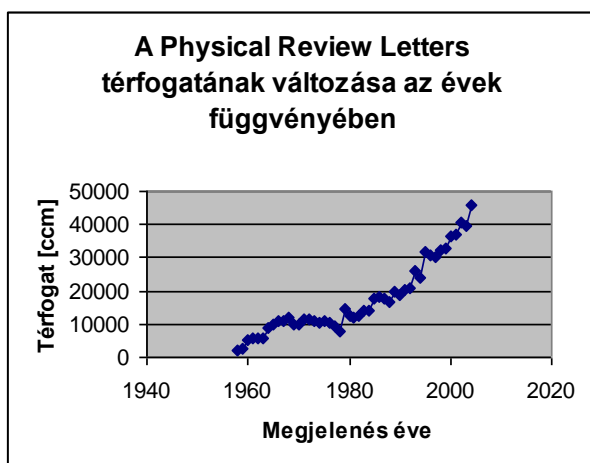


Quo vadis, tudományos publikációs adattár?

Dolgozatunkban öt reprezentatív hazai tudományos publikációs adattárat hasonlítottunk össze abból a célból, hogy felderítsük fejlődésük megtorpanásának okait. Vizsgálódásunk eredménye azt mutatja, hogy az adattárak működésében mutatkozó anomáliák a tudományometriai értékelés lehetőségének elmaszátolása és a tervezési szakaszban történt, a működtetés feltételeit figyelembe nem vevő aránytévesztések következményei.

Bevezetés

Napjaink egyik, talán nem is megértett közhelye a tudomány viharos fejlődése. Ennek az állításnak megértését számos megjelent tudományos közlemény próbálja elősegíteni, mi itt szemléltetésül az 1. ábrát mutatjuk be, amely a fizika tudományának legszínvonalasabb, a fizika minden ágában cikket közlő folyóiratának, a *Physical Review Letters* térfogatának változását mutatja a folyóirat 1958-as indulása óta. Tekintettel arra, hogy a folyóirat kezdetől fogva négyoldalas „Letter”-eket közöl, ez a szám valószínűleg jól tükrözi a területen megjelent publikációk számának növekedését.



1. ábra A *Physical Review Letters* c. folyóirat bekötött évfolyamai térfogatának változása a megjelenési év függvényében

A tudományos-társadalmi hatásakon túlmenően a megjelenő tudástömeg elkerülhetetlenné tette egyrészt a művek cím szerinti nyilvántartását, amely a közelmúltban lejátszódó informatikai forradalom nélkül valószínűleg megoldhatatlan feladat lett volna, másrészt pedig a művek és a rájuk vonatkozó hivatkozások adatainak nyilvántartását abból

a célból, hogy ezekből az adatokból képet kaphassunk a szerzők munkásságáról (teljesítményéről). Részben ezeknek az igényeknek a kielégítésére jöttek létre az ún. publikációs adattárak, melyeknek első és legismertebb képviselője az amerikai *Institute for Scientific Information (ISI)*, mai nevén: *Thomson Scientific*) által összeállított *Science Citation Index (SCI)* (a weben ma elérhető a *Web of Science (WoS)*, illetve tágabban a *Web of Knowledge (WoK)* szolgáltatás keretében).

Intézményi publikációs nyilvántartások hazánkban az 1960-as évektől léteztek (az általunk ismert legrégebbi, az *MTA Atommagkutató Intézetének* széllyukkártyás megoldása volt). Jelentős késéssel, a hazai informatikai háttér megerősödésével jelentek meg a korszerű, webes felülettel bíró hazai publikációs adattárak (legkorábban, 1998-ban, az *APHA*, az *MTA Atomki publikációs-hivatkozási adatbázisa*)¹. Ma hazánkban a többé-kevésbé működőnek mondható adattárak száma mintegy húszra tehető. Jelen dolgozat motivációját az adja, hogy több hazai publikációs adatbázis fejlődése megtorpanni látszik. Időszerű áttekinteni a hazai publikációs adatbázisokat, hogy a továbblépéshez irányokat adhassunk².

A *Miskolci Egyetem* publikációs adattára (*MEPA*) és az *MTA Köztudományi Publikációs Adattára (KPA)* mellett részletesen foglalkozunk az *MTA Tudományos Publikációs Adattárával (ATPA)*, az *MTA Atommagkutató Intézet Publikációs-Hivatkozási Adatbázisával (APHA)* és a *Pázmány Péter Katolikus Egyetem Bölcsészettudományi Karának Oktatói Publikációs Adatbázisával (PPKE BTK OPA)*. Véleményünk szerint választásunk reprezentálja a hazai tudományos publikációs adattárak főbb típusait. Az *APHA* részletesebb ismertetését az indokolja, hogy bemutatása a hazai szakirodalomban még nem történt meg, a *PPKE BTK* adatbázisa

pedig jól működő példa kisebb bölcsészettudományi adatbázisra.

A tudományos publikációs tevékenység összefüggéseinek tudományos vizsgálatára külön tudományág, a *tudománymetria (bibliometria)* alakult ki. A hazai tudománymetriai kutatások legfrissebb áttekintését *Vinkler Péter* jegyzi³.

Minden bizonnyal történetileg a legelső teljesítményértékelési szempont a szerzők publikációinak tartalma volt. Az egyéni publikációs szerzőlisták hosszabbodásával előtérbe került a publikációk számának figyelembe vétele (egyes nekrológok ma is ezt használják). A tudománymetria kialakulásával, illetve az informatika fejlődésével vált gyakorlattá a hivatkozások számbavétele. Ettől az időtől kezdve egy sor tudománymetriai statisztikai jellemző definiálására került sor. Ezek közül számos megtalálható az APHA-ban, országokra vonatkozóan I. a 4. jegyzetet.

Használatának népszerű volta miatt meg kell említeni az *impaktfaktort*^{5,6}. Az idő előrehaladtával az *impaktfaktort*, illetve annak egy adott publikáció-halmazra felösszegzett (esetleg szerzőszámmal beosztott) változatát elkezdtek használni egy szerző, intézmény vagy ország tudományos teljesítményének jellemzésére⁷. Ebben az értelemben használják az általunk vizsgált adatbázisok is az egyetemi adattárak kivételével. Időközben a folyóiratok értékelésére kitalált számos mutatót (frissességi index, idézési félidő, idézettségi félidő⁸ stb.) kezdték a kutatói teljesítményértékelés területén is alkalmazni. Teret nyertek olyan számszerű jellemzők is, amelyek a publikálási szokásokkal vannak kapcsolatban (aktivitási index, produktivitás⁹ stb.). A legutóbbi idők terméke az egyes szerzők tudományos teljesítményének jellemzésére kifejlesztett Hirsch-index (h-index)¹⁰, amelyet hazánkban az elmondottaktól függetlenül kezdtek el alkalmazni, ami a KPA hiányosan feltöltött adatain alapul¹¹.

Az *evaluatív* (értékelő) tudománymetria széles körben elfogadott álláspontja szerint a tudományos teljesítmény értékelésekor körültekintően kell eljárni, és minden esetben ki kell egészíteni ún. *peer-review* értékeléssel. Hazánkban a szakértői (*peer-review*) értékelés gyakorlatilag végrehajthatatlan az egyes szakterületeken dolgozók kis száma és függősége miatt. Még ha lennének is független szakértők, a tartalmi áttekintés a szóban forgó művek nagy száma miatt nem lehetséges a gyakorlatban. Ezen okok miatt – bár senki sem szereti – elkerülhetetlen a tudománymetriai statisztikai

jellemzők mind szélesebb körű használata. A számszerű jellemzők használata elsősorban makroszinten (országok közötti összevetés) vezet megbízható eredményre, mezo- (intézményi) és mikroszinten (egyéni teljesítményértékelés) alkalmazása kockázatos és nagy körültekintést igényel¹².

Az *evaluatív* tudománymetria alkalmazását az alsóbb szintek felé haladva egyre fokozódó éleségű viták kísérik. Ennek az a magyarázata, hogy a felsőbb szervek (mezoszinten: intézményfenntartók, mikroszinten: munkáltatók) a tudománymetriai adatokat esetlegesen egzisztenciális következményekkel járó következtetések levonására használhatják. A viták általában az egyéni és a csoportérdekek kifejezői, tudatosabb esetekben a *tudománypolitikai marketing* (vö.¹³) lépéseinek folyamatai. Ez utóbbiak rendszerbe foglalása a jövő kutatásainak témája lehet.

A publikációs adattárak a láttatás és az értékelés szándékával jöttek létre. A hazai adattárak létrehozásának közvetlen oka az volt, hogy a létrehozást finanszírozó szervezetek nem tartották kielégítőnek a nagy nemzetközi adattárak (WoK, Scopus, Medline stb.) által nyújtott szolgáltatásokat. Ez annak ellenére így történt, hogy az ISI megteremtette a társadalom- és bölcsészettudományok területén is az SCI megfelelőit (*Social Science Citation Index* és *Arts & Humanities Citation Index*). Míg a természettudományok hivatkozási indexei jelentős részben tartalmazzák a kiemelkedő magyar kutatók legutóbbi évekbeli munkásságát, azokon a területeken, ahol a magyar nyelvű közlés népszerűbb, illetve a tudományos teljesítmény nyomai a nemzetközi adatbázisok által nem indexelt művekben találhatóak, nagy igény mutatkozik a közlemények adatainak felmutatására. Ennek az igénynek a felismerése a dolgozatunkban vizsgált valamennyi publikációtár létrehozásának az oka. Megjegyzendő, hogy javarészt a létrehozó szervezetek eltérő érdekei vezettek a viszonylag nagyszámú adattár létrejöttéhez (többek között majdnem minden felsőoktatási intézmény törekszik valami hasonló kialakítására). Az egyetemi publikációs adattárak többnyire megelégszenek (valószínűleg azért, mert a legtöbb helyen a már működő könyvtárak kapták meg ezt a feladatot) a publikációk adatainak láthatóvá tételével.

Új korszakot jelentett a tudományos teljesítmények értékelésének nemzetközileg, majd hazánkban is kialakuló gyakorlata. Ez direkt módon hatással volt az egyes kutatók/intézmények sorsára. Közvetle-

nül ez – az MTA intézeteinek 1996-os értékelése¹⁴ – volt a kiváltó oka az APHA weben való megjelenésének, és egy majdani intézeti teljesítményértékelés terve¹⁵ vezetett az ATPA megalapításához. Az ATPA-t szervezetenként az MTA főtitkára alá rendelték. A KPA létrehozásának az MTA elnökétől kiinduló indítéka pedig az volt, hogy az MTA levelező és rendes tagjain kívül a köztestületi tagoknak (akik az MTA „résztevői”) lehetőséget nyújtsanak munkásságuk nyilvántartására, a weben való megjelenéséhez. Később, vélhetőleg a gyér érdeklődés miatt, ez az eredeti cél módosult. Egyre szélesebb körben ajánlották – attól függetlenül, hogy másutt már lehetséges munkásságuk figyelemmel kísérése – a KPA használatát ösztöndíjak, doktori cím, akadémiai tagság, Bolyai-ösztöndíj megpályázása esetén. Néhány éve az *Országos Tudományos Kutatási Alap (OTKA)* is a KPA adatait teszi ajánlottá, legújabbán pedig az *Országos Doktori Tanács (ODT)* adatbázisa is ennek használatát favorizálja.

