

Quo vadis, tudományos publikációs adattár?

**TMT Kézirat
2010. 02. 12.**

Írta: Adománné Zolnai Dóra
könyvtáros (dzolnai@atomki.hu)
és
Zolnai László
fizikus (zolnai@atomki.hu)

**MTA Atommagkutató Intézete
4026 Debrecen, Bem tér 18./c
Tel: 06-52-509-220
2009.**

Összefoglaló:

A szerzők tanulmányukban, amely az első szerző diplomamunkájának átdolgozott változata, arra a kérdésre keresnek választ, hogy mi az oka a hazai tudományos publikációs adatbázisok fejlődése megtorpanásának. Ebből a célból, a hazai kínálatból öt reprezentatív mintát vesznek szemügyre, kitekintve a nemzetközi adattárak működésére is. Az adattárak létrejöttének rövid áttekintése után több szempontból összehasonlítják az adattárak működését. Ezután sorra vesznek néhány olyan problémakört, amelyek megoldása nélkül nehéz lenne az előrelépés. Vizsgálatuk eredménye azt mutatja, hogy a nem megfelelő működés okai többnyire a tervezésnek abban a gyakorlatában keresendők, amely egyrésztől nem definiálja pontosan az adatbázisok tudományometriai felhasználásának módját másrésztől nem veszi figyelembe a színvonalas működtetés erőforrás igényét.



Adománné Zolnai Dóra
Az MTA Atomagkutató Intézete
könyvtárának könyvtárosa
E-mail: dzolnai@atomki.hu)



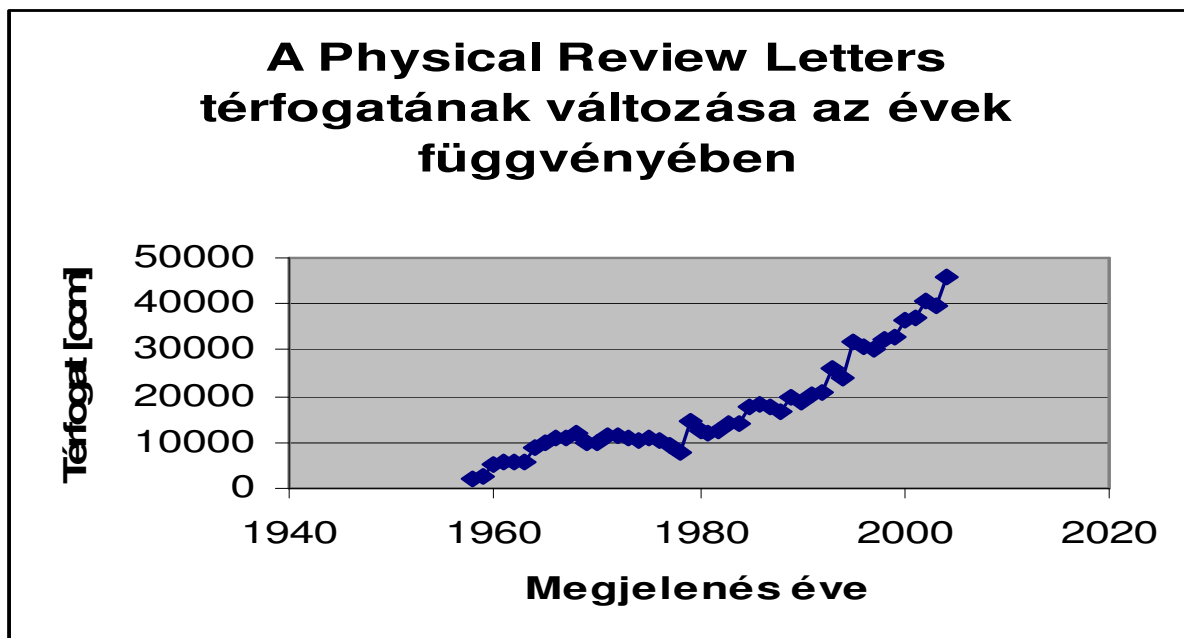
Zolnai László
Az MTA Atommagkutató Intézete
Kísérleti Magfizikai Osztályának
fizikus kutatója
E-mail: zolnai@atomki.hu

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 elmaszatólása és a tervezési szakaszban történő, a működtetés feltételeit figyelembe nem vevő aránytévesztések következményei.

1.0 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 cikkeket közlő folyóiratának, a Physical Review Letters térfogatának változását mutatja a folyóirat 1958-as indulása. Tekintettel arra, hogy a folyóirat kezdettől fogva négy oldalas „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 hatásokon túlmenően a megjelenő tudástömeg elkerülhetlenné tette:

- A művek legalább címszerinti nyilvántartását, amely a közelmúltban lejátszódó informatikai forradalom nélkül valószínűleg megoldhatatlan feladat lett volna.
- A művek és a rájuk vonatkozó hivatkozások nyilvántartását abból a célból, hogy ezekből az adatokból képet kaphassunk a művek szerzőinek munkásságára (teljesítményére). E célra külön tudományág fejlődött ki (evaluatív tudománymetria)

Részben ezeknek az igényeknek a kielégítésére jöttek létre az u. n. publikációs adattárak, melyeknek első – és talán legismertebb képviselője az amerikai Institute of Scientific Information (ISI, ma a Thomson cég) által összeállított Science Citation Index (SCI) (a weben ma elérhető formáját Web of Science-nek (WoS), később Web of Knowledge-nek nevezik) (WoK).

Intézményi publikációs nyilvántartások hazánkban a 60-as évektől léteztek (az általunk ismert legrégebbi, az MTA Atommagkutató Intézetének szélyukkártvás megoldása volt).

Jelentős késéssel – a hazai informatikai háttér megerősödésével – megjelentek a korszerű, web-es felülettel bíró hazai publikációs adattárak (a legkorábban 1998-ban az MTA Atomki publikációs-hivatkozási adatbázisa (APHA)¹. Mára 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 ³-ban ismertetett adattárakon túl dolgozatunkban részletesen foglalkozunk az MTA Tudományos Publikációs Adattárával (ATPA), az MTA Atommagkutató Intézete 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 adatbázisa pedig egy jól működő példa kisebb bölcsészettudományi adatbázisra.

Vizsgálataink során elsősorban az interneten elérhető információkra támaszkodtunk. Egyes esetekben személyesen is megkerestük az illető adatbázisok kezelőit. Ugyanezen a módon szereztünk – igaz, hézagos – információt az egyes adattárak jövőbeli terveit illetően

A tudományos publikációs tevékenység összefüggéseinek tudományos vizsgálatára egy külön tudományág, a *tudománymetria* alakult ki. Az erről a tudományágról a magyar

nyelven olvasó először Dobrov, 1973-ban fordításban megjelent művéből⁴ tájékozódhatott. Körülbelül ebben az időben indultak a hazai tudományometriai kutatások, melyeknek úttörője Braun Tibor volt, A hazai tudományometriai kutatások legfrissebb áttekintését az⁵ közlemény tartalmazza.

Minden bizonnyal a történetileg a legelső teljesítmény értékelési 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 számos tudományometriai statisztikai jellemző (angolul: Science Indicator) definiálására került sor. Ezek közül számos megtalálható az APHA-ban, országokra vonatkozólag ld.⁶

Használatának népszerű volta miatt meg kell említeni az impakt faktort (impact factor)^{7,8}.

Az idő előrehaladtával az impakt faktort, illetve annak egy adott publikáció halmazra felösszegzett (esetleg szerzőszámmal beosztott) változatát elkezdték használni egy szerző, intézmény vagy ország tudományos teljesítményének jellemzésére⁹. Ebben az értelemben használják az általunk vizsgált adatbázisok is az egyetemi adattárak kivételével. Ebben az időszakban a folyóiratértékelés számos mutatóját (frissességi index, idézési félidő, idézettségi félidő¹⁰, stb.) kezdték a kutatói teljesítményértékelés területén is alkalmazni. Teret nyertek olyan számszerű jellemzők alkalmazása is mely a publikálási szokásokkal vannak kapcsolatban (aktivitási index, produktivitás¹¹, stb.)

A legutóbbi idők terméke az egyes szerzők tudományos teljesítményének jellemzésére kifejlesztett Hirsch-index (h-index)¹².

