



## A tudományos adattárak helyzete 2005 tavaszán

### Bevezetés

2005. május 10–11-én szervezte meg Amszterdamban a CNI (*Coalition for Networked Information*, <http://www.cni.org>), a JISC (*Joint Information Systems Committee*, <http://www.jisc.ac.uk>) és a holland SURF-alapítvány (<http://www.surf.nl.home>) azt a nemzetközi konferenciát, amelyen egyetemi adattárak (institutional repositories) helyzetét és a velük kapcsolatos országos stratégiákat értékelték. Előkészületként ugyanazzal a kérdőívvel mérték fel az adattárak helyzetét 13 országban: Ausztráliában, Kanadában, az Egyesült Államokban, Belgiumban, Franciaországban, az Egyesült Királyságban, Dániában, Norvégiában, Svédországban, Finnországban, Németországban, Olaszországban és Hollandiában. Bár a kérdések egy részére nem tudtak megfelelő választ adni, ez a felmérés képezte a konferencia alapját, ahol a résztvevők – a többi ország válaszainak ismeretében – pontosíthatták, illetve pótolhatták adataikat. Ezzel a lehetőséggel azért is éltek, mert a válaszok alapján az egyes országok helyzete sok esetben összehasonlíthatatlan volt, továbbá a gyors fejlődés és növekedés következtében néhány helyen kiigazításra szorult. A CNI az amerikai tagok körében részletesebben, több kérdéssel végezte el a felmérést, ezért a válaszok az USA tekintetében sokszor bővebbek voltak, és pontosabb képet festettek az adattárak helyzetéről. A szerzők szerint a témában ez volt az első komoly méretű és mélységű felmérés, amely értékes tapasztalatokkal szolgálhat a felsőoktatási, a kutatói és a politikai döntéshozó testületek számára, valamint pontosíthatja a hasonló felméréseket a jövőben.

### A tudományos adattárak helyzete

A következő adatok alapján próbálták meg felmérni az adattárak helyzetét országokként:

- az adattárak száma,
- az adattárak elterjedtsége,
- az adattárak tételeinek száma és típusa,

- az adattárak szakterületek szerint,
- az adattárak megvalósításához használt szoftverek.

### Az adattárak száma

A kérdőív az országban található tudományos adattárak számára kérdezett rá, a válaszadóra bízva annak eldöntését, hogy mi tekinthető tudományos intézmény adattárának. A válaszként kapott értékek (1. táblázat) országokként nagymértékben eltérnek, mind összességében, mind az egyetemek számához viszonyítva.

1. táblázat  
Egyetemi adattárak 13 országban

Ország	Adattárak száma	Egyetemek száma	Adattárak aránya az összes egyetemhez képest	Átlagos tételszám
Ausztrália	37	39	95%	na*
Belgium	8	15	53%	450
Kanada	31	na	–	500
Dánia	6	12	50%	na
Finnország	1	21	5%	na
Franciaország	23	85	27%	1000
Németország	103	80	100%	300
Olaszország	17	77	22%	300
Norvégia	7	6	100%	na
Svédország	25	39	64%	400
Hollandia	16	13	100%	3,000/12,500
Egyesült Királyság	31	144	22%	240
Amerikai Egyesült Államok	na	132	–	na

\*na: nincs adat.

## Az adattárak elterjedtsége

A felmérés a következőképpen próbál meg becslést adni az adattárak elterjedtségére: az adattárral rendelkező egyetemek számát az egyetemek számához viszonyítja. Ez csak becslés, mert országonként eltérő az összes egyetem száma (egyres válaszadók a főiskolákat és az üzleti iskolákat is hozzászámolták). Az USA-ban például 2364 felsőfokú képzést adó intézmény található, és ezek közül 250 minősül doktori iskolával is rendelkező kutatóegyetemnek. Minthogy a 132 válaszból 97 sorolható közéjük, így e kör felülreprezentált a felmérésben.