Az ATPA és a KPA egyre több tudományometriai statisztikai jellemzőt közöl, azonban az APHA anyaga sokkal részletesebb ilyen szempontból. Itt kell megemlíteni, hogy a dolgozatban vizsgált adattárakkal kapcsolatban célszerű elkülöníteni a szűkebben vett adatnyilvántartást, az adattárak működtetését, és az esetleges tudományometriai célokat szolgáló tudományometriai eszköztárat. Ezek a részletek első látásra függetlennek látszanak, azonban mindegyik mögött eltérő érdekek állnak, amelyek az egyes összetevők eltérő színvonalú megvalósításához vezethetnek.

A jelen dolgozatban vizsgált hazai tudományos publikációs adattárak széles spektrumot ölelnek fel. Egyfelől, főleg az egyetemi adattárak (valószínűleg a kezelőként funkcionáló egyetemi könyvtárak befolyására), a helyben keletkező publikációk (dokumentumok) könyvtárszakmai feltárását (kereshetőségét) helyezik előtérbe, a többiek viszont inkább a szerzők és intézmények publikációs (hivatkozási) listáit tárják a felhasználó elé.

A Magyar Tudományos Akadémia Tudományos Publikációs Adattára (ATPA)

Az adattár az *Akadémiai Kutatóhelyek Tanácsának* javaslatára az MTA főtitkárának 2001. február 15-én kelt határozatával jött létre. Az MTA Tudományos Publikációs Adattára az *MTA Kémiai Kutatóközpontjának* önálló szervezeti egységként mű-

ködik. Az ATPA célja, hogy az MTA természet- és társadalomtudományi kutatóhelyeinek tudományos közleményeit és az azokat idéző publikációkat nyilvántartsa. Adatokat szolgáltat az MTA bizottságai és szervezetei részére a kutatóhelyek tudományos publikációs tevékenységének áttekintéséhez, valamint különböző szempontok szerinti értékeléséhez. Segítséget nyújt a kutatóhelyeknek a publikációikat és azok idézeteit tartalmazó jegyzékek különböző szempontok (pl. kutatási pályázatok) szerinti összeállításához.

Az ATPA bekapcsolódik a hazai kutatóhelyek és más intézmények tudományos kutatóinak szakirodalmi információval történő ellátásába, segítve ezzel a kutatóhelyek tudományos együttműködését. Hozzájárul a kutatóhelyi, az akadémiai és az országos tudományos kutatási koncepciók, tervek kimunkálásához, a magyar tudomány nemzetközi helyzetének felméréséhez a szükséges információk megszerzésével és elektronikus formában való szolgáltatásával.

Az intézetek adatai 1992-től, az akadémiai kutatócsoportok adatai pedig 1996-tól szerepelnek az adattárban. A 2008-as adatok feldolgozása után az adattár több mint 300 ezer tételt tartalmazott (~100 000 publikáció, ~200 000 hivatkozás). Az MTA ATPA-ban megtalálható bibliográfiai adatok és listák:

- az MTA kutatóhelyek tudományos és ismeretterjesztő közleményei,
- a publikációkra érkezett idézetek,
- az Akadémia doktora értekezések és az MTA kutatóhelyeken készült PhD disszertációk,
- az MTA kutatóhelyek jegyzéke,
- a publikációkat tartalmazó tudományos folyóiratok jegyzéke.

Az akadémiai kutatóhelyek tudományos közleményeinek bibliográfiai adatai felhasználói név és jelszó nélkül lekérdezhetőek. Keresni lehet a kutatóhely vagy a szerző(k) neve szerint. A kérdés tovább szűkíthető (év, megjelentető folyóirat, címben előforduló szavak, illetve a publikáció típusa vagy nyelve szerint). Az eredménylisták több szempont szerint rendezhetőek, megjeleníthetőek és menthetőek. Az idéző publikációk adatainak lekérdezéséhez felhasználói név és jelszó is szükséges.

Annak a felhasználónak, aki saját maga akarja adatait bevinni az adattárba, az ATPA két programot bocsát rendelkezésére. Ezek a forráspublikációk és idézetek adatainak bevitelére szolgáló

„TPA Forrás” és „TPA Idézet” nevű programok. Azok az intézmények, amelyek saját adattárat működtetnek (pl. APHA), az említett programokat szerkesztőprogramként alkalmazzák az ATPA felé történő adatexportálás céljából.

A Magyar Tudományos Akadémia Köztudományi Közlövése Adattára (KPA)

A KPA célja az akadémiai köztestületi tagok publikációira és azok idézettségére vonatkozó adatok összegyűjtése és tárolása. Elsőként a 2004. évi levelezőtag-választás tagajánlási szakaszában jelölést kapott köztestületi tagok adatai kerültek be az adatbázisba. Erről az MTA Elnöksége a 2003. február 25-i, 1/2003. sz. állásfoglalásában határozott. A KPA felállításával és működtetésével a Köztudományi Adattár keretein belül az *MTA Kutatásszervezési Intézete* foglalkozik.

A KPA gyűjtőköre kiterjed az érintettek olyan publikációira és nyilvánosan hozzáférhető olyan egyéb tudományos teljesítményeire (pl. szabadalmak), valamint a rájuk vonatkozó olyan hivatkozásokra is, amelyeket a nemzetközi publikációs adatbázisok nem figyelnek. Például könyvek, a nemzeti tudományok magyar nyelvű publikációi, tankönyvek. Sajnos az adattár túlnyomó része globálisan nem kereshető, így külföldről használata erősen korlátozott.

A KPA személyekhez kötődően gyűjti az adatokat, és kiterjed az érintett személyek munkásságának teljes időszakára. A publikációs és idézettségi adatok eljuttatásának a módja az adattárba: 1.) Az adatbevitel első lépése: telefonon, vagy e-mail útján történő bejelentkezés a KPA illetékes munkatársainál. 2.) A bejelentkező felhasználói nevet és jelszót kap. A köztestületi tag nevében csak az általa megbízott személy vihet be adatokat. 3.) Az adatok bevitelére az adattárba.

Ha az adatszolgáltatónak van a Web of Science-ben, vagy más hasonló nemzetközi adattárban nyilvántartott publikációja és idézettségi adata, úgy ajánlatos az adatgyűjtést ezeknek az adatoknak a letöltésével kezdeni. Az így kapott adatfájlokat az internetes adatbevitel megfelelő helyén fel kell tölteni az adattárba. Szövegszerkesztővel készített publikációs és idézettségi listát az adattár nem tud fogadni. A kutatóintézetben dolgozó köztestületi tagok kérhetik publikációs és idézettségi adataiknak az ATPA-ból való áttöltését.

A rendszerben történő végleges tárolás előtt az adatok nyilvánvaló ellentmondás-mentességét a KPA munkatársai ellenőrzik, és az adatokat az adatszolgáltatónak ellenőrzésre visszaküldik. Csak az adatszolgáltató jóváhagyása után válnak az adatok honlapjukon mindenki számára elérhetővé.

Megjegyzendő, hogy a KPA több kлона működik az országban. Így a *Budapesti Műszaki Egyetem (BME-PA)*, az *MTA Kísérleti Orvostudományi Intézetében (KOKI-PA)* és a *Szegedi Biológiai Központban (SZBK-PA)*. Ezek működése kis mértékben, de eltér a KPA-étól.

A Pázmány Péter Katolikus Egyetem Bölcsészettudományi Karának Oktatói Publikációs Adattára (PPKE BTK OPA)

A PPKE BTK publikációs adattárjának – az egyetemi bibliográfiának – célja az oktatók tudományos közleményeinek nyilvántartása és egységes szerkezetű bemutatása¹⁶. Az adatgyűjtéskor a hangsúly a tudományos tevékenység során keletkezett publikációk dokumentálására helyeződik, ebből következik az, hogy az adattár nem törekszik a lehető legteljesebb személyi bibliográfia regisztrálására. A szakirodalmi tevékenység nyomán született publikációk körébe sorolják az önálló műveket (monográfiák, könyvrészletek, gyűjteményes kötetek tanulmányai, folyóiratcikkek, és más, nem nyomtatott médiában való önálló szereplés, illetve társalkotás, művészi produkció) és a különféle közreműködések során keletkezett műveket (fordítás, szerkesztés, szöveggondozás stb.), beleértve a folyóirat-szerkesztést és a könyvsorozat-szerkesztést is. A könyvek lektorálását és a referált folyóiratok számára végzett referálási tevékenységet nem tekintik publikációnak, így azok már nem kerülnek az adattárba. Nem regisztrálják a konferencia-előadások adatait és a konferenciafüzetekben megjelent kivonatokat sem.

A PPKE BTK adattára két változatban él: az adattár 2002 októberében, a webváltozat fél évvel később, 2003 tavaszán kelt életre. Az adattár lehetővé teszi a több szempontú kereshetőséget, a megfelelő adatszegmentációt és a strukturált rekordszerkezetet, az adatkapcsolatok kialakítását, az automatizált sorfolytonos bibliográfiai tételek listázását a meglévő segédprogrammal, és a statisztikai adatszolgáltatások bázisa is. A webes közzététel külön eljárással készül. A statikus weboldalak szöveges szerkesztett formában mutatják be az oktató publikációs tevékenységét. A jegyzé-

keket a keresőrobotok rendszeresen indexelik és a találati listák elejére sorolják. Az adatbázis kézi lekereséssel és adatbányászattal bizonyos tudomány-metriai jelzőszámokat is számol és tesz közzé.

A Miskolci Egyetem Publikációs Adatbázisa (MEPA)

A *Miskolci Egyetemi Publikációs Adatbázis (MEPA)*¹⁷ 2006 óta építi a Miskolci Egyetem számítóközpontja és könyvtára. Az adatbázis-alkalmazás elsősorban az egyetem oktatói, kutatói számára készült, elsődleges célja összegyűjteni és több szempontból kereshetővé tenni azokat a publikációkat, amelyek az intézmény dolgozóinak nyomtatott vagy elektronikus formában megjelent, tudományos és egyéb jellegű munkái. Szintén az adatbázisból nyerhetők statisztikai adatok az egyetemi beszámolók számára.