A Hirsch-indexet hazánkban az elmondottaktól függetlenül elkezdték alkalmazni, ami a KPA hiányosan feltöltött adatain alapult¹³.

Az evaluatív tudománymetria széles körben elfogadott álláspontja szerint a tudományos teljesítmény értékelésekor körültekintően kell eljárni, és minden esetben ki kell egészíteni ú.n. *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 az egyes esetekben, szóban forgó művek nagy száma miatt nem lehetséges a gyakorlatban. Ezen okok miatt – bár senki sem szereti – elkerülhetetlen a tudományometriai statisztikai jellemzők mind szélesebb körű használata. A számszerű jellemzők használata elsősorban makroszinten (országok közti összevetés) vezet megbízható

eredményre, mezo- (intézményi) és mikroszinten (egyéni teljesítményértékelés) alkalmazása nagy körültekintést igényel¹⁴.

Az evaluatív tudománymetria alkalmazását, az alsóbb szintek felé haladva egyre fokozódó élességű 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 csoportérdekek kifejezői, tudatosabb esetekben a *tudománypolitikai marketing* (v. ö. ¹⁵) lépéseinek folyamányai. Ez utóbbiaknak rendszerbe foglalása a jövő kutatásainak témája lehet.

A publikációs adattárak kimondva vagy kimondatlanul a láttatás és értékeltetés 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 a WoK megteremtette a társadalom- és bölcsész tudományok területén az SCI megfelelőit (SSCI és A&HCI). Míg ez előbbi adattárak, nagy részben tartalmazzák a hazai vezető természettudományi területen dolgozó kutatók legutóbbi évekbeli munkásságát, azokon a területeken, ahol a magyar nyelvű vagy a perifériális folyóiratokban való közlés a népszerűbb, vagy a tudományos teljesítmény nyomai a nemzetközi adatbázisok által nem indexelt művekben található, nagy igény mutatkozik a tudományos – és nem ritkán egyéb – publikációk adatainak felmutatására. Ez az igény felismerése a dolgozatunkban vizsgált valamennyi publikációs adattár létrehozásának az oka. Megjegyzendő, hogy a létrehozó szervezetek eltérő érdekei – meg az interneten való realizálhatóság viszonylag egyszerű volta – vezettek a viszonylag nagyszámú, hazai publikációs 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 – és egyes helyeken a hivatkozások – adatainak láthatóvá tételével.

Új korszakot jelentett – és ez indokolja a viszonylag hosszú fejtegetést a 2.1 fejezetben – a tudományos teljesítmények értékelésének nemzetközileg, és majd hazánkban kialakuló gyakorlata. Ez direkt módon hatással volt az egyes kutatók/intézmények sorsára. Közvetlenül ez – az MTA intézeteinek 1996-os értékelése (az értékelésnél alkalmazott módszer leírását ld.¹⁶, - volt a kiváltó oka az APHA web-en való megjelenésének, és egy majdani intézeti teljesítményértékelés kimondott vagy kimondatlan terve – amint az adattár ismertetőjéből kitűnik („Az összegyűjtött adatok hozzájárulnak a szakirodalmi információs folyamatok

törvényszerűségeinek feltárásához is.”) – vezetett az ATPA megalapításához (ezt szervezetileg az MTA főtitkára alá rendelték). A KPA létrehozásának (az MTA elnökétől kiinduló) indítéka az volt, hogy az MTA levelező és rendes tagjain kívül a köztestületi tagoknak (akik az MTA „résztevői” az MTA-nak az 1993-as akadémiai törvény által való felduzzasztása óta) lehetőséget nyújtson munkásságuk nyilvántartására, a web-en való megjelenéséhez. Később ez az eredeti cél – vélhetőleg a gyér érdeklődés miatt – 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, legújabbban pedig az Országos Doktori Tanács (ODT) adatbázisa¹⁷ is ennek használatát favorizálja.

Az ATPA és 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 az 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 realizálásához vezethetnek.

A fenti fejtegetésekkel célunk az volt, hogy megpróbáljuk bemutatni, hogy a hazai tudományos publikációs adattárak fejlesztésének kérdései milyen érdekek folytán vált fontossá.

2.0 A vizsgált tudományos publikációs adattárak ismertetése

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 különböző egyetemen keletkező publikációk (dokumentumok) könyvtárszakmai feltárását (kereshetőségét) helyezik előtérbe, a többiek viszont főleg a szerzők és intézmények publikációs (hivatkozási) listáit tárják a felhasználó elé. Ez utóbbiak több, kevesebb tudományometriai statisztikai jellemzőt is produkálnak, sőt egyikük ezek idősorait is képes előállítani. A vizsgált adattárakat egy kivétellel röviden ismertetjük. A viszonylag hosszan előtárt APHA ismertetését az indokolja, hogy bár az országban a legrégebbi, kevésbé ismert, a hazai irodalomban még nem került leírásra.

2.1 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 működik.

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 összegyűjtött adatok hozzájárulnak a szakirodalmi információk folyamatok törvényszerűségeinek feltárásához is.

Az intézetek adatai 1992-től, az akadémiai kutatócsoportok adatai pedig 1996-tól szerepelnek az adatbázisban. A 2008-as adatok feldolgozása után az Adattár több mint 300

ezer tételt tartalmazott (100000 publikáció, 200000 hivatkozás). Az MTA TPA-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 lekérdezhetők. 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 rendezhetők, megjeleníthetők és menthetők. Az idéző publikációk adatainak lekérdezéséhez felhasználói név és jelszó is szükséges.

Annak a felhasználónak, aki saját maga akarja adatait bevinni az adatbázisba az ATPA két programot bocsájt rendelkezésére. Ezek a forráspublikációk és idézetek adatainak bevitelére szolgáló Windowsos TPA Forrás és TPA Idézet nevű programok.

Azok az intézmények, amelyek saját adattárakat működtetnek (pl. APHA), az említett programokat szerkesztő programként alkalmazzák az ATPA felé történő adatexportálás céljából.

2.2 A Magyar Tudományos Akadémia Köztestületi Publikációs 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 2003. február 25-i, 1/2003. sz. állásfoglalásában határozott.

A KPA felállításával és működtetésével a Köztestületi Adatbázis 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 adatbankok nem figyelnek.

Például könyvek, a nemzeti tudományok magyar nyelvű publikációi, tankönyvek. Sajnos az adatbázis 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 módja az adattárba: Az adatbevitel első lépése: telefonon, vagy e-mail útján történő bejelentkezés a KPA illetékes munkatársainál

A bejelentkező felhasználói nevet és jelszót kap. A köztestületi tag nevében csak az általa megbízott személy vihet be adatokat.

Az adatok bevitele az adattárba:

Amennyiben az adatszolgáltatónak van a Web of Science, vagy más hasonló nemzetközi adatbázisban nyilvántartott publikációja és idézettségi adata, úgy ajánlatos az adatgyűjtést ezen adatok letöltésével kezdeni. Az így kapott adatfile-okat az internetes adatbevitel megfelelő helyén fel kell tölteni az adatbázisba. Szövegszerkesztővel készített publikációs és idézettségi listát az adatbázis 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 ellentmondásmentességét a KPA munkatársai ellenőrzik, és az adatokat az adatszolgáltatónak ellenőrzésre visszaküldik. Csak az adatszolgáltató jóváhagyása után válnak az adatok honlapjukon mindenki számára elérhetővé¹⁸.

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

2.3 A Pázmány Péter Katolikus Egyetem Bölcsészettudományi Karának Oktatói Publikációs Adatbázisa (PPKE BTK)

A PPKE BTK publikációs adatbázisának – az egyetemi bibliográfiának – célja az oktatók tudományos közleményeinek nyilvántartása és egységes szerkezetű bemutatása. 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 adatbázisba. Nem regisztrálják a konferencia-előadások adatait és a konferenciafüzetekben megjelent absztraktokat.

A PPKE BTK adattára két változatban él: az adatbázis 2002 októberében, a webváltozat fél évvel később, 2003 tavaszán kelt életre. Kialakítása során a már működő adatbázisok körében felmérést is végeztek, melynek eredményeit hasznosították.

Az adatbázis 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 jegyzékeket a keresőrobotok rendszeresen indexelik, és a találati listák elejére sorolják.

Az adatbázis kézi lekereséssel és adatbányászattal bizonyos tudományometriai jelzőszámokat is számol és tesz közzé.