A másik ok, ami a becslést pontatlanná teszi, az a feltételezés, hogy egy intézmény csak egy adattárat üzemeltet. Az intézmények szakterületenként, vagy a tartalom típusa alapján külön adattárakat építenek, nem beszélve arról, hogy egy intézményhez több főiskola, üzleti iskola is tartozik. A becslés eredménye nagy szórást mutat Finnország 5%-os eredményétől a 100%-os Németorszáig, Norvégiáig és Hollandiáig, ahol az adattárak már szerves részét alkotják a felsőoktatás közös infrastruktúrájának. Az egyesült államokbeli felmérés az intézmények 40%-ában mutat ki adattárat, míg a többiek 88%-a tervezi létrehozásukat.

## A tételek száma

Az adattárban lévő tételek száma is nagy eltérést mutat országonként és országokon belül: néhány rekordtól a több százézes rekordszámig terjed az adattárak mérete. Azokban az országokban, amelyeknél megfelelő adatok birtokában ki lehetett számolni, az átlagos rekordszám ezer alatti, kivéve Hollandiát, ahol ez az érték 12 500.

Válaszadónként eltérő az adattárakban lévő rekordok értelmezése is. Míg például az USA-ban az adattárakban képek, szövegek, egész tartalmak is megtalálhatók, addig pl. Hollandiában az adattárak egy része csak metaadatokat tartalmaz, amelyek figyelembevétel nélkül csak 3000 lenne az átlagos rekordszám. A jövőbeli felmérések szempontjából tehát lényeges kérdés, hogy mit is tekintsenek egy rekordnak, mivel a rekordok egyre összetettebbek, és egyre több verziójuk születik. Az adattár tartalmának sokfélesége miatt (videoanyagoktól kezdve a szoftvereken át a szövegfájlokig) nem lehet az adattárakat összehasonlítani a felhasznált tárterületek alapján sem. Későbbi felméréseknél talán az adattárak növekedési ütemét

összehasonlítva lehetne következtetést levonni az adattárak méretére.

## A tételek típusa

A felmérésben csak a tudományos adattárakban leggyakrabban előforduló tételek száma szerepelt. A legszembevetőbb a szöveges tartalmak óriási hányada (2. táblázat). Eltérés csak a szöveges tartalmak típusai között észlelhető. Norvégiában pl. a tételek 90%-a (tan)könyv vagy szakdolgozat, míg Franciaországban a cikkek aránya 80%. Ausztráliában az elsődleges kutatási adatok aránya 83%, bár a válasznál jelezték, hogy ez a szám önkényes becslés. Az USA adattáraiban sokkal szélesebb a tételek megjelölhető típusainak köre, mint a kérdőívben felsoroltaké, így az USA-ban kimagasló a nem szöveges tartalmak aránya.

A számokból egyértelmű következtetés vonható le az adott ország adattáraitra vonatkozó elképzelésekről. Ausztrália és az USA kivételével a válaszadó országok adattárai elsősorban szöveges tartalmakat tárolnak: tudományos publikációkat, cikkeket, könyveket, szakdolgozatokat és disszertációkat, illetve kutatási eredményeket. Vagyis, ezen országok tudományos adattárainak építésében nagyobb szerepet kapott a tudományos publikációk szabad hozzáférése, míg Ausztráliában és az USA-ban az oktatással és a kutatással kapcsolatos elektronikus adatok is előtérbe kerültek. Ez az arány idővel eltolódhat az adattárak fejlesztésével. A különbségek megértéséhez az országok tudományos tartalmainak kezelésével kapcsolatos intézményi és országos stratégiát is ismerni kell. Az Egyesült Királyságban pl. egy publikációt szívesen helyeznek el a megbízható és értékes nemzeti adattárban, és utána egy helyi adattárban is. Az USA-ban azonban nincs megfelelő szintű és elterjedtségű nemzeti adattár, ezért a szerzők inkább a helyi adattárakat részesítik előnyben.