A gyűjtőkör meghatározása miatt az adattárban nagyon sokféle dokumentum található. Mivel sem minőségi, sem tartalmi kritériumok nem kellene az állományba kerüléshez, nehéz áttekinteni a szolgáltatás jelentőségét. Véleményünk szerint egy ilyen adattárnak jól körülhatárolt gyűjtőkörének (és céljainak) kell lennie, így lehet egyáltalán teljességre és pontosságra törekedni. A sokféleségből fakadóan igen nagy mélységű feltáráshoz alkalmas rendszert dolgoztak ki a fejlesztők. Nyomtatott és elektronikus dokumentumok, tanulmányok, szabadalmak, beszámolók, multimédiás alkalmazások leírására szolgáló adatmezők is feltüntethetők a rekordokban. Nemcsak bibliográfiai adatokat érhetünk el az összegyűjtött publikációkról, de részben lehetőségünk van teljes szövegek megtekintésére is.

Az adatbázist az *Egyetemi Könyvtár* segítségével az *Egyetemi Számítóközpont* üzemelteti. Az adatbevitelt könyvtárosok végzik az oktatók által átadott publikációs listák alapján, ugyanakkor az oktatók maguk is feldolgozhatják saját listáikat, ha előzőleg regisztrálják magukat. Az adatbázis 2005-ben kezdte meg működését, de visszamenőleg is feldolgozzák az egyetemi oktatók, dolgozók tudományos publikációit.

Az adatbázis a bibliográfiai feldolgozás mellett a teljes szövegek tárolására is alkalmas. Azok az egyetemi oktatók, akik digitális formában is hozzáférhetővé teszik publikációik teljes szövegét itt, írásaikkal automatikusan bekerülnek a *Nemzeti Digitális Adattár (NDA)* adatbázisába, ezzel is nö-

velve munkáik ismertségét, kereshetőségét. Az adatbázisban az egyetemi tudományos élet legfrissebb dokumentumai válhatnak kereshetővé, alkalmat teremtve például az azonos területen dolgozó esetleges kutatási partnerek megtalálására¹⁸.

Az MTA Atommagkutató Intézete Publikációs-Hivatkozási Adatbázisa (APHA)

A Magyar Tudományos Akadémia Atommagkutató Intézetét 1954-ben alapították Debrecenben. Az intézet könyvtárának¹⁹ feladata elsősorban az intézet dolgozóinak szakirodalommal való ellátása és az intézet publikációs tevékenységének, valamint a publikációkra történő független hivatkozások regisztrálása.

Az intézet fejlődésével egyre fontosabbá vált az intézeti munka produktumainak, a publikációknak és a rájuk vonatkozó hivatkozásoknak a nyilvántartása. Ennek szükségessége a múlt század 70-es éveinek végén nyomatékosan vetődött fel, amikor a felügyeleti szerv (az MTA) erőfeszítéseket tett az intézetek munkájának a publikációs tevékenység alapján történő értékelésére.

Az intézeti publikációs-hivatkozási nyilvántartás – amely a 80-as évek eleje óta számítógépesítve van (PC/DOS-on) – megléte különösen nagy szolgálatot tett az akadémiai intézetek 1996–98-as „konszolidációja” idején, amikor főleg az intézetek tudomány-metriai mutatói alapján döntöttek az intézetek további sorsáról.

Jelenleg, számítógép nélkül kilátástalan lenne a nyilvántartási feladat. A teendők egyre bővültek, főleg a rendszerváltozás után megnövekedett számú nemzetközi együttműködés miatt. Szemléltetésül: jelenleg az intézetnek évente megközelítőleg 1000 db írás- és szóbeli publikációja van, amiből a „kemény” írásbeli publikációk száma kb. 300. (Az egyéb, évente rögzítésre kerülő publikációk száma 5-700). Mivel az intézetnek jelenleg nagyjából 100 kutatója van, ez azt jelenti, hogy egy kutatóra évente kb. három „kemény” publikációval lehet számolni. Az intézet elmúlt 50 évének produktumaira évente kb. 2500-3000 független hivatkozást kap. Ezeknek a hivatkozásoknak a begyűjtése külön is hatalmas feladat, annak ellenére, hogy az utóbbi években nagy segítséget nyújtanak a különböző elektronikus, interneten keresztül hozzáférhető adatbázisok. Az adatbázis szerzői

gyűjtőköre elsősorban az intézet alkalmazásában álló kutatóira terjed ki, azzal a bővítéssel, hogy intézeti alkalmazásuk előtti publikációikat és az azokra vonatkozó idézeteiket is bevihetik az adatbázisba, valamint gyűjtik az intézetben készült diplomamunkák adatait. A *Debreceni Egyetemmel* való együttműködés keretében helyet ad a *DE „Fizikai Tudományok” Doktori Iskolája* hallgatói publikációinak és az azokra való hivatkozások adatainak is. Ez a gyakorlat nagyon megkönnyíti a PhD fokozat megszerzése után az intézet alkalmazásába kerülő munkatársak publikációinak és hivatkozásainak későbbi nyilvántartását.

Az adatbázis a publikációk fajtáit tekintve – egyedülállóan a vizsgált adattárak között – helyet ad a gyűjtőkörbe tartozó szerzők kutatási jelentéseinek, előadásainak, évkönyvekben megjelent cikkeinek, és az esetlegesen ezekre történő hivatkozások adatainak is.

Végül néhány kerekített adat az intézeti adatbázis jelenlegi nagyságára: az adatbázisban szereplő szerzők száma 19 000, megjelenési helyek száma

14 000, publikációk száma 22 000, a hivatkozásoké 44 000²⁰.

Az adatbázis a napi használaton kívül arra is szolgál, hogy különböző, itt nem részletezett adatokat szolgáltasson:

- 1) Az intézet vezetésének különböző döntések meghozatalához;
- 2) Az MTA vezetésének az intézet előző évi teljesítményéről (februárban);
- 3) Az MTA TPA-nak részletes publikációs-hivatkozási adatokat²¹ (március-május).

Fontos tudni, hogy az adatbázis az evaulatív tudományometriai eszközökön túl számos, a dinamikus és strukturális tudományometriai vizsgálatokat elősegítő eszközöket is szolgáltasson. Az előbbire a különböző idősorok lekérdezési lehetőségét, az utóbbira az „influence of the coauthors” és a „who is cited by whom” menüpontok szolgálnak például. Lehetőség van az intézeti szerzők különböző mutatók szerinti rangsorának a lekérdezésére is („ranking”).

Táblázatos összehasonlítás

A következő táblázatokban összehasonlítjuk a dolgozatban vizsgált adattárak tulajdonságait. A vizsgálat a 2009. szeptember-október havi időszakban történt.

Teljes név	Magyar Tudományos Akadémia Tudományos Publikációs Adattár	Magyar Tudományos Akadémia Köztudományi Köztudományi Publikációs Adattár	Miskolci Egyetem Publikációs Adattár	Pázmány Péter Katolikus Egyetem Bölcsészettudományi Kar Oktatói Publikációs Adattár	MTA Atomki Publikációs-Hivatkozási Adattár
Használt rövidítés	(MTA TPA, ATPA)	(MTA KPA)	(MEPA)	(PPKE BTK OPA)	(APHA)
URL	http://www.mtatpa.hu/	http://www.mtakoztes.t.hu/kpa.htm	http://publikacio.uni-miskolc.hu/user/	http://www.btk.ppke.hu/cikk.php?cikk=602	http://www.atomki.hu/p2/
Létrehozás dátuma	2001	2003	2005	2002	1986
Nyilvánosság	Publikációs adatok nyilvánosak, hivatkozási adatok jelszóval láthatók	Nyilvános	Nyilvános	Nyilvános	Nyilvános
Kitöltött rekordok száma	Publikáció: ~110 000 Hivatkozás: ~300 000	Publikáció: ~100 000 Hivatkozás: ~200 000	Publikáció ~30 000	Publikáció ~12 500 Hivatkozás ~200	Publikáció: ~22 000 Hivatkozás: ~42 000
Célidőszak	Intézetek 1992– kutatócsoportok 1996–	Életművek	2005– visszamenőleg is gyűjtik	2002– visszamenőleg is gyűjtik	Életművek ~ az Atomki-ból való távozásig

Célcsoport	MTA kutatói	MTA köztestületi tagjai (ajánlott jelleggel)	Az ME oktatói és dolgozói	A PPKE BTK oktatói	Az MTA Atomki kutatói, ill. az Atomki neve alatt publikálók
Célcsoport becsült létszáma	~3000	~14 000	~1500	~250 (eddig 212 szerző)	~500
Gyűjtőkör	Publikációk és <i>hivatkozások</i> Az MTA kutatóhelyek tudományos és ismeretterjesztő közleményei a publikációkra érkezett idézetek, az Akadémia doktora értekezések és az MTA kutatóhelyeken készült PhD disszertációk.	Publikációk és <i>hivatkozások</i> Kiterjed az érintettek olyan publikációira és nyilvánosan hozzáférhető olyan egyéb tudományos teljesítményeire (pl. szabadalmak), valamint a rájuk vonatkozó olyan hivatkozásokra is, amelyeket a nemzetközi publikációs adatbankok nem figyelnek. Például könyvek, a nemzeti tudományok magyar nyelvű publikációi, tankönyvek.	Publikációk <i>Hivatkozás: tervezői</i> Publikáció: az egyetem valamely oktatója vagy dolgozója által készített, szerkesztett vagy létrejöttében jelentős mértékben közreműködött, tudományos tevékenységéhez kapcsolódik, és amely a nagy nyilvánosság számára, tartósan hozzáférhetően, kiadó által nyomtatásban vagy elektronikusan megjelent és elérhető. Disszertációk, szakdolgozatok, kutatási jelentések, előadások nem részei az adatbázisnak.	Publikációk és <i>hivatkozások</i> Publikáció: önálló művek (monográfiák, könyvrészletek, gyűjteményes kötetek tanulmányai, folyóiratcikkek, és más, nem nyomtatott médiában való önálló szereplés, illetve társalkotás, művészi produkció) és a különféle közreműködések során keletkezett műveket (fordítás, szerkesztés, szöveg-gondozás stb.), beleértve a folyóirat-szerkesztést és a könyvsorozat-szerkesztést is.	Publikációk és <i>hivatkozások</i> Reportok, szabadalmak, előadások, annual reportok Szakdolgozatok és disszertációk.
Angol nyelvű felület	részben	részben	nincs	nincs	igen
Feltöltő személye	Kutatók, oktatók, önálló adatbázissal rendelkező intézmények	Kutatók, oktatók	Bejelentés alapján a könyvtár dolgozói, az egyetemi oktatók saját kezűleg	Bejelentés alapján a könyvtár dolgozói	Bejelentés alapján a könyvtár dolgozói
Adatforrások ²²	Másodlagos	Kisebb mértékben elsődleges	Könyvtáros bevitele: másodlagos. Oktatói bevétel: elsődleges.	Publikációk: másodlagos	Publikációk: elsődleges, hivatkozások: másodlagos
Ellenőrzés	Adatok nyilvánvaló ellentmondásmentességét ellenőrzik	Adatok nyilvánvaló ellentmondásmentességét ellenőrzik	Publikációk: elsődleges és másodlagos forrásokból, esetleges	Publikáció: elsődleges forrásból	Publikáció: elsődleges forrásból Hivatkozás: esetlegesen
Globális keresethez ²³	nincs	nincs	nincs	kiváló	kiváló
Lokális keresethez	közepes	kiváló	jó	csak név szerint	kiváló Intézményen belül ennél is részletesebb.
Naprakésztség (a feltöltés 90%-a)	Publikációk adatbevitelhez képest is 1-1,5 év csúszással jelennek meg	Publikációk 1-1,5 év csúszással	Pár hónap	1 év maximum	Publikációknál ~megjelenéstől számított 1 hónap, Hivatkozásoknál ~megjelenéstől számított 1 év.