További részletes ismertetése megtalálható Dudás Anikó cikkében¹⁹.

2.4 A Miskolci Egyetem Publikációs Adatbázisa (MEPA)

A *Miskolci Egyetemi Publikációs Adatbázis (MEPA)* 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 oktatóinak, 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 kellenek 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, amennyiben 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. Azon egyetemi oktatók, akik digitális formában is hozzáférhetővé teszik publikációik teljes szövegét itt, írásaikkal automatikusan bekerülnek a Nemzeti Digitális Adattár (NDA) adatbázisába, ezzel is növelve munkáik ismertségét, kereshetőségét.

Az adatbázisban az egyetemi tudományos élet legfrissebb dokumentumai válhatnak kereshetővé, alkalmat teremtve például az azonos területen dolgozó esetleges kutatási partnerek megtalálására.

Az adatbázis részletes ismertetése megtalálható a ²⁰ közleményben, bizonyos szempontok szerinti értékelése pedig a ²¹ irodalomban.

3.5 Az MTA Atommagkutató Intézet Publikációs-Hivatkozási Adatbázisa (APHA)

A Magyar Tudományos Akadémia Atommagkutató Intézetét 1954-ben alapították Debrecenben. Az intézet könyvtárának fő gyűjtőköre a fizika, ezen belül az atom- és atommagfizikai alapkutatáshoz és ezek ipari, valamint egyéb tudományokban történő alkalmazásához szükséges elméleti és kísérleti tárgyú szakirodalom. A könyvtár feladata elsősorban az intézet dolgozóinak szakirodalommal való ellátása valamint az intézet publikációs tevékenységének, valamint a publikációkra történő független hivatkozások (olyan hivatkozások, ahol a hivatkozó és hivatkozott cikknek nincs közös társszerzője) regisztrálása. A könyvtárra vonatkozó fontosabb információk megtalálhatók a könyvtár honlapján²².

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. Meg kell jegyezni, hogy az intézetben az előtt is létezett publikációs nyilvántartás, először egyéni listák, majd központi széllyukkártyás nyilvántartás formájában.

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 tudományometriai 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 ~ 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 ~ 100 kutatója van, ez azt jelenti, hogy az egy kutatóra évente kb. 3 „kemény” publikáció esik. Az intézet elmúlt 50 évének produktumaira évente kb. 2500-3000 független hivatkozást kap. Ezeknek a hivatkozásoknak a begyűjtése külön is hatalmas feladat, annak dacára, hogy az utóbbi években nagy segítséget nyújtanak a különböző elektronikus, interneten keresztül hozzáférhető adatbázisok (pl. az Oktatási Minisztérium által előfizetett Web of Science).

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 betehetik az adatbázisba. Ezen kívül az adatbázis gyűjti az intézetben készült diplomamunkák adatait is. A Debreceni Egyetemmel való együttműködés keretében helyet ad a DE „Fizikai Tudományok” Doktori Iskolája hallgatóinak publikációit és az azokra való hivatkozások adatainak is. Megemlíthető, hogy ez a tény nagyon megkönnyíti a PhD fokozat megszerzése után az intézet alkalmazásába kerülő munkatársak – egyes esetekben több százra rúgó számosságú - publikációinak és hivatkozásainak későbbi nyilvántartását.

Az adatbázis publikációk fajtáit tekintve – egyedülállólag a vizsgált adattárakat tekintve, helyet ad a gyűjtőkörbe tartozó szerzők kutatási jelentései (report), előadásai és évkönyvbeli cikkeinek, és az esetlegesen ezekre való hivatkozások adatainak is.

Végül néhány adat az intézeti adatbázis jelenlegi nagyságára: Az adatbázisban szereplő szerzők száma ~19.000, megjelenési helyek száma ~14.000, publikációk száma ~22.000,

hivatkozások száma ~44.000. Az adatbázissal kapcsolatos részletes statisztikák megtalálhatók a ²³ web-lapon.

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

1. Az intézet vezetésének különböző vezetési döntések meghozatalához,
2. Az MTA vezetésének az intézet előző évi teljesítményéről (februárban),
3. Részletes publikációs-hivatkozási adatokat szolgáltasson az MTA központi Tudományos Publikációs Adattárának²⁴ (március-május).

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

3.0 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 Köztestületi Publikációs Adattár	Miskolci Egyetem Publikációs Adattár	Pázmány Péter Katolikus Egyetem Bölcsészettudományi Kar Oktatói Publikációs Adatbázisa	MTA Atomki Publikációs-Hivatkozási Adatbázisa
használt rövidítés	(MTA TPA, ATPA)	(MTA KPA)	(MEPA)	(PPKE BTK OPK)	(APHA)
URL	http://www.mtatpa.hu/	http://www.mtakoztest.hu/kpa.htm	http://publikacio.uni-miskolc.hu/user/	http://www.btk.ppke.hu/cikk.php?cikkek=602	http://www.atomki.hu/p2/
Létrehozás dátuma	2001	2003	2005	2002	1986
Platform	mySQL, PHP	mySQL, PHP	mySQL, PHP	T-Series HTML	FORTRAN-77, Perl, HTML
Nyilvánosság	Publikációs adatok nyilvánosak, hivatkozási adatok jelszóval láthatók	Nyilvános	Nyilvános	Nyilvános	Nyilvános
Kitöltött rekordok száma	Publikáció: ~110000 Hivatkozás: ~300000	Publikáció: ~100000 Hivatkozás: ~200000	publikáció ~30000	Publikáció ~12500 Hivatkozás ~200	Publikáció: ~22000 Hivatkozás: ~42000
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 becsl létszáma	~3000	~14000	~1500	~250 (eddig 212 szerző)	~500
Gyűjtőkör	Publikációk és hivatkozások Az MTA kutatóhelyek tudományos és ismeret-terjesztő közleményei a publikációkra érkezett idézetek, az Akadémia doktora értekezések és az MTA kutatóhelyeken készült PhD disszertációk	Publikációk és hivatkozások 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. Hivatkozás: tervezi 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 hivatkozások 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 hivatkozások Reportok, szabadalmak, előadások, annual reportok Szakdolgozatok és disszertációk.
Angol nyelvű felület	részben	részben	nincs	nincs	igen
Feltöltő személye	Kutatók, oktatók Önálló adatbázissal	Kutatók, oktatók	Bejelentés alapján a könyvtár dolgozói,	Bejelentés alapján a könyvtár dolgozói	Bejelentés alapján a könyvtár dolgozói

	rendelkező intézmények.		az egyetemi oktatók sk.		
Adatforrások	Másodlagos	Kisebb mértékben elsődleges	Könyvtáros bevitele: Másodlagos. Oktatói bevétel elsődleges.	Publikációk: másodlagos	Publikációk: elsődleges, hivatkozások: másodlagos
Ellenőrzés	adatok nyilvánvaló ellentmondás-mentességét ellenőrzik	adatok nyilvánvaló ellentmondás-mentességét ellenőrzik	Publikációk: elsődleges és másodlagos forrásokból, esetleges	Publikáció: elsődleges forrásból	Publikáció: elsődleges forrásból Hivatkozás: esetlegesen
Globális kereshetőség ²⁵	nincs	nincs	nincs	kiváló	kiváló
Lokális kereshetőség	közepes	kiváló	jó	Csak név szerint	kiváló Intézményen belül ennél is részletesebb.
tárgyszavazás	nem megoldott	nem megoldott	mód van rá, hiányos, nem egységes	nem megoldott	nem megoldott
Naprakésztség (a telítés 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	Intézeti éves listák, Szerzők publikációs-hivatkozási listái Téma szerinti listák Listához csatolt tudományometriai

				jellemzők	jellemzők, Tudománymetriai jellemzők idősorai
Tudománymetriai jellemzők	Számos	Számos	Alig	Néhány	Számos Idősoros is
látogatottsági statisztika	nincs	nincs	Van (böngészés, gyorskeresés, összetett keresés)	Időnként, manuálisan készítik	Van (honlapon elhelyezett számláló)

4.0 Szöveges összehasonlítás, következtetések

Az előző fejezetbeli táblázat sokféle szempontból 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 adatok alkalmazásra kerülnének ez egyes szervezetek által végrehajtott értékelésekben.

