

**Debreceni Egyetem
Informatikai Kar
Könyvtárinformatikai Tanszék**

Tudományos publikációs adattárak

Magyarországon

Szakedolgozat

Konzulens:

Eszenyiné Dr. Borbély Mária PhD

Egyetemi adjunktus

Készítette:

Adománné Zolnai Dóra

Informatikus-könyvtáros szak

Debrecen

2009

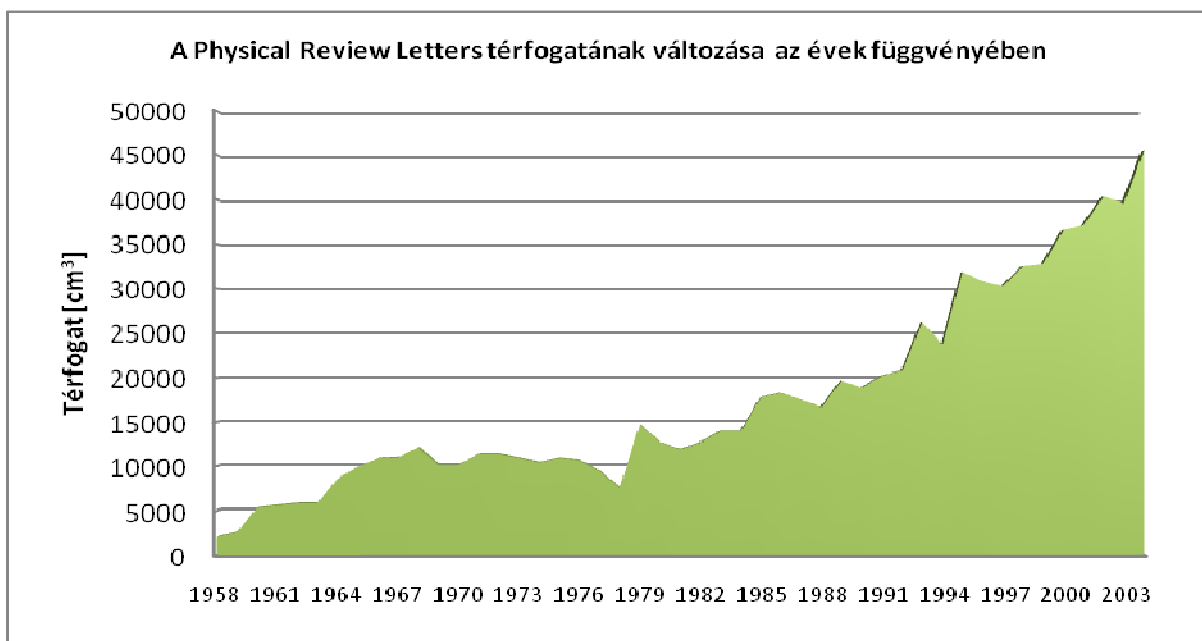
Tartalomjegyzék

1.0 Bevezetés	3.
2.0 Előzmények, történeti áttekintés	6.
2.1 Előzmények, áttekintés	6.
2.2 A publikációs adattárak létrejötte	11.
3.0 A vizsgált tudományos publikációs adattárak ismertetése	14.
3.1. A Magyar Tudományos Akadémia Tudományos Publikációs Adattára	14.
3.2. A Magyar Tudományos Akadémia Köztudományi Publikációs Adattára	16.
3.3. A Pázmány Péter Katolikus Egyetem Bölcsésztudományi Karának Oktatói Publikációs Adatbázisa	17.
3.4. A Miskolci Egyetem Publikációs Adatbázisa	18.
3.5. Az MTA Atommagkutató Intézet Publikációs-Hivatkozási Adatbázisa	19.
4.0 A dolgozatban használt fogalmak definíciói, alkalmazott vizsgálati módszerek	25.
4.1 Definíciók	25.
4.2 Vizsgálati eszközök és módszerek	30.
5.0 Táblázatos összehasonlítás	31.
6.0 Szöveges összehasonlítás, következtetések	36.
6.1 Az adatfolyamok összehasonlító elemzése	36.
6.2 Az „inaktív” szerzők esete	40.
6.3 Az azonos nevű szerzők problémája	41.
6.4 Sok társszerzős publikációkkal/hivatkozásokkal kapcsolatos problémák	42.
6.5 Önhivatkozások	44.
6.6 ISI/nem ISI, SCI/nem SCI	45.
6.7 A teljesítmények additivitásának kérdése – az egyéni/intézményi teljesítmény elkülönítése	45.
6.8 Tárgyszavazás lehetőségei	48.
6.9 Publikációs adattár és/vagy repozitórium	48.
6.10 Az adattárak működtetésének gazdasági háttere	49.
6.11 Következtetések és javaslatok	51.
7.0 Összefoglalás	54.
8.0 Köszönetnyilvánítás	56.
9.0 Irodalomjegyzék	57.
10.0 Függelék	60

1.0 Bevezetés

Napjaink egyik, talán nem is megértett közhelye a tudomány viharos fejlődése.