Szolgáltatások	Éves listák, Szerzők publikációs- hivatkozási listái Tudományometriai jellemzők	Éves listák, Szerzők publikációs- hivatkozási listái Tudományometriai jellemzők	Listák Szerzők publikációs- hivatkozási listái Tanszékek listái	Éves listák, Szerzők publikációs- hivatkozási listái Tanszékek listái Korlátozottan tudományometriai jellemzők	Intézeti éves listák, Szerzők publikációs- hivatkozási listái Téma szerinti listák Listához csatolt tudományometriai jellemzők, Tudományometriai jellemzők idősorai
Tudomány- metriai jellemzők	Számos	Számos	Alig	Néhány	Számos Idősoros is

A táblázat módot nyújt a dolgozatban vizsgált adatbázisok összehasonlítására. Tárgyunk szempontjából a legszembeötlőbb az egyes adatbázisok szerzőkörü és időköri feltöltöttségének a valamikori célokhoz viszonyított alacsony volta. Főleg ennek a jelenségnek az okait fogjuk elemezni a következőkben. A tárgyalt kérdések egy része túlmegy az egyszerű „adattári” problematikán, fontos voltuk akkor válna jelentőssé, ha az adattárak anyaga és/vagy a belőlük származtatott tudományometriai statisztikai adatokat alkalmaznák az egyes szervezetek által végrehajtott értékelésekben.

Az adatfolyamok összehasonlító elemzése

Az adatbázisok használhatósága szempontjából legfontosabbnak tartjuk, hogy az egyes adatbázisok a felvállalt gyűjtési tartomány mekkora részét képesek ténylegesen lefedni. Ennek a tulajdonságnak tagadhatatlanul legfontosabb összetevője az adatbázist tápláló folyamat természete.

A nemzetközi élvonalban lévő publikációs adatbázisokban az adatok forrásai maguk a feldolgozni kívánt művek (folyóiratok, konferenciakiadványok stb.), amelyeket megfelelő tőkeerő birtokában valószínűleg bér munkában dolgoztatnak fel az üzemeletető cégek. Ez elsődleges adatfeltöltést jelent. A hazai adatbázisok részben használják ezt a módszert, csak itt bér munka helyett a különböző eszközökkel rávett oktatók, kutatók, köztestületi tagok végzik az adatbevitelt, részben az eredeti művekre, nagyobb részben előzetesen gyűjtött listákra, és az előbb említett nemzetközi adatbázisokra támaszkodva. Ezeket a módszereket a vizsgált adatbázisok jellemzően keverten használják.

Az ATPA főleg az intézetektől kapott – tehát másodlagos – adatokból dolgozik, bár hivatkozások

tekintetében igénybe veszi a WoK szolgáltatásait is. A KPA eredeti koncepciója szerint a köztestületi tagok adatbevitelére támaszkodik, bár a kezdeti feltöltésre felajánlja a WoK, a Medline, a Scopus és az ATPA-beli adatok átvételét. Ez utóbbi, a KPA-kezelők által végzett átvétel rendkívül időigényes, és jellemzően csak a magasabb akadémiai ranggal bíró köztestületi tagok anyagának felvitelkor működik. A KPA-t megpróbálják a nyomásgyakorlás egyéb eszközeivel is igénybe vetetni (ajánlják használatát a Bolyai-ösztöndíj, az MTA Doktora és az akadémiai címekre való pályázás esetében, továbbá az OTKA és az Országos Doktori Tanács adatbázisa használatakor). A MEPA jellemzően felsőbb utasításokkal próbálja rávenni az oktatókat legalább az utóbbi évek publikációs adatainak feltöltésére. Tapasztalatunk szerint ezekben az esetekben az adatok forrása legtöbbször az oktatók régebről vezetett publikációs/hivatkozási listája. Az egyetemi adatbázisok egy része nem is törekszik a hivatkozások gyűjtésére. Az APHA a publikációs adatokat illetően kizárólag elsődleges forrásból dolgozik, ugyanis az adatbázis kezelői még a kutatók által beküldött adatokat is ellenőrzik.

Az adatfolyamot illetően az alábbi kérdéseket érintjük:

- Nyilvánvaló követelmény az adatbázisokat illetően az azokban tárolt adatok pontossága. Ebből a szempontból legjobb az elsődleges adatforrások használata, ami persze a nyomdahibákat és az ugyanazon szerzők különböző névhasználatát tekintve nem zárja ki az adathibákat. Sajnos, e hibáktól a leginkább „megbízható” adatforrásul használt nemzetközi adatbázisok sem mentesek, mivel azok is a nyomtatásban, elektronikus formában (CD, DVD) vagy a világhálón megjelent művekre támaszkodnak.
- Valamennyi vizsgált adattár egyre jobban támaszkodik az ismert nemzetközi adattárakra (WoK, Medline, Scopus), mintegy hazai infra-

strukturális adottságként véve ezek hozzáférhetőségét. Ezek azonban nem egyszeri beszerzéssel megkapott műszerek, hanem az éves költségvetési helyzet függvényében hozzáférhető lehetőségek. Ennek a problémának a meglétére példa a 2009 elején előállott helyzet, amikor rövid időre (a KPA már konkrét terveket alapított rá) hozzáférhető volt a konferenciaanyagok adatainak feltárását lehetővé tevő WoK-os szolgáltatás. Hasonló kontextusban említhető az SCI CD-s kiadásának az *MTA Könyvtára* által 2007 júniusában történt meglepetésszerű lemondása.

- Az előző pontokhoz szorosan kapcsolódik az adattárak minőségével kapcsolatos azon jellemző, hogy az illető adattár lényegileg tartalmaz-e többszörösen előforduló rekordokat. Ennek eldöntésére az ATPA és a KPA kiterjedten használ számítógépes programokat, azonban a nemzetközi adattárak által nem tartalmazott publikációk/hivatkozások esetében csak a kézi módszer járható. Ennek véghezvitelére az említett adattárak nem rendelkeznek elegendő kapacitással, azt a bejelentő szerzőkre bízzák.
- Mivel a vizsgált adatbázisok általában (az ATPA kivételével, ami túlnyomórészt eleve másodlagos adatokkal dolgozik) az érintettek (oktatók, kutatók, köztestületi tagok) alapoznak, az adatbeviteli motiváltságot nagy részben befolyásolja az adatbeviteli felület bonyolultsága. Ebből a szempontból az APHA a legbarátságosabb, ennek azonban történeti oka és ára van: az egységesen kezelt „Megjelenési hely” mezőt az adatbázisba tényleges bevitelt végző kezelőknek manuálisan kell pontosítani, illetve a más adatbázisok felé irányuló adatexport esetében szétválasztani, finomítani. A sorban következő bonyolultságú ATPA beviteli felület szintén barátságos, az APHA-hoz képest némileg tagoltabb, illetve több információt kérdez. Ez mondható el a MEPA beviteli felületeiről is. A PPKE BTK OPA adatbázisok használatával és a szerzők bevonásával korrigálja az adatokat. A legbonyolultabb adatbeviteli felülete KPA-nak van, amely társulva a vidéki hozzáférés lassúságával, feltehetően sokakat elriaszt a használatától (nem számítva az „elég magas” pozícióban levő személyeket, akiknek módjuk van e munka kiadására, vagy az egyéb módon adatbevitelre szorított oktatókat/kutatókat).

Úgy tűnik, hogy az adatfeltöltés hatékonysága ott a legjobb, ahol az adatbejelentésre kötelezettek szervezeten belül közel állnak az adatbázis kezelőjéhez. Ez a jelenség leginkább az APHA és az ATPA esetében figyelhető meg. Ez utóbbi úgy oldja meg

a közelséget, hogy intézeti szinten közbeiktat olyan egységeket (rendszerint a könyvtárakat), amelyek közvetlenül gyűjtik az adatokat a kutatóktól. Szintén ebbe az irányba mutat a KPA „klónjainak” megjelenése a Budapesti Műszaki és Gazdaságtudományi Egyetemen, a Szegedi Biológiai Központban és az MTA Kísérleti Orvostudományi Kutató Intézetében.

A feltöltöttséget alapvetően meghatározza az adatokat szolgáltató személyek motiváltsága. Ilyen szempontból a KPA helyzete a legreménytelenebb, hiszen még a tudományos hierarchia csúcsán álló szerzők között sem teljes az adattár feltöltöttsége. Az egyes modellek közül – igaz nem nagy szerzőszámot felölelő esetben – az APHA-é mondható a leghatékonyabbnak, ugyanis az adatbázis kezelői tevékeny részt vállalnak a „nem motivált” szerzők publikációinak – régebben hivatkozásainak is – felkutatásában. A motiváltság hiánya egyrészt a már a kívánt fokozatot elért szerzők, másrészt a „termékeny” szerzők sorában érhető tetten. Gondoljuk meg, hogy egy közel 500 tételre kiterjedő publikációs lista esetében mit jelent az „életmű” hivatkozásokkal együtt való karbantartása.

Végül meg kell említeni, hogy a nagy nemzetközi adatbázisokból való tömeges „adatleszívás” és az adatoknak a weben való megjelenítése valószínűleg jogi aggályokat is felvet.

Az „inaktív” szerzők esete

Mint a fenti fejezetben tárgyaltuk, abban a szerzői körben, amelytől a primer adatközlést elvárják, mindig jelentős számmal fordulnak elő kevésbé motivált szerzők, akiktől szinte lehetetlen adatokat „behajtani”. Ugyanebbe a csoportban sorolhatók az elhunytak, illetve a valami okból nem az illető adatbázis illetékességi körébe tartozó munkahelyre távozottak.

Az utóbbi csoport első látásra elhanyagolható lenne, azonban az intézményi érdekek szükségessé teszik ezeknek a szerzőknek a rendszer elhagyása után megjelent publikációinak és/vagy hivatkozásainak nyilvántartását. Erre a jelenségre a publikációk területén jó példák a nagy kollaborációkban született publikációk, amikor is a projekt egyik szakaszában a szerző részt vesz a kutatásokban, amely kutatások eredményeit tartalmazó publikációk csak a szerzőnek az elsődleges anyaintézetétől való távozása után jelennek meg. Az inaktív szerzők problémája még fontosabbá válik a hivat-

kozások esetében, amikor is a hivatkozások sokasága akkor jelenik meg, amikor a bejelentésre kötelezett szerző már nem kötődik a rendszerhez.

