudományi és filozófiai publikációkat is gyűjt). Ennek a jelenségnek többféle magyarázata is lehet: az eltérő publikációs arány, a preprint műfaj elfogadottsága, vagy a digitális médiához való viszony.



1. ábra A repozitóriumok adatai

Öt repozitórium saját rendszert használ, a másik öt pedig a DSpace (AgEcon Search és Policy Archive) illetve az EPrints (E-LIS, AEI és Organic Eprints) nevű ingyenes szoftverekre épül. Előbbiek között jellemzően régi gyűjteményeket találunk, ami érthető is, hiszen a nyílt forrású programok csak 1997 után kezdtek megjelenni ezen a területen. (Az AgEcon Search is saját fejlesztésű szoftverrel működött 1995-től 2008-ig, és csak ezután váltott a DSpace-re.) Ezzel magyarázható valószínűleg az is, hogy a legöregebb és legnagyobb archívumok jóval kevesebb dokumentumtípust gyűjtenek, mint a fiatalabbak, mert a régebbi rendszerek inkább csak a cikkek és hivatkozások nyilvántartására alkalmasak, míg az újak kezelik a jelentéseket, a konferenciaanyagokat, a disszertá-

ciókat, a multimédia dokumentumokat, a szoftveket, az adatállományokat és még sok minden mást is. Az egyetlen típus, amely mind a tíz repozitóriumban megtalálható, az a folyóiratcikk.

Ami a beadási szabályokat illeti: szinte mindenhol van valamilyen „moderálás” a minőség biztosítása érdekében. Az AEI, a Policy Archive, az Organic EPrints és az E-LIS rendszerekben a beadóknak regisztrálniuk kell, és a beküldött dokumentumok is csak jóváhagyás után kerülhetnek a gyűjteménye. Az SSRN és az AgEcon Search esetében is kötelező a regisztráció, de a feltöltéskor nincsen ellenőrzés. A PMC csak lektorált cikkeket fogad el a *National Library of Medicine* által élvonalbelinek definiált élettudományi folyóiratokból, illetve olyan szerzői kéziratokat, amelyek valamely nagy szervezet vagy intézet által finanszírozott kutatás keretében készültek. A RePEc esetében a beadás intézményi vagy részegység szintű helyi archívumokon keresztül történik. Az arXiv megint másfajta politikát alakított ki: egy publikáció akkor kerülhet az archívumba, ha azt egy „kezes” előbb jóváhagyta. Ilyen kezes státuszt azok érdemelnek ki, akik egy adott számnál több publikációt tölthettek már fel az illető témában. A számítástechnikára és informatikára specializálódott CiteSeer^x robotja pedig maga gyűjti be a webről az anyagokat, illetve a szerzők is megadhatnak egy e-mail címet és egy linket, ahonnan azután a rendszer automatikusan letölti és leindexeli a dokumentumot. Az, hogy a négy legnagyobb repozitóriumnak teljesen eltérő gyarapítási politikája van, azt jelzi, hogy a különféle gyűjteményépítési stratégiák egyaránt sikeresek lehetnek. Tanulságos az is, hogy bár ezeknél az archívumoknál természetesen nincsen olyan szakmai lektorálás, mint a tudományos folyóiratoknál, de egyéb módszerekkel: például a beadók megszűrésével, a tematikus korlátozással, a részletes metaadatok megkövetelésével mégiscsak képesek a minőséget fenntartani. Vagyis az *Open Access* mozgalom mellett, hogy elhárítja a hozzáférési akadályokat, nem jelenti egyben a szűrők teljes megkerülését is.

Egyik repozitórium sem igényli a szerzői jogok átruházását, de a többségük megkívánja, hogy a beadó ne adjon nekik kizárólagos jogot a dokumentum szolgáltatására. A PMC esetében a folyóiratok beküldhetik a teljes tartalmukat, vagy csak a NIH által finanszírozott kutatások publikációit, vagy csak válogatott cikkeket, anélkül, hogy a copyrightot átengednék az archívumnak. Az esetleges jog sértésekből származó következményeket mind egyik szolgáltatás a beadókra hárítja. Nem is le-

hetne elvárni természetesen, hogy szerzők és kiadók százainál lenyomozzák minden egyes cikk jogi státuszát, vagy hogy egyedileg segítséget nyújtsanak ehhez a beadónak. Ehelyett általában csak egy tájékoztatót tesznek közzé, amely a szerzők jogait ismerteti. Az arXiv az egyetlen a tíz közül, ahol nem vonható vissza a szolgáltatásra adott felhatalmazás, a többiekénél kérésre eltávolítják a korábban beküldött dokumentumot, bár többnyire nem szívesen teszik, tekintve hogy ez elmentmond az archívum céljának és pluszmunkával is jár.