Ennek megértését számos megjelent tudományos közlemény próbálja elősegíteni, itt szemléltetésül csak egy ábrát közlök, amely a fizika tudomány legszínvonalasabb, a fizika minden ágában cikket közlő folyóiratának, a Physical Review Letters-nek térfogatváltozását mutatja a folyóirat 1958-as indulása óta (az utóbbi években nincs adat, mert könyvtárunk áttért az on-line előfizetésre). Tekintettel arra, hogy a folyóirat kezdettől fogva négy oldalas „Letter”-eket közöl, ez a szám valószínűleg jól tükrözi a területen megjelent publikációk számának növekedését. Az egyes tudományágak és témák növekedésének ütemében természetesen jelentős különbségek lehetnek, általánosságban azonban elmondható, hogy a természettudományok területén a cikkek száma 5-15 évenként duplázódik. [Marton, 1999]



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

Jelenleg a szakértői lektorálással kiadott folyóiratok száma 20 000 körül van. [Eger, 2009]

A tudományos-társadalmi hatásokon túlmenően a megjelenő tudástömeg elkerülhetetlenné tette:

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

Részben ezeknek az igényeknek a kielégítésére jöttek létre az úgynevezett publikációs adattárak, melyeknek első és talán legismertebb képviselője a Science Citation Index (SCI)

Intézményi publikációs nyilvántartások hazánkban a 60-as évektől léteztek. Az általam ismert legrégebbi, az MTA Atommagkutató Intézetének szélyukkártyás 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) (<http://www.atomki.hu/p2/>). Mára hazánkban a többé-kevésbé működőnek mondható adattárak száma mintegy húszra tehető, de számos próbálkozás van, amelyek feltöltöttség tekintetében kezdeti szinten vannak.

Dolgozatom motivációját az adja, hogy több hazai publikációs adatbázis fejlődése megtorpanni látszik, időszerű áttekinteni a hazai publikációs adatbázisokat, hogy a továbblépéshez irányokat adhassunk. Személyes motivációm az is indokolja, hogy munkahelyem (MTA Atommagkutató Intézet) publikációs-hivatkozási adatbázisának – technikai okok miatt – lényeges átalakításon kell keresztül mennie.

A közelmúltban megjelent egy összehasonlítás [Drótos et al., 2008], 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 programozástechnikai (megjelenési felület, kereshetőség, stb...) szempontokból.

Dolgozatomban a publikációs adattárak használhatósága, üzemeltethetősége és rendszerszervezése képezik vizsgálataim tárgyát, az előző szempontokat inkább csak vázlatosan tekintem át. Ennek oka, hogy véleményem szerint egy adatbázis kialakításakor fontosabb a szerkezet és feltöltöttség, mint a tetszetős, sokfunkciós külső megjelenítés,

ugyanis azt módosítani a későbbiekben is lehetséges, az egyszer már feltöltött pontatlan, hiányos adatokat azonban jóval nehezebb – gyakran kilátástalannak tűnő feladat – javítani, vagy az adatfeltöltőket kiegészítésükre rábírni.

A fent említett cikkben [Drótos et al., 2008] két adattáron (MEPA és MTA KPA) túl dolgozatomban foglalkozom az MTA Tudományos Publikációs Adattárával (ATPA) a Pázmány Péter Katolikus Egyetem Bölcsészettudományi Karának Oktatói Publikációs Adatbázisával (PPKE BTK OPA), és a Magyar Tudományos Akadémia Atommagkutató Intézetének Publikációs – Hivatkozási Adatbázisával, az APHA-val. Folyamatosan működő adatbázisokat választottam, amelyek reprezentálják a fellelhető típusokat. Így került bele az anyaggyűjtésbe és összehasonlításba a két nagy akadémiai adatbázis mellé egy egyetemi, általános gyűjtőkörű adatbázis illetve a bölcsészettudományi PPKE BTK OPA és az APHA Utóbbinak a részletesebb ismertetését személyes okokon kívül részben az indokolja, hogy bemutatása a hazai szakirodalomban nem történt még meg, másfelől pedig számos problémára vetít fényt az adatbázisok rendszerszervezésével kapcsolatban. A PPKE adatbázisa pedig egy jól működő példa kisebb bölcsészettudományi adatbázisra.