Belátható, hogy a szóban forgó szerzők publikációs/hivatkozási adatainak begyűjtése kezelői szinten pótlólagos kapacitások beépítését teszi szükségessé, ha valamennyire is ragaszkodunk az illető adatbázis „teljességének” látszatához. A problémát valamelyest enyhíti, hogy a társszerzői struktúra abban az irányban tolódik el, hogy egyre többen vesznek részt a publikációkban azonos munkahelyről. Bár meg kell jegyezni, hogy a szerzők bejelentési hajlandósága erősen csökken a tőlük „személyileg távolabb eső” adatok bejelentésének elvárása esetén.

Az előbbi bekezdésben vázolt jelenség érzékeltetéséül megjegyezzük, hogy az APHA 2008-as évfolyamában a publikációk adatainak mintegy 20, a hivatkozások adatainak mintegy 70 százalékát nem a publikációk/hivatkozások „intézeti tulajdonosai” gyűjtötték be, vagy legalábbis ebben az esetekben – a publikációk létezésének ismeretében – az adatbázis kezelőinek felszólítására jelentették be.

Az azonos nevű szerzők problémája

A publikációs adattárak feladata nemcsak a publikációk adatainak a gyűjtése, hanem azoknak a gyűjtési célcsoportjaikban található megfelelő társszerzőkhöz való társítása is. Mivel technikai okokból a gyűjteni kívánt adatokat legfeljebb a megjelent művekből lehet beszerezni, nyilvánvaló problémát jelent az azonos nevű szerzőkhöz tartozó publikációk szétszórása a „megfelelő” szerzők között. Hasonló azonosítási probléma jelentkezik a publikációknak az egyes szerzők által használt különböző névváltozatokhoz való hozzárendelésekor. Erre a problémára idáig nem sikerült könnyen járható megoldást találni, bár a Scopus lehetővé teszi az egyedi szerzők biztos azonosítását azok számára, akik vállalják az ezzel járó online adminisztrációt. A legmegbízhatóbbnak az APHA kialakított gyakorlata tűnik, amelyben az adatbázis kezelői az adatbázisba való feltöltés előtt megpróbálják azonosítani az azonos nevű társszerzőket, akiket a megfelelő kódszámmal kódolnak²⁴. Az ATPA régebben a születési évvel próbálta azonosítani a szerzőket, jelenleg – nyilván a KPA igényeinek megfelelően – a köztestületi azonosító megadásával kísérletezik.

A KPA a szerzőkre bízta ennek a problémának az eldöntését, egy viszonylag munkaigényes folyamaton keresztül. Ennek a lényege az, hogy a nyilvánosan látható webes felületre való közzététel előtt minden publikáció és idézet kap egy „kvázitulajdonos”-t, aki a továbbiakban a szóban forgó publikáció és az ahhoz tartozó hivatkozások adatait szerkesztheti. Az újonnan jelentkező résztvevő a már bent levő publikációkat bejelölheti: az adatbázison kívüli azonos nevű társszerzőkkel nem törődik.

Mindegyik felsorolt metódus igénybe veszi a célcsoport tagjainak ismeretanyagát – talán az APHA kíméli meg legjobban a szerzőket ettől a feladattól.

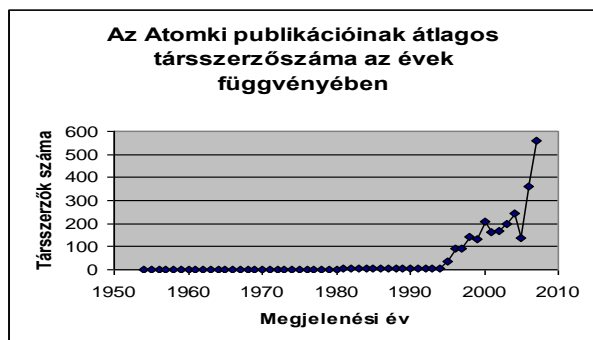
Sok társszerzős publikációkkal/hivatkozásokkal kapcsolatos problémák

A XIX. század végéig a tudománytörténet alig ismer több társszerzős tudományos művet. A legújabb kor kutatásainak jellemzője a sok résztvevős, főleg nemzetközi együttműködések megjelenése, amelyeket legtöbbször a kitűzött kutatási feladat nagysága és/vagy bonyolult volta tesz indokolttá. Ezekhez az együttműködésekhez természetesen társul a sok társszerzős publikáció megjelenése. A nagy létszámú együttműködések először a fizikában jelentek meg, de mára a biológiában, az orvostudományban és a csillagászatban is elterjedtek. Az APHA gyakorlatából kitűnik, hogy a társszerzőszám több esetben eléri a néhány ezres nagyságrendet. Kisebb mértékben más természettudományokban, de a társadalomtudományok területén is növekedésnek indult az átlagos társszerzőszám. Külön problémát jelent, sok esetben (már a néhány tízes szerzőszámnál is) ennek a számnak a megállapítása és/vagy feltüntetése a publikációk adatainál. Számos publikációban nem írják ki a teljes szerzőnévsort, a szerzők listájának leírása „XY for the/on behalf of the N Collaboration” alakú. Az együttműködés szerzőlistáját, ha egyáltalán kinyomozható, rendszerint egy másik publikáció tartalmazza (vagy az utóbbi évek szokása szerint az illető kötet végén szereplő lista). A probléma sajátosan érzékeny területet érint; a szóban forgó publikációk a tudomány frontvonáiban lévő témákat érintenek, ezért hivatkozásgeneráló képességük különösen nagy.

A sok társszerzős publikációk problémájára az APHA próbálkozott először megoldással: csak 30 szerzőt (köztük az első és az intézeti szerzőket) tüntetnek fel, a társszerzők számát (amelyet vagy

leszámolnak, vagy becsülnek) külön mezőben rögzítik. Ennek a megoldásnak az előnye, hogy a később szükségessé váló tudományometriai statisztikai adatok számításánál felhasználható. Az ATPA az APHA megoldását vette át, azzal a módosítással, hogy a szerzőszámot az utolsó szerző után tárolják. A KPA ideiglenes megoldásként a „Megjegyzések” mezőben fogadja be a társszerzők számát. Az egyetemi adattárak (PPKE BTK OPA, MEPA) megelégszenek az „et al.” jelölés használatával. A következőkben néhány, a sok társszerzős publikációk megjelenésével kapcsolatos problémát veszünk sorra:

- A hazai adattárak általában a „független” idézetek gyűjtését célozzák meg. Ezek olyan hivatkozások, amelyeknél a hivatkozó és hivatkozott mű szerzőinek halmaza egymástól elkülönül (diszjunkt). A függetlenség kérdése kulcsfontosságúnak számít az egyes kutatói és intézményi teljesítményértékelési eljárásokban. A függetlenség eldöntését az adattárak általában a hivatkozott publikáció szerzőire bízják. Egyes adattárak (ATPA és KPA) komoly erőfeszítéseket tettek a függetlenség gépi detektálására, a WoK-ból megszerezve a hivatkozott és hivatkozó művek teljes szerzőnévsorát. Természetesen ez sem teszi lehetővé a hivatkozások függetlenségének megnyugtató megállapítását, mivel az esetlegesen többes társszerzőgárdában a társszerzőszámmal együtt növekvő valószínűséggel fordulnak elő azonos nevű (főleg távol-keleti) szerzők.
- A sok társszerző neve rögzítésének néhol még technikai problémái is lehetnek.
- Speciális kérdés a sok társszerzős publikációkkal kapcsolatban az, hogy a publikációt végül is ki „számolja” el magának.
- A probléma akut voltát az APHA-ból nyert adatokat ábrázoló 2. ábrán szemléltetjük:



2. ábra Az Atomki azon publikációinak átlagos társszerzőszáma – a megjelenési évek függvényében –, amelyek fejezetében szerepel az Atomki neve

ISI/nem ISI, SCI/nem SCI

A tudományos kutatókra ráömlő rengeteg publikáció a tudományometriai értékelő eljárásokat is tartalmazó feladat elé állította. Célszerűnek látszott a megjelent publikációkat szűrni a megjelenési hely „kiválósága” szerint (hazai/nemzetközi, magyar/idegen nyelvű, referált/nem referált, impaktfaktoros/nem impaktfaktoros stb.). Ennek a szűrésnek egyik formája azon alapul, hogy az ISI-féle citációs indexek tartalmazzák-e az illető publikációt. Ezt a megkülönböztetést alkalmazza az ATPA és a KPA, ahol az ISI/nem ISI adat külön mezőt foglal el. Meg kell jegyeznünk, hogy jelenleg e két adattár bajban van, mivel az ISI-SCI utóda, a WoK elkezdte gyűjteni a konferenciakiadványokat is visszamenőleg. Ettől némileg eltér az APHA gyakorlata, amely az SCI megjelölést használja, továbbá – technikai okokból – megkülönböztetést tesz publikáció és hivatkozás között²⁵.

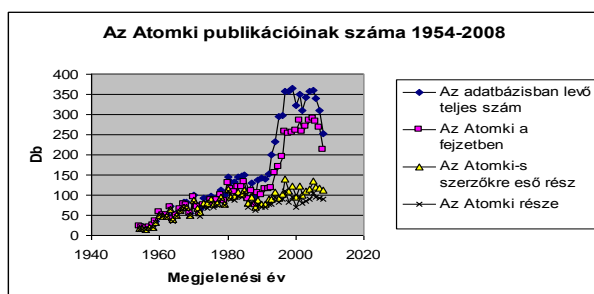
A teljesítmények additivitásának kérdése – egyéni/intézményi teljesítmény elkülönítése

Több társszerzős szerzőlista, vagy a publikáció fejezetében levő több intézmény esetén felvetődik a kérdés, hogy valójában a publikáció mekkora hányada tulajdonítható az illető egyes szerzőnek vagy intézménynek és/vagy országnak. Ha más támpont nincs – mint a szabadalmak esetén –, célszerű úgy felosztani a szóban forgó „teljesítmény”-t (publikációt, és a hozzá tartozó hivatkozásokat), hogy a különböző szerzőkhöz, intézményekhez tartozó hányadok összege végül is egy publikációt, illetve hivatkozás esetén a hozzá tartozó hivatkozások teljes számát eredményezze. (Az evaluatív tudománymetria alapvető statisztikai egységei a publikáció és az idézet^{26,27}). A szóban forgó felosztás történhet egyenletesen, vagy nem egyenletesen. Ez utóbbira példa a szabadalmaknál alkalmazott százalékos felosztás. Ilyen lehetőséget ad a KPA, azonban ott ezek a mezők publikációkra nézve általában nincsenek kitöltve. Valószínűleg ez a megoldás lenne kielégítő az egyéni teljesítmények megítélésakor is. Jelenleg ez a kérdés nagyon függ a szakterülettől. Van ahol az első, de van ahol az utolsó szerző „számolja” el a publikációt. Sok területen társszerzők nevének feltüntetésében az ábécésorrend a szokásos. Véleményünk szerint a több publikációt tartalmazó lista esetén az „érdem” kiátlagolódik, ugyanis a több résztvevő együttműködések nem tűrik meg a „lázslást”, ha valaki rajta van a szerzőlistán, az valamikor azt

kiérdemelte. Az egyéni szinttől eltávolodva, nagyszámú publikáció esetén alkalmazható az „érdek” egyenletes elosztása. Ennek előnye egyszerűségében rejlik, elég csak a társszerzők számát, illetve az illető intézménynek az előfordulási hányadát rögzíteni. Például, ha n a társszerzők száma, és egy intézeti szerző van, aki a tekintett intézeten kívül még m intézmény neve alatt jegyzi a cikket, akkor az egyéni saját részarány²⁸ $1/n$, míg az intézetének a részaránya $1/(n*(m+1))$. Az itt vázolt megoldást alkalmazza, és viszi végig az APHA.