## Az adattárak szakterületek szerint

A felmérés az adattárakat öt szakterület szerint csoportosította a kérdőívben: humán és társadalomtudományok, orvostudomány, természettudomány, műszaki tudományok és művészet. A válaszoló országok között óriási különbségek láthatók (3. táblázat). Míg Ausztráliában és Olaszországban a humán és társadalomtudományok dominálnak, addig az Egyesült Királyságban a műszaki és természettudomány fedi le az adattárak kétharmadát. Svédországban és Hollandiában az adattárak aránya szakterületenként kiegyenlítettnek mondható.

2. táblázat  
A tételek megoszlása (%)

Ország	Cikk	Szakdolgozat	Könyv	Elsődleges kutatási adatok	Videó	Oktatási segédanyag	Egyéb
Ausztrália	8	8	1	83		0	
Belgium	33	66					
Kanada							
Dánia							
Finnország							
Franciaország	80	20					
Németország	20	40-50			5	1	25
Olaszország	70	5				20	5
Norvégia	10	90					
Svédország	30	70					
Hollandia	20	40					40
Egyesült Királyság	74	16		1	4		4
Amerikai Egyesült Államok							

3. táblázat  
Az adattárak megoszlása szakterület szerint (%)

Ország	Humán és társadalomtudományok	Orvostudomány	Természet-tudomány	Műszaki tudomány	Művészet	Egyéb
Ausztrália	49	19	17	9	3	3
Belgium	33	39	16	11		
Kanada						
Dánia						
Finnország						
Franciaország	18	1		4		67
Németország	5	5-10	20	10-20	5	25
Olaszország	55	10	20	15		
Norvégia						
Svédország	30	20	30	20		
Hollandia	20	20	30	20	1	19
Egyesült Királyság	16	12	25	41		6
Amerikai Egyesült Államok						

Franciaország az egyéb kategóriára 67%-ot jelölt meg, amelybe főleg fizikai és matematikai tartalmakat sorolt, míg Németországnak az egyéb kategóriába írt 25%-a elsősorban informatikai témájú tartalmat jelent.

#### Az adattárak szoftverei

Az adattárakban használt szoftverek tekintetében is jelentős eltérés mutatkozik az egyes országok között (4. táblázat), bár néhány szoftvercsomagot sok helyen alkalmaznak. Az *EPrints* 13-ból 7 országban használatos, a *DSpace* pedig 11-ben. Az

adattárak egy részében saját fejlesztésű szoftvereket alkalmaznak. Az *OPUS* Németországban 44 helyen van jelen, Ausztráliában pedig a *Virginia Tech* csomagja terjedt el: 27 helyen alkalmazzák. Az USA-ban egyértelműen a *MIT*-en kifejlesztett *DSpace* az elterjedtebb (58%), bár sok helyen saját fejlesztésű rendszereket használnak, összefüggésben a tárolt tartalmak sokszínűségével.

#### Oktatói részvétel

A felmérés kitért azon oktatók számára is, akik legalább egy rekordot elhelyeztek valamelyik adat-

tárban. Ez az adat értékes kiegészítés az adattárakkal foglalkozó korábbi felmérésekhez képest, mivel azok csak egy-egy tanszék adattárát és digitalizálási, publikálási stratégiáját vizsgálták. A kevés helyről érkezett válaszok mindegyike csekély, kivéve Hollandiát és Németországot. Hollandiában kb. 40% azoknak az oktatóknak az aránya, akik legalább egy rekordot elhelyeztek egy adattárban. Németország is kivételt jelent, mivel disszertációknak a szakterülettől függően 2–62%-át elhelyezik legalább egy adattárban.