A 2. fejezetben áttekintem a publikációs adattárak kialakulásához vezető történeti előzményeket, a 3. fejezetben röviden ismertetem a vizsgálatba bevont adattárakat. A 4. fejezet a vizsgálatban alkalmazott fogalmakat, ill. a vizsgálat módszereit tartalmazza. Az 5. fejezet táblázatos összehasonlítást, a 6. szöveges értékelést és néhány következtetést és javaslatot tartalmaz. A dolgozatot a 7. fejezetbeli összefoglalás, köszönetnyilvánítás, irodalomjegyzék és függelék zárja.

2.0 Előzmények, történeti áttekintés

2.1 Előzmények, áttekintés

A tudományos munka minősítésének egyik igen fontos, talán a legfontosabb alapja a publikáció. Minden eredményt lehetőség szerint mihamarabb publikálni kell azért, hogy minél többen tudjanak róla és minél többen alkalmazzák az elért eredményeket. A folyóiratbeli publikálás igen korán, már 1665-ben megkezdődött. Különböző tudósok munkásságának listaszerű összegyűjtése még korábbi időkre tekint vissza. A kutatók eredményessége a publikációkon kívül azok idézettségével is jellemezhető. A hivatkozások gyűjtése újabb tevékenység. 1927-ben a Gross és Gross szerzőpáros végezte az első idézetelemzéses vizsgálatot, kémiai tárgyú folyóiratokkal kapcsolatban, az idézettségi adatok alapján fontossági sorrendet állítva fel. A 30-as évekre esik a publikált anyagokra való idézetek, hivatkozások feltérképezésének kezdete, egy nagyon egyszerű könyvtári probléma eldöntésére: nevezetesen milyen folyóiratokat kell rendelni, ha a beszerzésre rendelkezésre álló pénzüsszeg – mint mindig – korlátos. Igazán nagy léptékre ez a tevékenység a számítógép közhasználatba vonásával szélesedett ki és Eugene Garfield nevéhez fűződik.

A világon minden idézeti index alapja az amerikai Institute of Scientific Information által összeállított Science Citation Index, SCI (jelenleg WoK). Ennek különlegessége, hogy az idézeteket figyeli, 1960-tól jelenik meg, de 1945-ig visszamenőleg tartalmaz adatokat. Azóta kb. 7-9000 folyóiratot figyelnek úgy, hogy a cikkek irodalomjegyzékeit rendszeresen, standard formában beírják az adatbázisukba. 2009-től kezdve lehetőséget biztosítanak viszonylag nagyszámú konferencia kiadvány által tartalmazott hivatkozás keresésére is. Bár ez a forrásszám korlátozott, – ma a szakértői lektorálással kiadott folyóiratok száma 20000 körül van – eléggé sok ahhoz, hogy hosszú ideig az SCI (vagy ma: WoK) legyen a hivatkozások szisztematikus keresési forrása. Az SCI elsősorban a matematika és természettudományok területét fedi le. Kapcsolódó kiadványa a társadalomtudományok területén a Social Science Citation Index (SSCI), ill. a bölcsészterületekre kifejlesztett Arts & Humanities Citation Index (A&HCI). Hazánkban jelenleg az említett adatbázisok 1975 utáni anyaga a hozzáférhető. Az utóbbi években a WoK számos kiegészítő szolgáltatással bővült. Az impakt faktorokat közlő Journal Citation Report köteteivel (1975-től), idézeti térképekkel (2008-től), Hirsch-index [Hirsch, 2005] számításával. Közismert még a Medline (orvosi

területeken) és a Scopus (az Elsevier kiadóvállalat terméke) program, amelyek szintén lehetővé teszik az idézetek keresését.

A tudományos publikációs tevékenység összefüggéseinek tudományos vizsgálatára egy külön tudományág, a *tudománymetria* alakult ki. Erről a tudományágról a magyar nyelven olvasó először 1973-ban Dobrov: A tudomány tudománya [Dobrov, 1973], Derek de Solla Price 1979-ben [Solla, 1979] és Nalimov és Mulcsenko, 1980-ban fordításban megjelent művéből [Nalimov, 1980] tájékozódhatott. Körülbelül ebben az időben indultak a hazai tudománymetriai kutatások, melyeknek úttörője Braun Tibor, eredetileg vegyész végzettségű kutató volt, aki később megalapította a tudományág első nemzetközi folyóiratát, a hazai kiadású, angolul megjelenő *Scientometrics*-et. A hazai tudománymetriai kutatások legfrissebb áttekintését Vinkler Péter cikke tartalmazza [Vinkler, 2008]. Mind a gyűjtött információforrások kiválasztásában, mind pedig az információ értékének mérésében nagymértékben támaszkodunk a tudománymetriára.

A tudományág jelenlegi állása szerint további részterületeket különíthetünk el [Schubert, 2007]:

- A *strukturális* (szerkezeti) tudománymetria a (publikációk és hivatkozási) szerzői hálózatok viselkedésének törvényszerűségeinek feltárására alakult ki.
- A *dinamikus* (változó) tudománymetria a tudománymetriai statisztikai jellemzők időbeli alakulásából leszűrhető törvényszerűségeket vizsgálja.
- Az *evaluativ* (értékelő) tudománymetria a tudománymetriai statisztikai jellemzők és az egyes szerzők, szerzőcsoportok tudományos teljesítménye közti kapcsolatokat kutatja.