4.1 Az adatfolyamok összehasonlító elemzése

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

A nemzetközileg elsősorban számon tartott publikációs adatbázisok (WoS, Scopus, stb.) az adatok forrásai maguk a feldolgozni kívánt művek (folyóiratok, konferencia kiadványok, stb.), amelyeket – megfelelő tőkeerő birtokában – valószínűleg (olcsó, távol-keleti) bér munkában dolgoztatnak fel az ilyen adatbázisokat üzemeltető cégek. Mint fentebb definiáltuk, ez elsődleges adatfeltöltést jelent. A hazai adatbázisok is részben használják ezt a módszert, csak itt a bér munka helyett a különböző – nem mindig hatásos – 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 – főleg a hivatkozások tekintetében – igénybe veszi a WoK szolgáltatásait is. A KPA eredeti koncepciója szerint a köztestületi tagok adatbevitelére támaszkodik, bár a kezdeti – nagyobb mennyiségű – feltöltésre felajánlja a WoK, Medline, Scopus és ATPA-beli adatok átvételét. Ez utóbbi, a KPA kezelők által végzett átvétel rendkívül sok időt vesz igénybe, és jellemzően csak a magasabb akadémiai ranggal bíró köztestületi tagok anyagának felvitelekor működik. A KPA-t megpróbálják a nyomásgyakorlás egyéb eszközeivel is igénybe vetetni (ajánlják használatát a Bolyai-ösztöndíj, MTA Doktora és akadémikusi 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 régebbiről 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 a szintén bekért, publikációkról készített másolatokról. Sajnos, a hivatkozásokat csak esetleg sikerül ellenőrizni, akkor is többnyire a WoK segítségével.

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, ezektől a hibáktól a leginkább „megbízható” adatforrásul használt nemzetközi adatbázisok (WoK, Medline, Scopus) sem mentesek, mivel azok is a nyomtatásban, elektronikus formában (CD, DVD) vagy a világhálón megjelent művekre támaszkodnak.
- Valamennyi (a KPA különösen) vizsgált adattár egyre jobban támaszkodik az ismert nemzetközi adattárakra (WoK, Medline, Scopus), mintegy hazai infrastrukturális adottságként véve ezek hozzáférhetőségét. Ezek azonban nem egy 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 biztosító 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 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) alapoz, 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, ill. a más adatbázisok felé irányuló export 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 felülettel (valószínűleg a fejlesztésre fordított nagyobb befektetés miatt) a KPA rendelkezik, 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 szervezetileg 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 ebben az irányban 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 APHÁ-é mondható a leghatékonyabbnak, ui. az adatbázis kezelői tevékeny részt vállalnak a „nem motivált” szerzők publikációinak – régebben hivatkozásainak is – felkutatásában. A motiváltság hiánya túlnyomórészt az 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 (WoK, Medline, Scopus) való tömeges „adatleszívás” és az adatoknak a web-en való megjelenítése valószínűleg jogi aggályokat is felvet.

2.2 Az „inaktív” szerzők esete

Mint a fenti fejezetben tárgyaltuk, abban a szerzői körben, amelytől a primér 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 hivatkozások 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, amennyiben 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 érzékeltetésül megjegyzendő, 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 inaktív szerzők munkásságának nyilvántartását indokolják tudománytörténeti szempontok is. Gondoljuk meg, hogy mennyivel egyszerűbb lesz a híressé vált kollégák

munkásságának áttekintése a jól feltöltött tudományos publikációs adatbázisok megteremtése esetén

4.3 Az azonos nevű szerzők problémája

A publikációs adattáraknak nemcsak a publikációk adatainak a gyűjtése a feladatuk, 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 szétosztása 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 on-line 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 kódolnak²⁶ Ennek az információnak a forrásai:

- A szerzők intézetének adata.
- A publikáció témája.
- A beazonosított intézeti szerző múltbeli társszerzőinek áttekintése az adatbázis alapján.
- Az intézeti szerzőkkel való telefonos vagy személyes konzultáció.

Ennek a módszernek a hátrányai:

- csak viszonylag kis létszámú célcsoport esetén alkalmazható könnyen,
- viszonylag munkaigényes,
- csak a szerzői célcsoportot tekintve megbízható, külföldi – főleg távol-keleti – szerzők esetén nem pontos.

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 web-es felületre való közzététel előtt minden publikáció és idézet kap egy „kvázi-tulajdonos”-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 a

saját listájába, de azok adatait nem szerkesztheti. Látható, hogy ez a rendszer a rendszeren kívüli azonos nevű társszerzőkkel nem törődik.

Mindegyik felsorolt metódus igénybe veszi a célcsoport tagjainak ismeretanyagát – talán az APHA, amely a legjobban megkíméli a szerzőket ettől a feladattól.

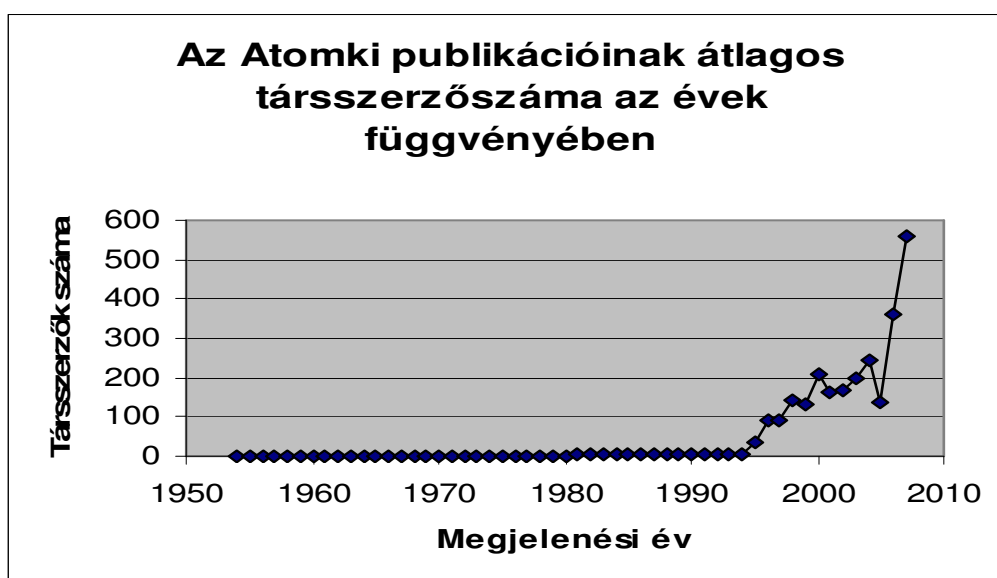
4.4 Sok társszerzős publikációkkal/hivatkozásokkal kapcsolatos problémák

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 publikációk megjelenése. A nagy létszámú együttműködések először a fizikában jelentek meg, de napjainkban a biológiában, 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 társszerzőszám⁹. 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. Egyes CERN-es (Conseil Européen pour la Recherche Nucléaire – Európai Magkutató Tanács) publikációk szerzőire csak a CERN bizonyos belső kutatási jelentéseiből (report) lehet következtetni.

A probléma sajátosan érzékeny területet érint; a szóban forgó publikációk a tudomány frontvonalában 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) megelőlszenek az et. al. jelölés használatával. Az alábbiakban néhány, a sok társszerzős publikációk megjelenésével kapcsolatos problémát veszünk sorra:

- Az 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 eldöntését az adattárak általában a hivatkozott publikáció szerzőire bízják. Egyes adattárak (ATPA és KPA) nagy erőfeszítéseket tettek a függetlenség gépi detektálására, a WoK-ból megszerezve a hivatkozott és hivatkozó művek teljes szerző névsorát. Természetesen ez sem teszi lehetővé a hivatkozások függetlenségének megnyugtató megállapítását, mivel az esetlegesen többeszerző gárdában a társszerzőszámmal együtt növekvő valószínűséggel fordulnak elő azonos nevű (főleg távol-keleti) szerzők. Meg kell jegyezni, hogy a függetlenség kérdése kulcsfontosságúnak számít az egyes kutatói és intézmény teljesítmény-értékelési eljárásokban.
- 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 (ld. 4.5 fejezet)
- 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.