Az APHA nyújtotta lehetőséget kihasználva megvizsgáltuk, hogy az előbb említett mechanizmus milyen szerepet játszik az intézeti publikációk számának alakításában.

A 3. ábrán négy adatsort láthatunk. A legfelső az adatbázisban az intézeti szerzők által bejelentett összes publikáció számát mutatja. A következő adatsor azoknak a publikációknak a számát ábrázolja, amelyekben a fejezetben megjelent az Atomki neve. A „permanent address...”, és az „on leave from...” lábjegyzetbeli kifejezéseknek ebből a szempontból csak informatív értéke van, ha az intézet neve nem szerepel a fejezetben, az azt jelenti, hogy az intézetek közti „erőviszonyok” ezt nem tették lehetővé. Az ATPA azokat a publikációkat is intézetinek tekinti, amelyeknél az intézet a lábjegyzetben szerepel, feltéve, hogy az illető szerző benne van az éves statisztikai létszámban – véleményünk szerint ez a felfogás túlságosan „rugalmas”. A harmadik görbe az intézeti szerzőkre eső, az intézetre összegzett publikációszámot mutatja. A negyedik adatsor az intézetre eső publikációk számát ábrázolja. Jól látható az egyes esetek közti különbség.



3. ábra Az Atomki publikációinak és azok saját részének változása az évek függvényében

Az irodalomban fellelhető országok és intézetek közti összevetések²⁹ általában a WoK adataira

alapozódnak, amelyben minden publikáció mellett megtalálható az az információ, hogy a cikk létrehozásában milyen intézetek kerültek a fejezetbe (attól függetlenül, hogy az illető intézetből hányan voltak a szerzők, vagy eredetileg honnan jöttek). A statisztikák készítésekor minden előforduló országnak, vagy intézetnek a teljes publikációt és/vagy összes hivatkozást elszámolják („full credit” módszer). Így egyes országok és/vagy intézetek a valóságos helyzetnél jóval kedvezőbb színben tűnhetnek fel, amint az már más szempontból is felvetődött³⁰. Természetesen a fentiek csak akkor igazak, ha mint az evaluatív tudománymetria általában teszi, a publikációkat és hivatkozásokat azonos statisztikai értékűnek fogadjuk el.

Publikációs adattár és/vagy repozitórium

Az informatika fejlődésével lehetővé vált, hogy a publikációk adatainak rögzítése mellett a publikáció teljes szövegét is elérhetővé tegyék a publikációs adattárakban, és ezeken keresztül az interneten. A tudományos folyóiratok szerzői jogi politikái³¹ azonban ezt nem mindig teszik lehetővé. A vizsgált egyetemi adattárak különálló elektronikus archívumot (repozitóriumot) tartalmaznak, amelyek működtetése szerzői jogi szempontból erősen aggályos. Az APHA a legtöbb lehetőség (html, pdf, link, DOI) használatát lehetővé teszi, azzal a kikötéssel, hogy a bejelentőnek a szerzői jogi feltételek ismeretében el kell döntenie, hogy a teljes szövegű változatot a teljes vagy intézetkörű nyilvánosság számára teszi-e láthatóvá. Meg kívánjuk jegyezni, hogy a feltöltöttség az APHA kivételével minden vizsgált adattár esetében minimális. A „feltöltésre ítélték” motivációjára vonatkozó fejtegetéseink értelmében a közeljövőben nem is várható pozitív változás ezen a téren.

Minőségbiztosítási kérdések

A megvalósított adattárak minőségéhez hozzájárul az egyes adattárak működtetői által követett minőségbiztosítási gyakorlat. A legalaposabbnak ebből a szempontból a KPA tűnik a honlapján vázolt eljárás alapján. E szerint az adattár kezelői ellenőrzik szintaktikai és egyéb szempontokból a szerzők által bevitt adatokat. A teljes nyilvánosságra a szerzők és kezelők által validált adatok kerülhetnek. Az ATPA hasonló eljárást követ. Úgy érezzük, hogy bár az előbbi eljárások időigényesek (a KPA a validációt 2-8 hetes határidőre vállalja, amit a gyakorlat szerint ritkán sikerül tartani), azonban

mégis csak a nagy nemzetközi adattárak (WoK, PubMed, Scopus) adataival való összevetésre szorítkoznak.

A jövőben a minőségbiztosítási kérdésekre még fokozottabb hangsúlyt kellene helyezni. A továbbfejlesztés érdekében lehetőséget kellene adni a felhasználóktól a fejlesztőkig/kezelőig, sőt a fenntartókig irányuló visszacsatolásra (*feedback*). Valamilyen módon képet kellene kapni az illető adattár iránti érdeklődés feltérképezésére. Ennek első lépése lehetne az APHA-ban található látogatottsági számlálók szélesebb körű alkalmazása.

Az adattárak működtetésének gazdasági háttere

Az egyes adattárak működtetésének gazdasági hátterére nehéz adatokat találni. Egy biztos, hogy a nagy nemzetközi adattárakkal szemben – amelyeket magánvállalkozások hoztak létre, és magántőkéből fejlesztenek – a hazai adattárak kivétel nélkül állami költségvetési pénzre támaszkodnak. Kérdéses, hogy hazánkban lehetne-e profitalapon létrehozni publikációs adattárat. Jelenlegi tudomásunk szerint a piacon nincs hozzáférhető, a fenti igényeket kielégítő, az egyéni és intézményi publikációkat és hivatkozásokat a pontosság igényével kezelni kész szoftver.

Ebből a tényből azonban a hazai publikációs adattárak egy sajátos tulajdonsága eredeztethető: míg a fejlesztési célok kitűzésénél az összefogni kívánt „aktorok” kívánságainak „legkisebb közös többszöröse” jut érvényre, a későbbi működtetést illetően az egyes helyen meglévő erőforrások „legnagyobb közös osztója” szab határt – amely erőforrásoknak a szükségleteket kielégítő felmérésére, illetve átcsoportosítására nagyon ritkán kerül sor. A mondat első felének igazságára példa, hogy a WoK állapota és fejlesztési irányai a kezdeti fejlesztés időszakától eltekintve, elsősorban a felhasználók (előfizetők) érdekeit tükrözik, a hazai adattárak fejlesztési irányai a fenntartók sajátos szempontjaiból táplálóknak. A szóban forgó mondat második felének igazságát pedig a működtetés során fellépő anomáliák igazolják.

A költségvetési pénzek korlátozott voltából ered az is, hogy az adattárak nagy célokat tűznek ki – amelyekhez szükséges eszközök fejlesztéséhez még, úgy-ahogy meg vannak az eszközök – a legmunkaigényesebb munkafolyamatot, a feltöltést

már az érintett szerzőkre bízzák rá. Ennek a körülménynek az eredménye az, hogy bizonyos felfutási periódus után több adattár közönybe fulladt.

A hazánkban a felsőoktatási-tudományos szférában megvalósuló szoftverfejlesztések (általában adatbázisok) működésében felmerülő zavarok okai:

- a forrásokhoz képest túl ambiciózus célok kitűzése,
- a fejlesztés során szükségessé váló döntések nem megfelelő időben és előkészítettséggel való meghozatala,
- a nem megfelelő szintű rendszertervezés,
- a tesztidőszak rövid időre való korlátozása, a tesztek nem elég széles körben való végrehajtása, a felhasználói visszacsatolások figyelmen kívül hagyása,
- az a hit, hogy a nemegyszer kötetekre rúgó használati utasítást elegenden elolvassák.

A fenntartók általában csak koncepcionális szempontok megadására szorítkoznak, az adatbázis megvalósításának formája, annak kezelhetősége a fejlesztők (informatikusok) és a kezelők (általában könyvtárosok) közötti erőviszonyok függvénye. Erre példa, hogy míg az egyetemi működtetésű adattárak a publikációk adatainak leírásakor a kéznél lévő integrált könyvtári rendszer (Corvina) szabványát preferálják, az informatikusoknak jobban kiszolgáltató ATPA és KPA ezeket a kérdéseket jóval lazábban kezeli. Szerencsésnek mondható az az eset, amelyben sikerül a munkafolyamat nagy részét egyetlen személynek átfogni (APHA).

Az egyes adattárak működésében mutatkozó zavarok magyarázata a szűkös források miatti korlátozott tervezésben keresendő. A szűkös források sokszor abból adódnak, hogy a projekt előterjesztői – a rendszerváltást megelőző időszakra jellemzően – alulbecsülik a megvalósítás költségeit. Ez oda vezet, hogy a matematikai/szervezeti modell realizálása folyamán nem veszik figyelembe a fogadó közeg szervezeti-képességi korlátait. Ráadásul a szervezeti feltételek tervszerű, előzetes módosítására alig-alig van példa. Ily módon az olyan-amilyen produktumot rákényszerítik a működtető szervezetre – személyekre –, egyfajta „viviszekcióra” kényszerítve az(oka)t. A működési anomáliák kiküszöbölése jobb esetben a realizált adattár szoftveres kiegészítését, rosszabb esetben plusz emberi erőforrás hozzárendelését teszi szükségessé.

A vizsgált publikációs adattárak működtetési költségeinél általában nem veszik figyelembe az adattárak által egyre intenzívebben használt nemzetközi publikációs adattárak előfizetési (bérleti) költségét. Kérdéses, hogy mindent összevéve, hogyan aránylik annak a döntésnek a költsége, hogy „ki az érdemesebb kutató, intézmény stb.”, ahhoz az összeghez, amely aztán a „valódi” kutatás céljaira marad.

A vizsgált adattárak működtetéséből – egy kivétellel – hiányzik a működtetés hatékonyságának mérésére szolgáló eszközök használata. Egyedül az APHA vezette be a leggyakrabban lekérdezett listák megtekintési gyakoriságának számlálását. Ez a kérdés minden vizsgált esetben nagyon fontos lenne a fejlesztés további irányainak kitűzéséhez. Az is lehet, hogy kiderülne, hogy a relatíve nagyszámú hazai publikációs adattár fejlesztését a valós igények nem indokolják.