Dolgozatom tárgya leginkább összefüggésben ez utóbbival van, ezért a következőkben ismertetek néhány alapfogalmat. A pontos definíciókat a 4.1 fejezetben közlöm.

Minden bizonnyal történetileg a legelső teljesítményértékelési szempont a szerzők publikációinak tartalma volt. Az egyéni publikációs szerzőlisták hosszabbodásával, előtérbe került a publikációk számának figyelembe vétele (egyres méltatások, nekrológok ma is ezt használják). A tudománymetria kialakulásával, illetve az informatika fejlődésével vált gyakorlattá a hivatkozások számbavétele. Ettől az időtől kezdve számos tudománymetriai statisztikai jellemző (science indicator) definiálására került sor. Ezek közül számos megtalálható az APHA-ban, országokra vonatkozólag [Tolnai, 2008b].

Használatának népszerű volta miatt meg kell említeni az impakt faktort (impact factor) [Garfield, 1963]. Ezt eredetileg a folyóiratok minősítésére dolgozták ki, abból a célból, hogy a könyvtárakban, egy esetleges folyóirat-megrendeléskor melyik folyóirattól kell „megszabadulni”. Definíciója a következő: A folyóiratok impakt faktora annak a mértéke, hogy milyen gyakran idézik a tárgyévben a folyóiratnak a megelőző két évben megjelent "átlagos cikkét". Az impakt faktor segít megítélni a folyóiratok relatív fontosságát, különösen az ugyanazon szakterületen működő többi folyóirat között.

Az impakt faktort úgy számítjuk ki, hogy a folyóirat előző két évben publikált cikkeire a tárgyévben kapott idézetek számát elosztjuk a folyóiratban az előző két évben publikált cikkek számával. Az impakt faktor értéke függ az évtől és a kiszámításához felhasznált adatbázistól. Általában a WoK-on alapuló értéket fogadják el. Ezeket az értékeket az ISI az évenként megjelenő Journal Citation Reports kiadványban teszi közzé.

Az idő előrehaladtával az impakt faktort, illetve annak egy adott publikáció halmazra összegzett (esetleg szerzőszámmal beosztott) változatát kezdték el használni egy szerző, intézmény vagy ország tudományos teljesítményének jellemzésére [Zádori, 2006a-b]. Ebben az értelemben használják az általam vizsgált adatbázisok is az egyetemi adattárak kivételével. Ebben az időszakban a folyóiratértékelés számos mutatóját (idézési félidő, idézettségi félidő, frissességi index, stb.) kezdték a kutatói teljesítményértékelés területén is alkalmazni. Teret nyert olyan számszerű jellemzők alkalmazása is, mely a publikálási szokásokkal vannak kapcsolatban (aktivitási index, produktivitás, stb.).

A legutóbbi idők terméke az egyes szerzők tudományos teljesítményének jellemzésére kifejlesztett Hirsch-index (h-index) [Hirsch, 2005], [Bencze, 2006]. Ennek a definíciója a következő: egy szerző Hirsch indexe az a h szám, amelyre vonatkozólag a szerzőnek van legalább h darab publikációja, amelyekre legalább h darab hivatkozás történt. A Hirsch-index elég jól használható kiemelkedő teljesítmények jellemzésére, ahol az önidézetek számossága eltörpül. Azonban nem minden tudományos munkásság kiemelkedő, mint a relativitáselmélet felfedezése vagy egy rák elleni új gyógyszer. Az eredmények ennél szerényebbek is lehetnek. Ekkor, kisebb jelentőségű munkásságnál – amennyiben az önidézetek esetleg túlnyomóak -, valamint például több társszerzős publikációhalmaznál illetve kevés számú, de magas hivatkozási számú publikáció esetén az index használhatósága kérdéses. A Hirsch-index

értéke jelentősen eltér a különböző adatbázisokból számolt esetekben, leginkább elfogadott a WoK-on alapuló érték, amelyet 2008-tól a WoK is közöl.

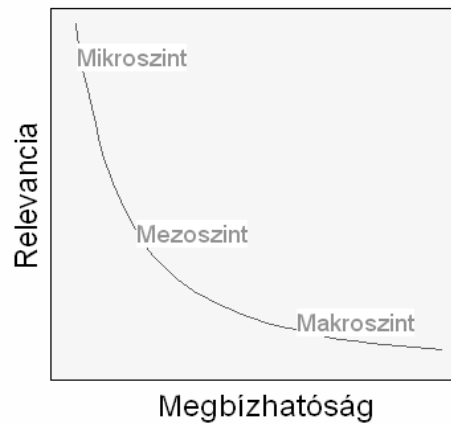
A Hirsch-indexet folyóiratokra is általánosították [Braun, 2006]. A Hirsch-index inspirálta többek között a társzerzői hálózatokra alkalmazható l-index (lobby index) kifejlesztését is [Korn et al., 2009], de vannak más „leszármazottjai” is.