4.5 ISI/nem ISI, SCI/nem SCI

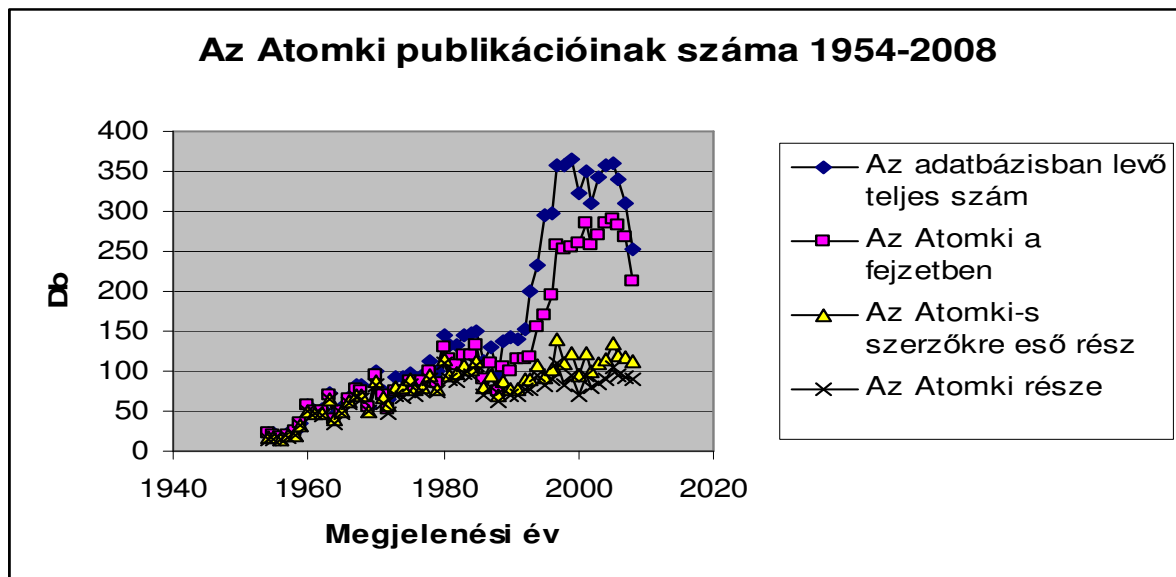
A tudományos kutatókra ráömlő rengeteg publikáció a tudományometriai értékelő eljárásokat is hatalmas feladat elé állította. Célszerűnek látszott a megjelent publikációkat szűrni a megjelenési hely „kiválósága” szerint (hazai/nemzetközi, magyar/idegen nyelvű, referált/nem referált, impakt faktoros/nem impakt faktoros, 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 konferencia kiadványokat is visszamenőleg. Ettől némileg eltér az APHA gyakorlata, amely az SCI megjelölést használja, továbbá – technikai okokból – megkülönböztetést tesz publikáció és hivatkozás között²⁷.

4.6 A teljesítmények additivitásának kérdése - az 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. Amennyiben 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 1 publikációt, illetve hivatkozás esetén a hozzá tartozó hivatkozások teljes számát eredményezze (az evaluatív tudománymetria alapvető egységei a publikáció és az idézet). Ennek a problémának hazai felvetése ²⁸-ben történt, további taglalása megtalálható a ²⁹ és ³⁰ -ban. A szóban forgó felosztás történhet egyenletesen, vagy nem egyenletesen. Ez utóbbira példa a szabadalmaknál alkalmazott százalékos felosztás. Ilyen lehetőséget biztosít a KPA, azonban ott ezek a mezők publikációkra nézve általában nincsenek kitöltve. Valószínűleg ez a megoldás lenne kielégítő az egyéni teljesítmények megítélésékor. 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 ABC-sorrend a szokásos. Véleményünk szerint számosabb publikációt tartalmazó lista esetén az „érdem” kiátlagolódik, u.i. a több résztvevő együttműködések nem túrik meg a „lazsálást”, ha valaki rajta van a szerző listán, az valamikor azt kiérdemelte. Az egyéni szinttől eltávolodva, nagyszámú publikáció esetén alkalmazható az „érdemek” egyenletes elosztása. Ennek előnye egyszerűségében rejlik, elég

csak a társszerzők számát, ill. az illető intézménynek a előfordulási hányadát rögzíteni. Pl. 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 részarány³¹ $1/n$, 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.



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

A 3. ábrán négy adatsort láthatunk. A legfelső az adatbázisban az intézeti szerzők által bejelentett összes publikáció számát mutatja (itt az intézeti szerzők kifejezést a szerzők azon halmazára használjuk, akik az intézettel valamilyen munkajogi kapcsolatban voltak/vannak). A következő adatsor azoknak a publikációk számát ábrázolja, amelyekben a fejtetben megjelent az Atomki neve (a „permanent address...”, és „on leave from...” lábjegyzetbeli kifejezéseknek ebből a szempontból csak informatív értéke van, ha az intézet neve nem szerepel a fejtetben, 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 publikáció számot mutatja. A negyedik

adatsor az intézetre eső publikációk számát ábrázolja. Jól látható az egyes esetek közti különbség.

Az irodalomban fellelhető országok és intézetek közti összevetések (ld. pl. ³²) általában a WoK adataira alapozódnak, amelyben minden publikáció mellett megtalálható az az információ, hogy a cikk létrehozásában milyen intézetek kerültek a fejtetbe (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” módszer¹⁴). Így egyes országok és/vagy intézetek a valóságos helyzetnél jóval kedvezőbb színben tűnhetnek fel, amint az már más szempontból is felvetődött³³. Természetesen a fentiek csak akkor igazak, ha mint az evaluatív tudománymetria általában teszi, a publikációkat azonos statisztikai értékűnek fogadjuk el.

4.7 Publikációs adattár és/vagy repozitórium

Az informatika fejlődésével (elsősorban az internet megjelenésével) 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 politikái (ld.:³⁴) azonban ezt nem mindig teszik lehetővé. Ennek megoldására került előtérbe a Digital Object Identifier (DOI) használata, amely egy olyan kapcsolat (link), amely születésétől kezdve az illető cikk digitális változatára mutat, és attól függően tudja látni a felhasználó, hogy az illető cikk teljes változatának megtekintésére van-e (rendszerint IP számhoz kötött) jogosultsága. A teljes szövegű hozzáférés lehetősége lehet integráltan az illető publikációs adattár része, vagy különálló objektum. A WoK és a KPA a DOI-t alkalmazza, az ATPA nem tartalmaz ilyen lehetőséget.

Jelenleg számos, ennek az igénynek kielégítését szoftver van forgalomban. A különböző hazai repertóriumok általában ezekre támaszkodnak. Az egyik legismertebb a MTA Könyvtára³⁵ által használt, és országosan ajánlott szoftver, a southamptoni egyetemen kifejlesztett EPrints3³⁶.

A vizsgált egyetemi adattárak különálló elektronikus archívumot (repozitóriumot) tartalmaznak, amelyek működtetése szerzői jogi szempontból erősen aggályos. Az APHA az előbb ismertetett összes lehetőség (.html, .pdf, link, DOI) használatát lehetővé teszi, azzal a kikötéssel, hogy a bejelentőnek a fentebb említett szerzői jogi feltételek ismeretében el kell döntenie, hogy a teljes szövegű változatot a teljes vagy intézetkörű nyilvánosság számára teszi láthatóvá.

Megjegyzendő, hogy a fenti lehetőségek használhatósága csak az utóbbi pár évben vált lehetővé, a WoK kivételével, ezeknek az opcióknak a feltöltöttsége minden vizsgált adattár esetében minimális, és a 6.1-ben tárgyalt, a „feltöltésre ítétek” motivációjára vonatkozó fejtegetéseink értelmében a közeljövőben nem is várható változás ezen a téren.

4.8 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 – nem részletezett – 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 hétre 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 biztosítani a felhasználóktól a fejlesztőkig/kezelőkig, sőt a fenntartókig irányuló visszacsatolásra (*feedback*). Valamilyen módon képet kellene kapni az illető adattár iránti érdeklődés feltérképezésére. Ennek első lépése lehet az APHA-ban található látogatottsági számlálók szélesebb körű alkalmazása.

4.9 Az adattárak működtetésének gazdasági háttere

Az egyes adattárak működésének 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 – amelyeknek a szükségleteket kielégítő felmérésére ill. reallokációjára 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 táplálkoznak. A szóbanforgó 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 fejlesztési 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,
- abban való hit, hogy a nemegyszer kötetekre rúgó használati utasítást elegenden elolvassák.