Azzal a becsléssel is kevés válaszadó próbálkozott, hogy az évente megjelenő kutatási eredmények mekkora hányadát helyezik el országosan valamilyen adattárban. Hollandiában a szakterületek széles körét figyelembe véve ez az arány 25%. Belgium francia nyelvű intézményei körében az arány 33% a humán és társadalomtudományoknál, 39% az orvostudományban, 11% a műszaki, és 16% a természettudományoknál. Németországban a természettudományok területén 10%, az Egyesült Királyságban a műszaki tudományokban és az informatikában 15%-os az arány. Több értékelhető válasz nem érkezett.

Egy rekord adattárban való elhelyezésének folyamatával kapcsolatban az tűnik ki a válaszokból, hogy a legtöbb esetben nem a szerző vagy a szer-

ző munkatársa látja el ezt a feladatot, hanem mások, például könyvtárosok. Kivétel ez alól néhány olasz egyetem, ahol maguk a szerzők végzik ezt a munkát. A felmérésből az is kiderült, hogy az adattárak többsége egy korábbi adattár továbbfejlesztése.

Az USA-ban a válaszokból leszűrhető, hogy az adattárak ma már többnyire inkább az egyetemi infrastruktúra szerves részévé válnak, illetve az adattárral nem rendelkező egyetemek 88%-a tervezi adattár kialakítását. A nem a kutatóegyetemek közé tartozó felsőoktatási intézmények elsősorban az oktatással kapcsolatos anyagokat tervezik elhelyezni adattárakban. Ezen intézményeknek aggályaik vannak egy adattár kialakítási és fenntartási költségeivel kapcsolatban, ezért egyre inkább több intézmény alkotta konzorciumok létrehozásában, és közös adattárakban gondolkodnak. A tapasztalatok egyelőre azt mutatják, hogy a konzorciumok projektjei hosszabb időt vesznek igénybe a tervezési és az implementálási idő miatt. Viszont amint elkészült egy ilyen adattár, máris több intézmény válik adattárral rendelkező intézménnyé. Az adattárak feltöltésébe egyre gyakrabban vonják be a hallgatókat. Őket sokszor könnyebb rávenni arra, hogy a szakdolgozatukat, vagy egyéb komolyabb munkájukat elhelyezzék egy-egy tanszék adattárban, mint magukat a tanszékeket.

#### 4. táblázat

##### Az adattárakhoz használt szoftverek

Ország	GNU Eprints	DSpace	CDSWare	ARNO	Fedora	DIVA	iTOR	Egyéb
Ausztrália	7	3			3			29
Belgium		2						
Kanada		**						
Dánia		2				1		3
Finnország								1
Franciaország	11	2						3
Németország	2	2						103
Olaszország	7	3	1					
Norvégia		1				1		3
Svédország	3					10		12
Hollandia		6		6			2	3
Egyesült Királyság	24	6						2
Amerikai Egyesült Államok	+	+			+			+

\* +: legalább egy intézmény.

## Az adattárak együttműködése

A kérdőív kitért az adattárak nemzeti vagy nemzetközi együttműködésére is (pl. különböző adattárakban való keresés, nyomtatás, vagy a metaadatok automatikus elhelyezése több adattárban). Ilyen együttműködések, több adattárra épülő szolgáltatások már léteznek (pl. *CompQuest*, <http://www.surf.nl/download/country-update2005.pdf>), ám egyelőre inkább csak kísérleti jelleggel, ezért még korai lenne megítélni, hogy melyik szolgáltatás sikeres, és válik majd széles körben elterjedté. A kérdések ezért az adattárak közötti szolgáltatás létrehozására, és az adatokhoz való hozzáférésére irányultak. A válasz arra, hogy milyen intézménynek kellene ilyen szolgáltatást létrehozni, vagy legalább támogatni, egyértelmű azokban az országokban, amelyek erős központi nemzeti szervezettel rendelkeznek, mint pl. a JISC Angliában vagy a SURF Hollandiában.