Következtetések és javaslatok

A publikációs adattárak jövőbeli fejlődésének/fejlesztésének kétség kívül legsarkalatosabb pontja a hazai tudománypolitikának a tudományos teljesítmény értékelésével kapcsolatos álláspontja. Fontos lenne, hogy a tudománypolitika döntse el, hogy az evaluatív tudományometriát a hosszú távú tudománypolitikai döntések segítőjeként vagy az egyes tudománybeli érdekcsoportok taktikai bűnböjtjeként használja. Ez, egyrészt a fenntartókra, másrészt a felhasználókra gyakorol nyomást.

A fentiekől függetlenül néhány javaslatot már most is megfogalmazhatunk:

1. Célszerű lenne a különböző adattárak közötti kapcsolatok elmélyítése:
 - Az eltérő helyen gyűjtött adatok összeférhetőségének javítása, vagy ha az nem kivitelezhető, az egyes publikációs adattárak közötti adatimport/-export lehetőségének megteremtése. Ennek egy lehetséges módja lehetne a RIS formátumszabvány³² használata.
 - Egységesíteni kellene az egyes adattárak által szolgáltatott tudományometriai statisztikai jellemzőket.
 - Célszerű lenne az intézményekre és azok szervezeti egységeire vonatkozó statisztikai idősorok bemutatása.
 - Célszerű lenne a rendelkezésre álló forrásokat koncentrálni a hatékonyabb fejlesztés érdekében. Annál is inkább, mivel például az

MTA kutatói nem egyszer három oldalról vannak adatfeltöltésre kényszerítve.

2. Célszerű lenne eltörölni bizonyos gyűjtőkori korlátokat (Egyrészt az ATPA-ban lehetővé kellene tenni az 1992 előtti anyagok keresését, megnyitva ezzel az utat az akadémiai intézetek publikációs története előtt, másrészt a KPA-ban a fokozattal még nem rendelkező fiatal tudósok számára is nyitottnak kellene lenni.) Belátható, hogy ezeknek a korlátoknak a tágitása nem jelentene jelentős kapacitásnövelési igényt.
3. Meg kellene találni a publikációs adattárak szofisztikált szerkezete/kezelhetősége és a felhasználók motiváltsága (az adatbázisok feltöltöttsége) közötti középutat. Jelenleg bonyolult szerkezetű adatbázis (KPA) működik alacsony feltöltöttséggel, és aránylag gyatra platformon működő adatbázis (APHA) ér el jó feltöltöttségi eredményeket. Növelni kell a megcélzott szerzőcsoport motiváltságát és/vagy bővíteni kell azt a kapacitást, amely az adatok előkészítését/bevitelét végzi. Ennek egyik eszköze lehet, hogy az adattár és az adatszolgáltatásra végül is kötelezettek közé gyűjtőszinteket telepítenek. Az „inaktív” szerzők munkájának kiváltására kezelői szinten pótlólagos kapacitásokat kell beépíteni.
4. Az előbbi pontban említett hierarchikus szervezeti felépítésnek szükséges eleme lenne, hogy a különböző szinteken lényegileg azonos programok (klónok) működjenek, lehetővé téve az egymás közti kommunikáció egyszerű módját.
5. Egyértelműen javítani kell az adattárak globális kereshetőségét. Ennek eszköze lehet az az APHA-ban alkalmazott megoldás, hogy legalább az egyéni publikációs listákat – az alkalmazott dinamikus weblap megoldás helyett – időnként statikus, Dublin Core³³ metaadatokkal jól ellátott weblapként újra kell generálni. A globális kereshetőség speciális kérdése az idegen nyelven való kereshetőség, ami fejlesztésre szorul. Ez vonatkozik az APHA kivételével mindegyik vizsgált adattárra.
6. Meg kellene oldani néhány *minőségbiztosítási* kérdést, lehetőséget kellene nyújtani a felhasználóktól a fejlesztőkig/kezelőkig, sőt a fenntartókig irányuló visszacsatolásra. Valamilyen módon meg kellene oldani az illető adattár iránti érdeklődés feltérképezését. Ugyancsak minőségbiztosítási kérdés a feltöltött adatok megbízhatósága, ami különösen az idézetek esetében fontos egy esetleges értékelés szempontjából. Erre a problémára megoldás lenne az egyéni/intézményi publikációs/hivatkozási listák *mintavétellel* történő minősítése.

7. Célszerű lenne az előző részben ismertetett problémák konszenzuson alapuló tisztázása, annak eldöntése egyáltalán, hogy a jövőbeli teljesítményértékelési szempontok kiterjednek-e ezekre a tényezőkre. Különösen az egyetemek területén kellene ezzel kapcsolatban határozott döntéseket hozni. Nem véletlen, hogy a publikációs adattári fejlesztésekben a BME jár elől, ahol valószínűleg van mit szerbe-számba venni.
8. Törekedni kell a matematikai modell (beleértve annak realizált formáját is) és a szervezeti modell (a végül is rendelkezésre álló személyi feltételeket is beleértve) összhangjának már a projekt kezdeti szakaszában való megteremtésére.

Az egyes adattárak jövőbeni tervei a fenntartó szerv megfogalmazott elvárásain túl, általában a fejlesztők és kezelők fejében alakulnak ki az érdekeknek megfelelően. Sajnálatos, hogy a felhasználók igényei csak minimális mértékben kapnak teret.

Összefoglalás

Dolgozatunk néhány hazai fejlesztésű tudományos publikációs adattár összehasonlító elemzésével foglalkozik. Az elemzés során megállapítottuk, hogy az adattárak működésének színvonala jelentősen elmarad az ismert nemzetközi adattárakétól. Úgy gondoljuk, hogy az anomális működésnek a szűkös költségvetési forrásokon túl két, a jelenlegi viszonyok között is finomítható pontja van: egyrészt tisztázatlan az adattárak esetleges tudományometriai értékelésre való felhasználása, másrészt átgondolatlan az adattárak működtetésének szervezeti kérdése, főleg ami az adatszolgáltatók motiváltságát illeti.

A dolgozat anyaggyűjtési időszakának lezárulta (2009. 06. 30.) után értesültünk arról, hogy a magyar tudományos élet vezető szervezetei, a Magyar Tudományos Akadémia (MTA), a Magyar Akkreditációs Bizottság (MAB), az Országos Tudományos Kutatási Alapprogramok (OTKA), a Rektori Konferencia, valamint az Országos Doktori Tanács (ODT) 2010. január 1-jétől létre kívánják hozni a *Magyar Tudományos Művek Tárát (MTMT)*, amely távolról egyesíti a publikációk és a hivatkozások nyilvántartására vonatkozó hazai igények kiszolgálását³⁴.

Csak remélni lehet, hogy ennek az új adatbázisnak a létrehozásakor hasznosulnak a jelen dolgozat megállapításai.

*

Időközben a KPA adattárát lezárták, a fenti tanulmány végén említett MTMT lépett a helyébe. Az MTMT honlapja 2010. február 3-án indult. (A szerk.)³⁵

Jegyzetek és hivatkozások

- ¹ APHA - <http://www.atomki.hu/p2/>
- ² DRÓTOS L. – KŐRÖS K. – SOMOGYI T.: Adattár-mustra. Publikációs adatbázisok. = TMT, 55. köt. 3. sz. 2008. p. 127–145. Ez a kritikai elemzés a Miskolci Egyetem publikációs adattárát (MEPA), a Szegedi Tudományegyetem publikációs adattárát (SZTPA) és az MTA Köztestületi Publikációs Adattárát (KPA) dolgozza fel főleg programozástechnikai szempontokból. A mi dolgozatunk a kereshetőség és megjelenési felület helyett a publikációs adattárak használhatóságát, üzemeltethetőségét és rendszerszervezését vizsgálja.
- ³ VINKLER P.: Tudományometriai kutatások Magyarországon. = Magyar Tudomány, 169. köt. 11. sz. 2008. p. 1372–1380.
- ⁴ TOLNAI M. – SCHUBERT A. – WOLF Gy.: Tudományunk mérhető teljesítménye az Essential Science Indicators mutatószámainak tükrében. = Magyar Tudomány, 169. köt. 8. sz. 2008. p. 989–997.
- ⁵ GARFIELD E. – SHER I. H.: New Factors in Evaluation of Scientific Literature through Citation Indexing. = American Documentation, 14. köt. 3. sz. 1963. p. 195–201.
- ⁶ Impaktfaktor/hatástényező: Egy folyóirat n -dik évi impaktfaktora egyenlő az illető folyóirat $n-1$ -dik és $n-2$ -dik évfolyamában megjelent cikkekre való összes (minden, az adatbázisban levő) tárgyévi hivatkozások száma osztva az ezekben az években az illető folyóiratban megjelent összes cikk számával. Az impaktfaktor arra ad becslést, hogy egy folyóiratban megjelent átlagos publikációra a megjelenést követő két év átlagában hány, az adatbázisbeli hivatkozás várható.
- ⁷ ZÁDORI Zs.: Mit ér a tudós, ha magyar. Hvg.hu – hírek szünet nélkül, 2006.05.17. <http://hvg.hu/itthon/20060516tudosmagyarmta> ; Letöltés dátuma: 2009. december 1. és ZÁDORI Zs.: Hódít a teljesítménymérés: számnak ingere. Heti világgazdaság, 2006.10.14. p. 68–71.
- ⁸ Idézetségi félidő: Az idézetségi félidő a tárgyévtől visszafelé számított annyi év, amennyiben megjelent cikkeire a folyóirat a tárgyévben kapott összes idézeteinek felét kapta. Idézési félidő: Az idézési félidő a tárgyévtől visszafelé számított annyi év, amennyiben megjelent cikkekre a folyóirat tárgyévi hivatkozásainak fele vonatkozott.
„Frissességi” (immediacy) index: Egy folyóirat frissességi indexe azt méri, hogy milyen gyorsan idézik a folyóirat egy „átlagos cikkét”. A frissességi index