A tízből nyolc szolgáltatás felsőoktatási intézményekhez kötődik, közülük négyet egyetemi könyvtárak üzemeltetnek (arXiv, AgEcon Search, Policy Archive, AEI). A PMC-t is egy könyvtár tartja fenn: a National Library of Medicine, vagyis az amerikai orvostudományi szakkönyvtár. Vannak olyanok is, amelyeket több intézmény működtet konzorciális formában: például a gazdaságtudományokra specializálódott RePEc archívum egyes részgyűjteményei több ország egyetemei és kutatóintézetei között vannak szétosztva. A társadalom- és bölcsészettudományi publikációkat gyűjtő SSRN pedig jelenleg egy kiadványt kezelésében van.

A finanszírozási források igen változatosak: állami, intézményi, szakmai szervezeti és piaci szereplőket egyaránt találunk a támogatók között, és gyakori, hogy egy repozitórium több lábon is áll, többféle pénzforrást is felhasznál. Egyes esetekben egy idő után megváltozott a szponzorálási forma: az SSRN-t például eredetileg néhány egyetem és kutatóintézet tartotta fenn. Az arXiv a fenntartó intézmény, a *Cornell University Library* saját költségvetése mellett 2010 elején egy újfajta támogatási mód bevezetésével is kísérletezni kezdett: felkérte az archívum 200 legaktívabb intézményi használóját, hogy önkéntes adományaikkal járuljanak hozzá a szolgáltatás működtetéséhez minden évben. A változatos finanszírozási formák és ezek átalakulásai az évek során azt jelzik, hogy a fenntarthatóság komoly kihívás a tematikus repozitóriumnak is, akárcsak a legtöbb digitális szolgáltatásnál.

A finanszírozást, az üzemeltetést és az irányítást ritkán végzi egyazon szervezet. Öt esetben külső irányítótestület segíti a döntéshozást. A PMC-nél például a NIH igazgatója által kinevezett kurátorok vannak. Az arXiv-ot működtető könyvtárosok egy fizikusokból, csillagászokból, matematikusokból és más szakterületek képviselőiből álló bizottságtól kapnak segítséget. Az SSRN kuratóriuma üzleti,

gazdaságtudományi, jogi és pszichológiai, az AgEcon Search-é pedig gazdasági, mezőgazdasági és kormányzati szakértőkből áll. A könyvtár- és információtudományra specializálódott E-LIS szolgáltatást három bizottság irányítja: egy adminisztratív, egy műszaki és egy szerkesztői.

A tíz legnagyobb tematikus repozitórium jellemzőinek összehasonlítása alapján a következő általános megállapítások tehetők:

- a legnagyobb gyűjtemények még 2000 előtt indultak (a PMC kivételével);
- a legtöbb ilyen archívum inter- vagy multidiszciplináris;
- a természet- és társadalomtudományok uralkodnak a témák között;
- az 1997 előtt született szolgáltatások többnyire saját fejlesztésű szoftvert használnak;
- a „cikk”, vagyis a pre- és a postprint jelenti az egyetlen közös dokumentumtípust;
- a beadási folyamat valamilyen módon moderált;
- a törlést/visszavonást ellenzik;
- a copyright ügyében a beadók a felelősek;
- a legtöbb szolgáltatást egyetemi könyvtárak vagy karok üzemeltetik.

Természetesen a tíz legnagyobb tematikus repozitórium nem reprezentálja azt a 150-400 további hasonló szolgáltatást, amelyeket a különféle *open archive* nyilvántartásokban eddig regisztráltak. Ez

a tíz is eléggé eltér egymástól, mert például az első és a második között méretben nagyobb a különbség, mint az utolsó öt összesített tételszáma, vagyis már a toplistán is észlelhető a *long tail* jelenség: a néhány nagy után sok kicsi következik. A méret nem lehet egyedüli mérőszáma a sikernek, hiszen vannak olyan új, vagy szűk, illetve interdiszciplináris szakterületekre specializálódott gyűjtemények, amelyekben ugyan nincsen sok publikáció, de felhasználói körük aktív és elkötelezett, és fontos szerepet játszanak az adott területen. Különösen ezeknek a kisebb archívumoknak volna hasznos, ha kialakulnának olyan szabványos és igény szerint alakítható értékelési módszerek, amelyekkel objektívabb összehasonlításokat lehetne végezni. Azt, hogy egy repozitóriumot mennyire ismer el az általa szolgált szakmai közösség, mérni lehetne például a közösség méretének és a feltöltött, illetve letöltött dokumentumoknak az arányával; vagy a szerzők beadási és archívumhasználati hajlandóságával; vagy azzal, hogy a repozitórium mennyire képes dokumentálni az adott szakterület növekedését és változásait.

/ADAMICK, Jessica – REZNIK-ZELLEN, Rebecca: Trends in Large-Scale Subject Repositories. = D-Lib Magazine, 16. köt. 11–12. sz. 2010./

(Drótos László)