A vizsgált publikációs adattárak fejlesztésében és működtetésében a következő érdekcsoportok vesznek részt:

- fenntartó szerv, vagy annak képviselői,
- fejlesztők,
- kezelők,
- felhasználók.

A felhasználókat még két csoportra lehet bontani, aszerint, hogy munkásságukat akarják a szélesebb nyilvánosság elé tárni, vagy pedig mások munkássága iránt érdeklődnek.

A fenntartók általában csak koncepcionális szempontok megadására szorítkoznak, az adatbázis megvalósításának formája, annak kezelhetősége a fejlesztők (informatikusok) és a kezelők (általában könyvtárosok) közti 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 által korlátozott tervezőmunka nem alapos voltában keresendő. A szűkös források sokszor abból adódnak, hogy a projekt előterjesztői – a rendszerváltást megelőző időszakra jellemzően – alulbecslik a megvalósítás költségeit. Ez oda vezet, hogy a matematikai modell realizálása folyamán nem veszik figyelembe a fogadó közeg szervezeti-képességi korlátait, nem beszélve arról, hogy 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 diszfunkciók kiküszöbölése jobb esetben a realizált adattár szoftveres kiegészítését, rosszabb esetben plusz emberi erőforrás hozzárendelését teszi szükségessé. Ez utóbbi – az esetleges szervezeti változtatásokkal együtt – költségvetési környezetben különösen nehezen valósítható meg. Megjegyzendő, hogy az előbb említett változtatások mindegyikére sor került az APHA esetében. Ebből a szempontból még érzékenyebb eset az, amikor új adattár indítása esetével szemben több, különböző modelleken alapuló adattár „in vivo” egyesítését célozza meg a projekt, amikor a felmerülő új forrásigények számosabbak lehetnek. Számos esetben a létező érdekviszonyok nem teszik lehetővé a racionális szervezeti változtatásokat. Ilyenkor a szoftveres megvalósítást teszik bonyolulttá a szervezeti sajátosságok feltétel nélküli tisztelete. Ebben az esetben a rendelkezésre álló források más területeken (pl. a becsületos dokumentáció, on-line sűgők előállítása) korlátozzák a projektcélok elérését.

A fentiekben vázolt igények felismerésénél rosszabb az az eset, ha egyfajta struccpolitika részeként az anomális működésről nem vesznek tudomást, évekig megkeserítve a közreműködők munkáját, ami végül is a projekt fiaskójához vezet. Általában hazai viszonyok között induló projektekre nézve jellemzőnek mondható egyfajta voluntarista megközelítés.

A vizsgált publikációs adattárak működtetési költségeinél általában nem kerül figyelembevételre az adattárak által egyre intenzívebben használt nemzetközi adattárak előfizetési (bérleti) költsége. Ennek a kérdésnek a vizsgálata annál jogosabb mivel ismereteink szerint hazánkban majdnem kizárólag a publikációs adattárak céljaival összefüggésben használják azokat. Kérdéses, hogy mindezt összevéve, hogy aránylik annak eldöntésének 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.

Belátható, hogy a publikációs adattárak létrehozására költött források, annál hatékonyabban hasznosulnak, minél inkább átfed a publikációk (és az arra való hivatkozások) szerzőlistája a megcélzott felhasználócsoporttal. Ennek napjainkban, amikor egyre jobban előtérbe kerülnek a sok társszerzős együttműködések, fokozottabb jelentősége van, a nagy nemzetközi adatbázisok valószínűleg sokkal hatékonyabbak ebből a szempontból.

A gazdaságossági szempontok további összetevője, hogy mennyire sikerül a már egyszer befektetett fejlesztési költségeket megőrizni, vagyis a fejlesztés mennyire időtálló. Félő, hogy a hazai fejlesztésű adattárak ki vannak szolgáltatva az illető fejlesztőcégeknek.

4.10 Következtetések és javaslatok

A publikációs adattárak jövőbeli fejlődésének/fejlesztésének kétség kívül legsarkalatosabb pontja a hazai tudománypolitikának a tudományos teljesítmény értékelésével kapcsolatos álláspontjának minéműsége. Fontos lenne, hogy a tudománypolitika döntse el, hogy az evaluatív tudományometriát a hosszútávú tudománypolitikai döntések segítőjeként vagy az egyes tudománybeli érdekcsoportok taktikai bunkósbotjaként 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 közti kapcsolatok elmélyítése:

- Az eltérő helyen gyűjtött adatok összeférhetőségének javítása, vagy ha az nem megy,
- Az egyes publikációs adattárak közti adat import/export lehetőségének megteremtése. Ennek egy lehetséges módja lehetne a RIS³⁷ formátumszabvány használata.
- Egységesíteni kellene az egyes adattárak által szolgáltatott tudományometriai statisztikai jellemzőket.
- Célszerű lenne az intézményekre és azok szervezeti egységeire vonatkozó statisztikai idősorok bemutatása.

2. Célszerű lenne eltörölni bizonyos gyűjtőkori korlátokat (Egyrészt az ATPA-ban lehetővé kellene tenni az 1992 előtti anyagok keresését, ezzel megnyitva 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 tágítása nem jelentene jelentős kapacitásnövelési igényt.
3. Meg kellene találni a publikációs adattárak szofisztikált szerkezete/kezelhetősége és a felhasználók motiváltsága (az adatbázisok feltöltöttsége) közti középutat. Jelenleg bonyolult szerkezetű adatbázis (KPA) működik alacsony feltöltöttséggel, és relatíve gyatra platformon (FORTRAN) működő adatbázis (APHA) ér el jó feltöltöttségi eredményeket.
4. Megállapítható, hogy a megcélzott szerzőcsoport motiváltsága és feldolgozáshoz rendelkezésre álló kapacitás megszab egy olyan bonyolultsági szintet, amelyen túlmenve a az adatbázisban illetve annak kezelői felületében beprogramozott befogadóképesség öncélúvá válik, és a befektetett eszközök nem hatékonyan hasznosulnak.
5. Az előbbi pontból következik, hogy növelni kell a megcélzott szerzőcsoport motiváltságát és/vagy bővíteni kell azt a kapacitást, amely az adatok előkészítését/bevitelét végzi. Ez utóbbi különösen a KPA esetében lenne fontos. Ennek egyik eszköze lehet, hogy az adattár és az adatszolgáltatásra végül is kötelezettek közé pótlólagos gyűjtőszinteket telepítenek.
6. Az előbbi pontban említett hierarchikus szervezeti felépítésnek szükséges eleme lenne – és ma már a megfelelő informatikai eszközök is könnyen hozzáférhetők -, hogy a különböző szinteken lényegileg azonos programok (klónok) működjenek, lehetővé téve az egymás közti kommunikáció egyszerű módját.
7. Az „inaktív” szerzők munkájának pótlására kezelői szinten pótlólagos kapacitásokat kell beépíteni.
8. Egyértelműen javítani kell a globális kereshetőséget. Ennek eszköze lehet az az APHA-ban alkalmazott megoldás, hogy legalább az egyéni publikációs listákat – az alkalmazott dinamikus web-lap megoldás helyett – időnként statikus, Dublin Core³⁸ metaadatokkal jól ellátott web-lapként újra kell generálni.
9. A globális kereshetőség speciális kérdése az idegen (legalább angol) nyelven való kereshetőség, ami fejlesztésre szorul. Ez vonatkozik az APHA kivételével mindegyik vizsgált adattárra.
10. Meg kellene oldani néhány *minőségbiztosítási* kérdést, lehetőséget kellene biztosítani a felhasználóktól a fejlesztőkig/kezelőkig, sőt a fenntartókig irányuló visszacsatolásra

(*feedback*). Valamilyen módon képet kellene kapni az illető adattár iránti érdeklődés feltérképezésére. Ennek első lépése lehet az APHA-ban található látogatottsági számlálók szélesebb körű alkalmazása.