- megmutatja, hogy milyen gyakran idézik a folyóirat cikkeit a megjelenéssel azonos évben. A frissességi indexet úgy számítjuk ki, hogy a folyóirat tárgyévben publikált cikkeire a tárgyévben kapott idézetek számát elosztjuk a folyóiratban ugyanezen évben publikált cikkek számával.
- ⁹ Aktivitási index: Egy adott időszakban, egy adott szakterületen, egy adott szerzőhalmaz által publikált cikkek száma a világban publikált cikkek számához képest. Az ~et például lehet tekinteni a Magyarországon dolgozó szerzőkre stb. nézve
 Produktivitás: Valamilyen szempontból definiált saját rész osztva az elkövetéshez felhasznált évek számával. Ennek használatban van az egy szerzőre eső része is.
- ¹⁰ Hirsch-index: Egy szerző Hirsch-indexe az a h szám, amelyre vonatkozólag a szerzőnek van legalább h darab publikációja, amelyekre legalább h darab hivatkozás történt.
- ¹¹ Szakterületi statisztikák magyar kutatók Hirsch-féle h -indexeire az MTA Köztudományi Publikációs Adattárának adatai alapján. MTA-KSzl, 2005 november; Heti válasz: A száz legidézettebb természetkutató. = 8. köt. 2008. 01. 17. p. 12–14.; Heti válasz: Tudósok a mérlegen. = 8. köt. 2008. 02. 21. p. 22–23. és
 TOLNAI M.: Tudósaink mérhető teljesítménye az MTA Köztudományi Publikációs Adattár adatainak tükrében. = Magyar Tudomány, 169. köt. 8. sz. 2008. p. 976–988.
- ¹² SCHUBERT A.: Tudománykommunikáció képzés. Előadás. Budapest, Magyar Tudományos Akadémia, 2007. okt. 3.
- ¹³ RIES, A. – TROUT, J.: A marketing huszoneköt éves törvénye. Budapest, Bagolyvár Kiadó, 1995.
- ¹⁴ ZOLNAI L. – GÁCSI Z.: Mérünk, de mit? Egy formula margójára. = Magyar Tudomány, 43. (105.) köt. 8. sz. 1998. p. 988–993. Ebben a tanulmányban olvasható az értékelésnél alkalmazott módszer leírása.
- ¹⁵ Amint az adattár ismertetőjéből kitűnik: „Az összegyűjtött adatok hozzájárulnak a szakirodalmi információs folyamatok törvényszerűségeinek feltáráshoz is.”
- ¹⁶ DUDÁS A.: Tudományos önismeret: az egyetemi bibliográfiák szükségességéről és egy bölcsészettudományi oktatói publikációs adatbázisról. = TMT, 55. köt. 3. sz. 2008. p. 107–125.
- ¹⁷ KISS A. – VITÉZ G.: Miskolci Egyetemi Publikációs Adatbázis. = TMT, 53. köt. 7–8. sz. 2006. p. 321–327.
- ¹⁸ DRÓTOS L. – KŐRÖS K. – SOMOGYI T.: Adattár-mustra. Publikációs adatbázisok. = TMT, 55. köt. 3. sz. 2008. p. 127–145.
- ¹⁹ MTA Atomki Könyvtára –
<http://www.atomki.hu/library/konyvtar.htm>
- ²⁰ APHA – <http://www.atomki.hu/p2/>
- ²¹ ATPA – <http://www.mtatpa.hu/>
- ²² Elsődleges/másodlagos adatforrás: Az adatforrás elsődleges, ha a publikációs adattár feltöltése közvetlenül a megjelent művek alapján történik. Ha nem így van, másodlagos adatforrásról beszélünk.
- ²³ Globális keresetőség: Ha egy weblap tartalma feltehető a világhálón létező keresőmotorok számára. Ehhez általában az szükséges, hogy az illető weblap el legyen látva metaadatokkal (l. Dublin Core).
- ²⁴ A címek és a megjelenési helyek is kapnak kódszámot, ami több feladat egyszerű megoldását – például a folyóiratok impaktfaktorral való ellátását – megkönnyíti.
- ²⁵ SCI Publikáció: Ha a megjelenés évében a megjelenési helynek van impaktfaktora (OTKA definíció); SCI Hivatkozás: Ha az illető megjelenési helynek valaha volt impaktfaktora.
- ²⁶ ZOLNAI L. – BERÉNYI D.: Kutatóintézeti tudomány-metria – ATOMKI 1954–1989. = Fizikai Szemle, 39. köt. 8. sz. 1989. p. 285–291.
- ²⁷ ZOLNAI L.: A „sötét teljesítmény” nyomában. = Magyar Tudomány, 169. köt. 7. sz. 2008. p. 870–874.
- ²⁸ Idegen rész: A saját résznek a felőlt publikáció, illetve hivatkozási számból való kivonásával adódik. Saját rész: Egy szerző a publikációnak, vagy az arra való meghatározott körű hivatkozások számának a társszerzők számával osztott része. Aszerint, hogy az így kapott mennyiségeket a szerzők milyen körére összegezzük fel, beszélhetünk intézeti, magyar stb. Idegen részről.
- ²⁹ BENCZE Gy.: Helyünk Európában. = Fizikai Szemle, 45. köt. 1. sz. 1995. p. 18–20.
 ; BRAUN T. – GLÄNZEL, W. – NÉMETHNÉ KOVÁCS É. – PERESZTEGINÉ SZABADI Zs.: Magyarország helyzete a természettudományi alap kutatás világában – tudomány-metriai tájkép a második évezred végén. = Magyar Tudomány, 47. (108.) köt. 7. sz. 2002. p. 935–945.
 ; TOLNAI M. – SCHUBERT A. – WOLF Gy.: Tudományunk mérhető teljesítménye az Essential Science Indicators mutatószámainak tükrében. = Magyar Tudomány, 169. köt. 8. sz. 2008. p. 989–997. és
 MARTON J.: Tizenkét futam az európai tudomány-metriai mezőnyben. Élettudományok. = TMT, 56. köt. 3. sz. 2009. p. 118–122.
- ³⁰ TÖRÖK Á.: Reális-e a magyar tudomány 20. helye a (képzelt) világranglistán? = Magyar Tudomány, 45. (108.) köt. 11. sz. 2000. p. 1307–1328.
- ³¹ Sherpa – <http://www.sherpa.ac.uk/romeo.php>
- ³² RIS formátum: A Thomson-Reuters cég Reference Manager nevű bibliográfiakezelő szoftvere által használt, a publikációk adatai import/export-jára szolgáló formátum szabvány.

³³ Dublin Core: A weblapok tartalmának leírására szolgáló metaadatok szabványa. A magyar gyakorlatban az MSZ ISO 15836 vezette be.

³⁴ Magyar Tudományos Művek Tára létesítése: http://mta.hu/index.php?id=634&no_cache=1&backPid=390&tt_news=11266&cHash=873f77c3c7

³⁵ MTMT – <http://mtmt.hu>



Adománné Zolnai Dóra

az MTA Atommagkutató Intézete
könyvtárának könyvtárosa.
E-mail: dzolnai@atomki.hu

Beérkezett: 2010. IV. 6-án.



Zolnai László

az MTA Atommagkutató Intézete
Kísérleti Magfizikai Osztályának
fizikus kutatója.
E-mail: zolnai@atomki.hu

A tartalomgyárak lehetnek a média sírásói

A Demand Media naponta több ezer szöveges bejegyzéssel és videóval árasztja el az internetet. A *Richard Rosenblatt* által életre hívott vállalat napjainkra lett meghatározó tényező a tartalomszolgáltatók között és a különböző kiadók aggodnának is a jövőjük miatt. „Minden nap elolvasom az újságokat és még több hírt a világhálón, szeretem a híreket” – jelentette ki a *Financial Times*-nak Rosenblatt, akit sokan nemes egyszerűséggel az újságírás és a média sírásójának neveznek.

A 40 éves szakember által felépített birodalomra jellemző, hogy a YouTube egyik legnagyobb bedolgozója és 1,3 millió szöveges bejegyzést, tudósítást szállít a különböző online sajtóorgánumoknak. A cég 7000 munkatársa naponta több mint 4500 új anyagot készít. Az alig négy éve alapított vállalat rövid időn belül a hagyományos média ellenségévé vált. *Jay Rosen*, a *New York-i Egyetem* professzora egyenesen ördöginek nevezte a Demand Media által alkalmazott üzleti modellt. Mások olyan szavakkal illették, mint „digitális tartalomfarm”, „válaszgyár” vagy „kihasználóüzem”.

Vannak, akik úgy vélték, hogy Rosenblatt afféle gyorséttermi internetet valósít meg, amely kizárólag a keresésekre adandó válaszokra és az internetes hirdetésekre összpontosít. Rosenblatt értetlenül áll a jelenség előtt és semmiképpen sem a hagyományos média sírásójaként tekint magára.

Mi a siker titka? A Demand Media minden olyan tartalmat szállít, amire az emberek a keresőkön keresztül kíváncsiak. Ezek lehetnek hírek, videók, szöveges anyagok. Az üzleti modell alapja egy kifinomult algoritmus, amelynek a segítségével pontosan tudják a szakemberek, hogy mik a Google és más keresők legnépszerűbb fogalmai. A különböző válaszokhoz akár videók is készíthetők és a tartalmakat szétosztják a YouTube, valamint további 45 platform között. A módszer működik, hiszen a Demand Media honlapjait havonta 50 millió felhasználó keresi fel. A cél egyszerű: minden anyaggal pénzt kell keresni. Ez sikerül, hiszen tavaly a cég forgalma elérte a 200 millió dollárt.

Rosenblatt igazi veteránnak számít az online piacon, hiszen 1994-ben létrehozta az iMallt, amit később 565 millió dollárért adott el. 2004 elején befektetők felfogadták az Intermix Media vezetőjéül. A vállalathoz tartozott többek között a MySpace nevű portál is. Rosenblatt a csőd szélén táncoló birodalmat felvirágoztatta, majd öt évvel ezelőtt eladta a MySpace-t 650 millió dollárért *Rupert Murdoch*nak. Az üzleten 23 millió dollárt keresett. Nem véletlen, hogy ha Rosenblatt neve felbukkan egy vállalat papírjain, akkor a befektetők is tolongani kezdenek. A Demand Media partnere már a *Goldman Sachs* és az *Oak Investment*, e két óriáscég összesen 355 millió dollár kockázati tőkét biztosít a vállalkozásnak.

Egy néhány száz szavas anyagért a vállalat 15-20 dollárt fizet, de a jó újságírók akár 80 dollárt is kereshetnek leadott anyagokként. Egy videokészítő 20 dollárt kaszálhat művenként, az adatellenőrzők szövegenként 1 dollárt kapnak. Bár némelyik összeg megalázóan alacsony, a korábban elbocsátott szakemberek kapva kapnak a lehetőségen és örülnek, hogy ismét alkalmazottak lehetnek valahol. A jelenséggel egyedül a médiaipar nem tud mit kezdeni. Aggódva, de ugyanakkor irigykedve figyelik Rosenblatt előretörését, próbálják megfejteni a titkát. Pedig talán nincs is titok, vagy ha van is, nagyon egyszerűen hangzik. „Mi tartalmakat adunk az oldalaknak és utána osztozunk a reklámbevételeken. Ezek olyan anyagok, amelyeket az újságok nem készítenének el maguk, de többletinformációkat jelentenek az olvasóknak. Olyan utat találtunk, ahol kiegészíthetjük a hagyományos médiatartalmakat. Nem vagyunk az újságírás megmentői, de a lerombolói sem” – összegezte cége filozófiáját Rosenblatt.

/SG.hu Hírlevél, 2010. március 23., <http://www.sg.hu/>

(SzP)