Az adattárak együttműködését megvalósító szabványok közül a legelterjedtebbnek a metaadatokat összegyűjtő *OAI-PMH* protokoll tűnik, illetve a keresőfelület megvalósításánál az *SRU/SRW/Z39.50*. Ezekkel még korántsem oldódott meg minden probléma. Sok kérdésre közösen kellene megoldást találni, mint pl. állandó és egyedi azonosító alkalmazása az adattárak elemeire.

Érdemes megemlíteni, hogy az USA adattárainak többsége felkészült más adattárakkal való együttműködésre (pl. az *OAI-PMH* protokollon keresztüli metaadatcserére, amellyel az *Illinois Egyetem* már gyűjt metaadatokat), a gyakorlatban ezek egyelőre csak elvétve valósulnak meg. Úgy tűnik, mintha az amerikai intézmények csak egy együttműködést koordináló intézményre, szervezetre várnának.

## Stratégiák és szervezetek

Bár csak kevés válaszadó országban létezik nemzeti (akár kormány szintű) stratégia az adattárak fejlesztésére, a nagyobb intézmények szintjén minden országban léteznek direktívák, deklarációk. Az országok válaszai a fejlesztési tervekről (*CompQuest*) értékes és érdekes képet mutatnak, amelyek nemcsak az adattárak fejlesztéséről szólnak, hanem a tudományos élet szereplőinek kommunikációjáról, a tudáshoz való szabad hozzáférésről éppúgy, mint a *Berlini deklaráció* adaptálásáról.

Az USA-ban sincs még nemzeti szintű stratégia az adattárak fejlesztésére, bár a *Nemzeti Egészségügyi Szervezet (National Institutes of Health)* és a *Nemzeti Tudományos Alap (National Science Foundation)* anyagi támogatásaival, támogatási rendszerével serkenti az adattárak fejlesztését. A fejlesztések munkacsoportok vagy konzorciumok általi támogatása, koordinálása, segítése máshol is megfigyelhető. Angliában és Hollandiában pl. nemzeti programokat hoztak létre adattárak fejlesztésére és a fejlesztések támogatására. Németországban egy állami szervezet hitelesíti az adattárakat a használt szabványok szerint. Máshol vannak hasonló törekvések, pl. Hollandiában, ahol egy nemzeti koordinációs, illetve finanszírozó szervezet igyekszik egységesíteni a szabványokat.

A felmérés a stratégiákkal kapcsolatban nem tért ki arra, hogy egy-egy adattár adminisztrációjáért és az adattárral kapcsolatos döntések meghozataláért ki a felelős – a jövőben ezt érdemes lenne pótolni –, az USA-ból mégis érkezett válasz ezzel kapcsolatban. Itt az intézmények 80%-ánál az adminisztráció egyedüli felelőse a könyvtár, 60%-ánál pedig a könyvtár az egyedüli döntéshozó. A többi intézménynél az adminisztratív és a döntéshozói felelősség megoszlik a könyvtár és az egyetem, vagy a könyvtár és a tanszék, vagy a könyvtár és a digitalizálási csoport, illetve egyéb egységek között. A könyvtárak persze sok esetben csak papíron az egyedüli a döntéshozók, legtöbbször az egyetem megfelelő egységeivel közösen döntenek. Bárhol is legyen megosztva a felelősség, a könyvtárak kiépített infrastruktúrája sokszor biztos alap egy adattár számára.

## Az adattárak fejlesztése

A válaszadók hosszú listát állítottak össze az adattárak kialakításának, feltöltésének és fenntartásának nehézségeiről. Ezek többségét a szűkös erőforrások és a szerzők, illetve az intézmények aggályai jelentik. Nehéz meggyőzni egy tanszéket, hogy érdemes kialakítani és használni egy adattárt. Félelmet kelt a nem pontosan ismert szerzői jog fogalma, valamint az a hiedelem, hogy az adattárak tartalmi rossz minőségűek. Gyakran említik hiányként az intézmények kutatási eredményeinek adattárakban való elhelyezési kötelezettségét. Világos tehát, hogy mindent meg kell tenni az intézmények segítésére, munkájuk könnyítésére. Az adattárak elterjedésének, fejlesztésének serkentésére több megoldást is felsoroltak az országok. Ausztrália és Belgium az implementálási és üze-

meltetési folyamatok egyszerűsítését hangsúlyozta, mások az intézmények, könyvtárak hatékony támogatását emelték ki.