11. Ugyancsak minőségbiztosítási kérdés a feltöltött adatok megbízhatósága, ami különösen az idézetek esetében fontos egy esetleges értékelés szempontjából. Erre a problémára egy lehetséges megoldás lenne az egyéni/intézményi publikációs/hivatkozási listák mintavétellel történő minősítése.
12. Célszerű lenne a 4. fejezet előző részeiben 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. Gondolunk itt elsősorban a jelenleg a közbeszédben megjelenő „kutatóegyetem” koncepciójára. Nem véletlen, hogy a publikációs adattári fejlesztésekben a BME jár elől, ahol valószínűleg van mit szerbeszámba venni.
13. Lehet, hogy célszerű lenne a rendelkezésre álló forrásokat koncentrálni a hatékonyabb fejlesztés érdekében. Annál is inkább, mivel pl. az MTA-beli kutatók nem egyszer három oldalról vannak adatfeltöltésre kényszerítve.
14. 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 érdeküknek megfelelően. Sajnálatos, hogy a felhasználók igényei csak minimális mértékben kapnak teret.

5.0 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 túlmenően két, - a jelenlegi viszonyok között is finomítható - pontja van:

- Az adattárak esetleges tudományometriai értékelésre való felhasználásának definiálatlansága,
- Nagyobb figyelmet kell fordítani az adattárak működtetésének szervezeti kérdéseire, 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-től létre kívánják hozni a Magyar Tudományos Művek Tárát (MTMT), amely távlatilag egyesíti a publikációk és hivatkozások nyilvántartására vonatkozó igények kiszolgálását³⁹.

Csak remélni lehet, hogy ezen új adatbázis létrehozásakor hasznosulnak a jelen dolgozat megállapításai.

Irodalom

- BENCZE Gy.: Helyünk Európában. = Fizikai Szemle, 45. köt. 1. sz. 1995. p. 18-20.
- BENCZE Gy.: H-index: Egy új javaslat az egyéni tudományos teljesítmény értékelésére. = Magyar Tudomány, 167. évf. 1. sz. 2006. p. 88-91.
- BRAUN T., GLÄNZEL, W., NÉMETHNÉ KOVÁCS É., PERESZTEGINÉ SZABADI Zs.: Magyarország helyzete a természettudományi alapkutatás világában – tudományometriai tájkép a második évezred végén. = Magyar Tudomány, (47. (108.) évf.). évf. 7. sz. 2002. p. 935-945.
- BRAUN T., GLÄNZEL W., SCHUBERT A.: A Hirsch-type Index for Journals. = Scientometrics. 69. évf. 1. sz. 2006. p. 169–173.
- BRAUN T.: Szellem a palackból, tudományometriai értékelések. = Magyar Tudomány, 169. évf. 11. sz. 2008. p. 1366-1371.
- DOBROV, G. M.: A tudomány tudománya. Budapest, Gondolat-Kossuth, 1973.
- DRÓTOS L., KÖRÖS K., SOMOGYI T.: Adattármustra. Publikációs adatbázisok. = Tudományos és Műszaki Tájékoztatás, 55. köt. 3. sz. 2008. p. 127-145.
- DUDÁS A.: Tudományos önismeret: az egyetemi bibliográfiák szükségességéről és egy bölcsészettudományi oktatói publikációs adatbázisról. = Tudományos és Műszaki Tájékoztatás. 55. évf. 3. sz. 2008. p. 107-125.
- EGER A.: Hogyan írjunk világszínvonalú tanulmányt? Praktikus tanácsok publikáláshoz. = Tudományos és Műszaki Tájékoztatás, 56. évf. 7. sz. 2009. p. 307-319.
- GARFIELD E., SHER I. H.: New Factors in Evaluation of Scientific Literature through Citation Indexing. = American Documentation, 14. évf. 3. sz. 1963. p. 195-201.
- GARFIELD E., SHER I. H., TORPIE R. J.: The use of citation data in writing the history of Science. Philadelphia: Institute for Scientific Information, 1964, p. 75.
- GARFIELD E.: Citation analysis as a tool in journal evaluation. = Science, 178. évf. 1972. p. 471-479.
- GYERGYAI A.: „Az el nem ért bizonyosság” = Védelem az esszé ügyében. Budapest, Szépirodalmi könyvkiadó. 1984. ISBN 963 15 2452 3
- HIRSCH J. E.: An Index to Quantify an Individual’s Scientific Research Output. = Proceedings of the National Academy of Sciences of the United States of America, 102. vol. 2005. p. 16569-16572.
- HETI VÁLASZ: A száz legidézettebb természetkutató. = 8. évf. 2008. 01. 17. p. 12-14.

- HETI VÁLASZ: Tudósok a mérlegen. = 8. évf. 2008. 02. 21. p. 22-23.
- KISS A., VITÉZ G.: Miskolci Egyetemi Publikációs Adatbázis. = Tudományos és Műszaki Tájékoztatás, 53. évf. 7-8. sz. 2006. p. 321-327.
- KOLLÁR I., MICHELBERGER P.: Hozzászólás Tolnai Márton *Tudósaink mérhető teljesítménye az MTA Köztudományi Publikációs Adattár adatainak tükrében* című cikkéhez. = Magyar Tudomány, 169. évf. 10. sz. 2008. p. 1262-1273.
- KORN A., SCHUBERT A., TELCS A.: Lobby index in networks. = Physica A, 388. évf. 11. sz. 2009. p. 2221-2226.
- MARTON J.: Bibliometria. = Könyvtárosok kézikönyve 1. kötet. Budapest, Osiris Kiadó, 1999. p. 83-151. ISBN 963 379 646 6
- MARTON J.: Tizenkét futam az európai tudományometriai mezőnyben. Élettudományok. = Tudományos és Műszaki Tájékoztatás, 56. évf. 3. sz. 2009. p. 118-122.
- NALIMOV V. V., MULCSENKO Z. M.: Tudománymetria. Budapest, Akadémiai Kiadó, 1980. ISBN 963 05 2352 3
- RÁB P., KADLECOVÁ I. L.: Evaluations in the Academy of Sciences of the Czech Republic: past, present and future. Előadás. Bibliometric event Workshop, Budapest, MTA Könyvtára, 2009. nov. 2.
- REEDIJK J., MOED H. F.: Is the impact of journal impact factors decreasing? = Journal of Documentation, 64. vol. 2008. p. 183-192.
- RIES A., TROUT J.: A marketing huszonkét vastörvénye. Budapest, Bagolyvár Kiadó, 1995. ISBN 963 7423 50 8
- RÓZSA L., PAPP L.: A csoportmunka, mint egyéni érdem és egyéb megjegyzések a Köztudományi Publikációs Adattár használatáról. = Magyar Tudomány, 170. évf. 1. sz. 2009. p. 95-102.
- SCHUBERT A.: Tudománykommunikáció képzés. Előadás. Budapest, Magyar Tudományos Akadémia. 2007. okt. 3.
- Szakterületi statisztikák magyar kutatók Hirsch-féle h-indexeire az MTA Köztudományi Publikációs Adattárának adatai alapján. MTA-KSzi, 2005 november.
- TOLNAI M.: Tudósaink mérhető teljesítménye az MTA Köztudományi Publikációs Adattár adatainak tükrében. = Magyar Tudomány, 169. évf. 8. sz. 2008. p. 976-988.
- TOLNAI M., SCHUBERT A., WOLF Gy.: Tudományunk mérhető teljesítménye az Essential Science Indicators mutatószámainak tükrében. = Magyar Tudomány, 169. évf. 8. sz. 2008. p. 989-997.
- TOLNAI M., MOSONYINÉ FRIED J., SOÓS S.: Fialat kutatók az MTA köztestületeiben. =

Magyar Tudomány, 170. évf. 3. sz. 2009. p. 334-344.

TÖRÖK Á.: Reális-e a magyar tudomány 20. helye a (képzeletbeli) világranglistán? =

Magyar Tudomány, (45. (108.) évf.) 11. sz. 2000. p. 1307–1328.

VINKLER P.: Tudománymetriai kutatások Magyarországon. = Magyar Tudomány, 169. évf.

11. sz. 2008. p. 1372-1380.

ZÁDORI Zs.: Mit ér a tudós, ha magyar. Hvg.hu – hírek szünet nélkül, 2006.05.17.

<http://hvg.hu/itthon/20060516tudosmagyarmta.aspx>

Letöltés dátuma: 2009. december 1. 10.00

ZÁDORI Zs.: Hódít a teljesítménymérés: számnak ingere. Heti világgazdaság, 2006.10.14. p.

68-71.

ZOLNAI L., BERÉNYI D.: Kutatóintézeti tudománymetria – ATOMKI 1954–1989. = Fizikai

Szemle. 39. évf. 8. sz. 1989. p. 285–291.