A Hirsch-indexet hazánkban az elmondottaktól függetlenül elkezdték alkalmazni, ami a KPA hiányosan feltöltött adatain alapult [MTA KSzI, 2005], [Heti Válasz, 2008a-b], [Tolnai, 2008a].

Az evaluatív tudománymetria széles körben elfogadott álláspontja szerint a tudományos teljesítmény értékelésekor körültekintően kell eljárni, és minden esetben ki kell egészíteni úgynevezett *peer-review* értékeléssel. Különösen vonatkozik ez az idézettségi adatok elemzésére. Az idézettségi adatok nem hivatottak helyettesíteni a szakértői véleményt. Figyelmet kell továbbá fordítani azokra a tényezőkre, amelyek az idézettséget befolyásolják: a nyelvre, a formai és történeti tényezőkre, a megjelenési gyakoriságra, valamint a szakterületre. A nem angol nyelven és kiváltképp a nem latin betűs írással publikáló folyóiratok kevésbé hozzáférhetőek a világ kutatóközössége számára, és ez befolyásolhatja az idézettségüket. A vizsgálatok pontossága természetesen függ a használt adatbázisok pontosságától. Ugyancsak befolyásolhatja az idézettséget az idézés oka, amelyet már számosan, számos szakterületről elemeztek, Garfield-től Gyergyai Albertig.

Hazánkban a *peer-review* értékelés gyakorlatilag végrehajthatatlan, egyrészt az egyes szakterületeken dolgozók kis száma miatt, másrészt mert azok nem függetlenek. Különösen így van ez intézményi és egyéni szinteken. Még ha lennének is független szakértők, a tartalmi áttekintés az egyes esetekben 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 közti összevetés) vezet megbízható eredményre, mezo- (intézményi) és mikroszinten (egyéni teljesítményértékelés) alkalmazása nagy körültekintést igényel [Schubert, 2007]. Ezt szemlélteti a következő oldalon látható ábra:

Relevancia és megbízhatóság



2.1 ábra. Az evaluatív tudományometriai teljesítményértékelés relevanciájának és megbízhatóságának összefüggése az alkalmazási szintek szerint [Schubert, 2007].

Itt kell kitérnünk arra a problémára, hogy dacára annak, hogy egyes – és egyre több – tudományág nemzetközi kiterjedésű, a mezo- és mikroszinten történő értékelések szűkebb keretek között mennek végbe. Ez nyilvánvalóan azzal jár, hogy a mikroszinthez közeledve egyre több szubjektív szempont kerül alkalmazásra az értékelésekben. Ezzel is magyarázható, hogy a tudománymetria hazai művelői egyre inkább távol tartják magukat a hazánkban kampányszerűen történő értékelésektől.

Természetesen a számszerű jellemzők használatának is megvannak a korlátai, kiváltképp nem szabad őket abszolutizálni, amint az Magyarországon az utóbbi időkben szokásos (főleg, ahogy az impakt faktorról történt – aminek használata szakmailag is aggályos – jobb lenne az idézettség valamilyen formáját használni). Ez ellen a jelenség ellen Braun Tibor – a hazai tudományometriai kutatások úttörője – is felemelte szavát [Braun, 2008]. A kutatásértékelésnek a szűkebben vett evaluatív tudományometriai szempontoknál jóval szélesebb szempontrendszerre kell támaszkodnia [Zolnai, 2008].

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 tudományometriai adatokat esetleg egzisztenciális következményekkel járó következtetések levonására használhatják. A

viták általában az egyéni és csoportérdekek kifejezői, tudatosabb esetekben a *tudománypolitikai marketing* lépéseinek folyamányai. Ez utóbbiaknak rendszerbe foglalása a jövő kutatásainak témája lehet.