Az adattárak iránti érdeklődést többféleképpen lehet serkenteni. Megoldás lehet az adattár tartalmának indexelése a Google keresőjében, vagy az a holland projekt (*CoS = Cream of Science Projekt*, <http://www.creamofscience.org>), amely a nemzeti adattárat népszerűsíti a tudomány vezetőinek publikációival. Belgiumban külön szolgáltatás létezik a szerzőknek, amely publikációs listáik létrehozását és karbantartását segíti, hasonlóan egy másik holland projekthez, az *E-Depot*hoz (<http://www.kb.nl/dnp/e-depot/e-depot-en.html>). Sok országban egy adattár nemcsak nagyobb hozzáférést jelent, hanem az anyagi támogatásokat is befolyásolja. Belgiumban egy csoport azért lobbizik, hogy az intézmények az adattárak alapján osszák el támogatásaikat az egyetem szervezetei között. Így a tanszékek adattárainak kialakítását egyfajta verseny is segítené. Az USA-n kívüli országok adattárainak fejlesztését, támogatáshoz való hozzájutását és elterjedtségét az adattárak kevés típusú tartalma is befolyásolja (az USA-ban az adattárak komplex tartalmakkal is rendelkeznek, mint például adathalmazokkal, szoftverekkel, kutatási szimulációkkal, egyetemi hallgatók digitális munkáival).

### **Összegzés**

A CNI, a JISC és a SURF 2005 tavaszán végzett felméréséből kiderül, hogy a tudományos adattárak manapság az egyetemi infrastruktúra szerves részeivé válnak, így felelve meg a digitális kor kihívásainak, biztosítva a szabad hozzáférést a

tudomány és az oktatás eredményeihez egyetemen belül és kívül is. A felmérésre választ adó országok jövőre vonatkozó elképzelései egyértelműek: a tudományos adattárak fejlődése és szélesebb körű elterjedése várható. Ez nem csak új adattárak felállítását jelenti: az USA-n és Ausztrálián kívül az adattárakban a jelenleg leginkább szöveges publikációk mellett a jövőben új típusú tartalmak megjelenése várható. Országoként eltérő, ám szinte mindenhol kezdetleges stádiumban vannak a nemzeti szintű szolgáltató-, szabványosító szervezetek, illetve az adattárak együttműködésére épülő szolgáltatások, ezért a jövőben ezek erősödése, fejlődése is várható.

A válaszadó országok az amszterdami konferencián szorgalmazták a felmérés megismétlését, esetleg rendszeressé tételét. Ezzel kapcsolatban a felmérést lebonyolító szervezetek egyeztetésbe kezdtek, hogy a jövőben egyértelműbbé (pl. az adattárak méretének kérdésében) és még szélesebb körűvé tegyék a felmérést.

**van WESTRIENEN, Gerard–LYNCH, Clifford A.: Academic institutional repositories: Deployment status in 13 nations as of mid 2005. = D-Lib Magazine, 11. köt. 9. sz. 2005. <http://www.dlib.org/dlib/september05/westrienen/09westrienen.html>**  
**LYNCH, Clifford A.–LIPPINCOTT, Joan K.: Institutional repository deployment in the United States as of early 2005. = D-Lib Magazine, 11. köt. 9. sz. 2005. <http://www.dlib.org/dlib/september05/lynch/09lynch.html/>**

(Somogyi Tamás)