ZOLNAI L., GÁCSI Z.: Mérünk, de mit? Egy formula margójára. = Magyar Tudomány (43.

(105.) évf.) 8. sz. 1998. p. 988-993.

ZOLNAI L.: Mennyire magyar a magyar? = Magyar Tudomány, (48.(108.) évf.) 12. sz. 2001.

p. 1497-1498.

ZOLNAI L.: Tudománymetria és intézeti kollaboráció. = Fizikai Szemle. 51. évf. 8. sz. 2001.

p. 264–265.

ZOLNAI L.: A „sötét teljesítmény” nyomában. = Magyar Tudomány, 169. évf. 7. sz. 2008. p.

870-874.

Jegyzetek

¹ <http://www.atomki.hu/p2/>

² A közelmúltban Drótos László és társszerzőinek tollából megjelent egy összehasonlítás, amely a Miskolci Egyetem publikációs adattárát (MEPA), a Szegedi Tudományegyetem publikációs adattárát (SZTPA) és a Magyar Tudományos Akadémia Köztudományi Központjának Publikációs Adattárát (KPA) dolgozza fel főleg programozás technikai (megjelenési felület, kereshetőség, stb...) szempontokból. Dolgozatunkban ezeket a szempontokat, csak vázlatosan tekintjük át, inkább a publikációs adattárak használhatósága, üzemeltethetősége és rendszerszervezése képezik vizsgálataink tárgyát.

³ U.o.:

⁴ DOBROV, G. M.: A tudomány tudománya. Budapest, Gondolat-Kossuth, 1973.

⁵ VINKLER P.: Tudományometriai kutatások Magyarországon. = Magyar Tudomány, 169. évf. 11. sz. 2008. p. 1372-1380.

⁶ TOLNAI M., SCHUBERT A., WOLF Gy.: Tudományunk mérhető teljesítménye az Essential Science Indicators mutatószámainak tükrében. = Magyar Tudomány, 169. évf. 8. sz. 2008. p. 989-997.

⁷ GARFIELD E., SHER I. H.: New Factors in Evaluation of Scientific Literature through Citation Indexing. = American Documentation, 14. évf. 3. sz. 1963. p. 195-201.

⁸ Impakt faktor/hatástényező: Egy folyóirat n -dik évi \sim -a 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. Valamilyen mértékben információt arra nézve, hogy mennyire használják a folyóiratban közölt információkat. Az \sim arra ad becslést, hogy egy folyóiratban megjelent átlagos publikációra a megjelenést követő két év átlagában hány, az adatbázisbeli hivatkozás várható.

⁹ ZÁDORI Zs.: Mit ér a tudós, ha magyar. Hvg.hu – hírek szünet nélkül, 2006.05.17.

<http://hvg.hu/itthon/20060516tudosmagyarmta.aspx>

Letöltés dátuma: 2009. december 1. 10.00 és

ZÁDORI Zs.: Hódít a teljesítménymérés: számnak ingere. Heti világgazdaság, 2006.10.14. p. 68-71.

¹⁰ Idézettségi félidő: Az idézettségi 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 cikkek 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.

¹¹ Aktivitási index: Egy adott időszakban, egy adott szakterületen, egy adott szerzőhalmaz által publikált cikkek száma a világban publikált cikkek számához képest. Az ~et pl. 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.

¹² 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.

¹³ Szakterületi statisztikák magyar kutatók Hirsch-féle h-indexeire az MTA Köztudományi Publikációs Adattárának adatai alapján. MTA-KSZI, 2005 november;

HETI VÁLASZ: A száz legidézettebb természetkutató. = 8. évf. 2008. 01. 17. p. 12-14.;

HETI VÁLASZ: Tudósok a mérlegen. = 8. évf. 2008. 02. 21. p. 22-23. és

TOLNAI M.: Tudósaink mérhető teljesítménye az MTA Köztudományi Publikációs Adattár adatainak tükrében. = Magyar Tudomány, 169. évf. 8. sz. 2008. p. 976-988.

¹⁴ SCHUBERT A.: Tudománykommunikáció képzés. Előadás. Budapest, Magyar Tudományos Akadémia. 2007. okt. 3.

¹⁵ RIES A., TROUT J.: A marketing huszonkét vastörvénye. Budapest, Bagolyvár Kiadó, 1995. ISBN 963 7423 50 8

¹⁶ ZOLNAI L., GÁCSI Z.: Mérünk, de mit? Egy formula margójára. = Magyar Tudomány (43. (105.) évf.) 8. sz. 1998. p. 988-993.

¹⁷ <http://www.doktori.hu/>

¹⁸ <http://www.mtakoztest.hu/>

¹⁹ DUDÁS A.: Tudományos önismeret: az egyetemi bibliográfiák szükségességéről és egy bölcsészettudományi oktatói publikációs adatbázisról. = Tudományos és Műszaki Tájékoztatás. 55. évf. 3. sz. 2008. p. 107-125.

²⁰ KISS A., VITÉZ G.: Miskolci Egyetemi Publikációs Adatbázis. = Tudományos és Műszaki Tájékoztatás, 53. évf. 7-8. sz. 2006. p. 321-327.

²¹ DRÓTOS L., KŐRÖS K., SOMOGYI T.: Adattármustra. Publikációs adatbázisok. = Tudományos és Műszaki Tájékoztatás, 55. köt. 3. sz. 2008. p. 127-145.

²² <http://www.atomki.hu/library/konyvtar.htm>

²³ <http://www.atomki.hu/p2/>

²⁴ <http://www.mtatpa.hu/>

²⁵ Globális kereshetőség: Ha egy web-lap tartalma fellelhető a világhálón létező keresőmotorok (Google, Yahoo, stb.) számára. Ehhez általában az szükséges, hogy az illető web-lap el legyen látva metaadatokkal (ld. Dublin Core).

²⁶ A címek és a megjelenési helyek is kapnak kódszámot, ami több feladat egyszerű megoldását – például a folyóiratok impakt faktoral való ellátását – megkönnyíti.

²⁷ SCI Publikáció: Ha a megjelenés évében a megjelenési helynek van impakt faktora (Országos Tudományos Kutatási Alap – OTKA definíció); SCI Hivatkozás: Ha az illető megjelenési helynek valaha volt impakt faktora.

²⁸ ZOLNAI L., BERÉNYI D.: Kutatóintézeti tudománymetria – ATOMKI 1954–1989. = Fizikai Szemle. 39. évf. 8. sz. 1989. p. 285–291.

²⁹ U.o. mint ²¹ és ¹⁸

³¹ Idegen rész: A saját résznek a fellelt publikáció, illetve hivatkozási számból való kivonásával adódik.

Saját rész: Egy szerző ~e 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. ~ről.

³² 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 alaputatás világában – tudománymetriai tájkép a második évezred végén. = Magyar Tudomány, (47. (108.) évf.). évf. 7. sz. 2002. p. 935-945.

; TOLNAI M., SCHUBERT A., WOLF Gy.: Tudományunk mérhető teljesítménye az Essential Science Indicators mutatószámainak tükrében. = Magyar Tudomány, 169. évf. 8. sz. 2008. p. 989-997. és

MARTON J.: Tizenkét futam az európai tudománymetriai mezőnyben. Élettudományok. =

Tudományos és Műszaki Tájékoztatás, 56. évf. 3. sz. 2009. p. 118-122.

³³ TÖRÖK Á.: Reális-e a magyar tudomány 20. helye a (képzeletbeli) világranglistán? = Magyar Tudomány, (45. (108.) évf.) 11. sz. 2000. p. 1307–1328.

³⁴ <http://www.sherpa.ac.uk/romeo.php>

³⁵ <http://www.mtak.hu/>

³⁶ <http://www.eprints.org/software/>

³⁷ RIS formátum: A Thomson Reuters cég Reference Manager nevű bibliográfiakezelő szoftvere által használt, a publikációk adatai import/export-jára szolgáló formátum szabvány³⁷.

³⁸ Dublin Core: A weblapok tartalmának leírására szolgáló metaadatok szabványa. A magyar gyakorlatban a MSZ ISO 15836 vezette be.

³⁹

http://mta.hu/index.php?id=634&no_cache=1&backPid=390&tt_news=11266&cHash=873f77c3c7