2.2 A publikációs adattárak létrejötte

A szerzők publikációinak bemutatására több irányból mutatkozik igény. A publikációs adattárak kimondva vagy kimondatlanul a láttatás és értékelés szándékával jöttek létre. Időben a legkorábbi amikor igény mutatkozott a kutatókban, oktatókban arra, hogy saját teljesítményüket megmutassák, reprezentálják, majd intézményük oldaláról is felmerült a szándék, hogy a falaikon belül elért eredményeket bemutassák. A tudományos kutatókat a modern társadalom tartja el, ezért is vetődött fel az elvárás, hogy rendszeresen adjanak számot teljesítményükről, bizonyítsák, hogy a kutatásra fordított támogatás nem hiábavaló. Ez a teljesítményértékelés alapvető fontosságú feladat. 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 bölcsészettudományok 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 találhatóak, nagy igény mutatkozik a tudományos – és nem ritkán egyéb – publikációk adatainak felmutatására. Különösen így van ez bizonyos bölcsészettudományi szakterületeken, de egyes orvostudományi témákban is. (A témáról bővebben: [Jehoda, 2006].) Ez az igény felismerése volt többek közt a dolgozatomban vizsgált valamennyi publikációs adattár létrehozásának az oka. Megjegyzendő, hogy a létrehozó szervezetek eltérő érdekei – és 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).

Új korszakot jelentett – és ez indokolja a viszonylag hosszú fejtegetést a 2.1 fejezetben – a tudományos teljesítmények értékelésének nemzetközileg, és a 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 [Zolnai, 1998] ld. a 2. függelékben) – volt a kiváltó oka az APHA web-en való megjelenésének. Egy majdani intézeti teljesítményértékelés kimondott vagy kimondatlan terve – amint az az adattár ismertetőjéből kitűnik („Az összegyűjtött adatok hozzájárulnak a szakirodalmi információs folyamatok törvényszerűségeinek feltárásához is.”) – vezetett az ATPA megalapításához, melyet szervezetenként 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 „résztvevői” az MTA-nak az 1993-as akadémiai törvény által való felduzzasztása óta) lehetőséget nyújtson munkásságuk nyilvánartására, a web-en való megjelenéséhez. Később ez az eredeti cél – vélhetőleg a gyér érdeklődés miatt – módosult. Egyre szélesebb körben és egyre erősebben ajánlották – attól függetlenül, hogy másutt már lehetséges munkásságuk figyelemmel kísérése – a KPA használatát ösztöndíjak, doktori cím, akadémiai tagság megpályázása esetén. Néhány éve az Országos Tudományos Kutatási Alap (OTKA) is a KPA adatait teszi ajánlottá, legújabbban pedig az Országos Doktori Tanács (ODT) adatbázisa (www.doktori.hu) is ezt részesíti előnyben.

A jelenlegi gyakorlat vegyes képet mutat. Számos helyen lehet publikációs listát találni, személyi, intézményi vagy több intézményt átfogó adatbázisokban, teljes vagy válogatott publikációs listákkal találkozhat az érdeklődő. A rendezési elvek úgyszintén változnak.

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.

Hasonlóképpen kevesen foglalkoznak a publikációk közvetlen hozzáféréseinek megteremtésével is, intézményi repozitórium kialakításával.

A hazai adattárak többsége a célcsoport szerzői kör publikációi adatainak közlésére szorítkozik. Egyes adattárak a publikáció teljes szövegét tartalmazó dokumentumok tárolására – keresésére (repozitórium) is módot adnak, bár szerzői jogi szempontok miatt a hozzáférést gyakran korlátozzák az intézményen belülre. Hasonló módon nem általános a publikációk alapadatai mellé más adatok (kulcsszavak, tárgyszavak, kivonatok, stb.) tárolása is.

Bonyolultabb esetben – és nem kis energiaráfordítással – a hivatkozások feltérképezése is lehetővé válik. Ennél is tovább mennek azok az adattárak, amelyek az adatokból tudományometriai statisztikai jellemzőket származtatnak, sőt ezek idősorait is megmutatják. A WoK bonyolultságát, és az általa nyújtott szolgáltatások színvonalát, a dolgozatomban vizsgált adattárak egyike sem éri el, a publikációk adatainak naprakészséget – igaz, kisebb méretekben – csak az APHA közelíti meg. Az ATPA és KPA egyre több tudományometriai statisztikai jellemzőt közöl, azonban az APHA anyaga számosabb. Az idézeteket tekintve rosszabb a helyzet, a vizsgált adattárak fél-két éves késéssel követik a WoK adatait. Ebben a tekintetben is az APHA tanúsítja a legkisebb lemaradást.

Az egyes adattárak jövőbeni tervei a fenntartó szerv megfogalmazott elvárásain túl, általában a fejlesztő informatikusok és kezelők fejében alakulnak ki az érdeküknek megfelelően. Sajnálatos, hogy a felhasználók igényei csak minimális mértékben kapnak teret. Ez utóbbira példa a túlságosan nehezen áttekinthető és kezelhető adatbeviteli felületek megléte. Valószínűleg a jövőbeni fejlődés gátja lehet a fejlesztésre fordítható források szűkös és szétaprózott volta.

A fenti fejtegetésekkel céloim az volt, hogy megpróbáljam bemutatni, hogy a hazai tudományos publikációs adattárak fejlesztésének kérdései milyen érdekek folytán váltak fontossá.

3.0 A vizsgált tudományos publikációs adattárak ismertetése

A 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 és kereshetőségét helyezik előtérbe, mások viszont főleg az szerzők és intézmények publikációs, egyes esetekben 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 ismertetem. A viszonylag hosszan előtárt APHA hosszabb ismertetését az indokolja, hogy bár az országban a legrégebbi, kevésbé ismert, a hazai irodalomban még nem került ismertetésre.

3.1. A Magyar Tudományos Akadémia Tudományos Publikációs Adattára (ATPA, MTA TPA)

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 adatok szolgáltatása és minősége az egyes intézetek igazgatóinak felelőssége. Az összegyűjtött adatok hozzájárulnak a szakirodalmi információs folyamatok törvényszerűségeinek feltárásához is.

Az intézetek adatai 1992-től, az akadémiai kutatócsoportok adatai pedig 1996-tól szerepelnek az adatbázisban. A 2007-es adatok feldolgozása után az Adattár több mint 410 ezer tévelt tartalmazott (110000 publikáció, 300000 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 publikáló 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 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. Szemléltetésül bemutatjuk a TPA Forrás program kezelői felületét, amelyhez a TPA Idézet program felülete nagyon hasonlít. A beviendő adatok leírása jól leolvasható az ábráról. A programok XML adatfájlokat állítanak elő.

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

3.1 ábra. Az ATPA-hoz tartozó, publikációk adatainak bevitelére szolgáló „TPA Forrás” nevű program kezelői felülete.

3.2. A Magyar Tudományos Akadémia Köztestületi Publikációs Adattára (KPA)

A KPA célja a 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. A KPA személyekhez kötődően gyűjti az adatokat, és kiterjed az érintett személyek munkásságának teljes időszakára.

A publikációs és idézettségi adatok eljuttatásának módja az adattárba: első lépésként telefonon, vagy e-mail útján történő bejelentkezés a KPA illetékes

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

Amennyiben az adatszolgáltatónak van 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 a KPA honlapján mindenki számára elérhetővé (www.mtakoztest.hu).

Megjegyzendő, hogy a KPA több klonja működik az országban. Így az MTA Kutatásszervezési Intézet (KSZI), MTA Kísérleti Orvostudományi Kutatóintézet (KOKI), Szegedi Biológiai Központ (SZBK), Budapesti Műszaki és Gazdaságtudományi Egyetem (BME), Semmelweis Egyetem (SOTE), Szent István Egyetem (SZIE). Ezek működése kis mértékben, de eltér a KPA-étól.

3.3 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 cikkében. [Dudás, 2008]

3.4. A Miskolci Egyetem Publikációs Adatbázisa (MEPA)

A *Miskolci Egyetemi Publikációs Adatbázist (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.

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. 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.

A Miskolci Egyetemi Publikációs Adatbázis célja, hogy egy helyen összegyűjtse, tárolja és visszakereshetővé tegye az egyetemen dolgozók tudományos publikációinak bibliográfiai adatait, valamint statisztikai adatokat szolgáltasson a különböző egyetemi beszámolók számára.

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 NDA-ban jelenleg elérhető, a Miskolci Egyetemen keletkezett publikációk listája az oldalról kiindulva megtekinthető. A publikációgyűjtemény segítségével, figyelemmel kísérhető az egyetem szerzőinek munkássága, az egyes tanszékeken, szervezeti egységeken belül folyó munka.

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 [Kiss – Vitéz, 2006] közleményben, bizonyos szempontok szerinti értékelése pedig a [Drótos, 2008] irodalomban.

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

A Magyar Tudományos Akadémia Atommagkutató Intézetét 1954-ben alapították Debrecenben. A korlátozott nyilvánosságú, gyűjtőkörét tekintve országos feladatkörű, tudományos szakkönyvtár az intézettel együtt jött létre. A könyvtár feladata elsősorban az intézet dolgozóinak szakirodalommal való ellátása, az intézet publikációs tevékenységének,

valamint a publikációkra történő független hivatkozások regisztrálása, az Atomki Annual Report terjesztése.

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 azelőtt is létezett publikációs nyilvántartás, először egyéni listák, majd központi széllyukkártyás nyilvántartás formájában.

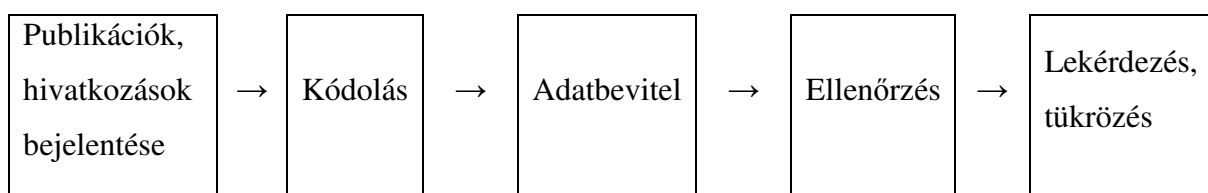
Az intézeti nyilvántartás – amely a 80-as évek eleje óta számítógépesítve van (PC/DOS-on) – meglete különösen nagy szolgálatot tett az akadémiai intézetek 1996-98-as konszolidációja idején, amikor főleg az intézetek tudományometriai mutatói alapján döntöttek az intézetek további sorsáról.

Jelenleg számítógép nélkül kilátástalan lenne a feladat. A feldolgozandó publikációk és hivatkozások száma egyre nőtt, főleg a rendszerváltozás után megnövekedett nemzetközi együttműködés miatt. Szemléltetésül: jelenleg az intézetnek évente ~ 1000 db írás- és szóbeli publikációja van, amiből a „kemény” írásbeli publikációk száma kb. 300. Mivel az intézetnek jelenleg ~ 100 kutatója van, ez azt jelenti, hogy az egy kutatóra 3,3 „kemény” publikáció/fő/év 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 gyűjti a DE „Fizikai Tudományok” Doktori Iskolája hallgatóinak publikációit és az azokra való hivatkozások adatait is.

Az adatbázis publikációk fajtáit tekintve, egyedülállóan a vizsgált adattárak közül, 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 ~42.000. Az adatbázissal kapcsolatos részletes statisztikák megtalálhatók a <http://www.atomki.hu/p2/> web-lapon.

A publikációk és hivatkozások feldolgozását, ill. az adatbázis működését szemlélteti a következő folyamatábra:



A publikációk és hivatkozások bejelentése háromféleképpen történhet:

- A függelékben mellékelt bejelentőlapok kézi kitöltésével, és a könyvtárhoz való eljuttatásával,
- A bejelentőlapok gépi kitöltésével és a könyvtárhoz e-mailen való eljuttatásával,
- On-line bejelentés a <http://www.atomki.hu/p2/> web-lapon keresztül.

Jelenleg a bejelentések, főleg a hivatkozásoké, túlnyomó része online történik. (A kézi bejelentést szeretnénk teljesen kiiktatni.) A tapasztalat azt mutatja, hogy a rendszer működésének kritikus pontjai a bejelentések. Ezért a számítástechnikai fejlesztések az utóbbi időben főleg ezek könnyebb kezelhetőségének biztosítására irányultak. Az emberi tévedések kiküszöbölésénél kevesebb a lehetőség. A kutatók által bejelentett adatok jelentős része, akár 20 százaléka is, valamilyen szempontból téves. Leginkább a szerzők neve és a publikáció címe bizonyul hibásnak, de nehézségek adódnak az egyéb bibliográfiai adatok (pl. kötetszám, terjedelem) kapcsán is. A hibák kiküszöbölését úgy próbáljuk megoldani, hogy bekérjük a cikk, könyv első oldalának fénymásolatát is, amelyen az adatok láthatók. Hogy nem egyedül küzdünk e problémával, bizonyítják a MEPA és a PPTÉ BTK OPA működéséről szóló [Kiss, 2006], [Dudás, 2008] cikkek is. Hasonló okokból fontos a könyvtár munkatársainak a kutatókkal való szoros együttműködése is. Összességében az a tapasztalatunk, hogy

nagyszámú csoport bibliográfiai ismereteiben illetve precizitásában bízni a bejelentések során nem szerencsés, mindenképp ellenőrzésre van szükség.

A *Kódolást* – amely alatt a címek, szerzők és a megjelenési helyek egy numerikus kódot kapnak – megosztva végzik a könyvtárosok, egymás munkájának keresztellenőrzésével. Ebben a munkaszakaszban az ugyanolyan nevű, de személyileg különböző szerzők okozzák a legnagyobb problémát. A probléma megoldása nem teljes körű, elsősorban az intézeti és magyar szerzők körében törekednek a megfelelő kódolásra. Ehhez segítséget kapnak egy, az intézetben 30 éve dolgozó fizikus munkatárstól, aki meglehetősen jól ismeri a munkatársakat, illetve az intézet publikációs történetét. Külön nehézséget jelentenek a távol-keleti szerzők nevei.

Az *Adatbevitelt* jelenleg mindig ugyanaz a könyvtáros végzi. Így biztosítva van a bevitt adatok helyességéért viselt felelősség.

Az *Ellenőrzéshez* a nagyobb csomagokban bevitt adatokról nyomtatott ellenőrző lista készül, amely az eredeti bejelentőlapok illetve a bejelentéssel párhuzamosan a publikáció első oldaláról/oldalairól bekért fénymásolatok tartalmával kerül összehasonlításra. Amennyiben eltérés van, a hibás adat javításra kerül. Évente egyszer a kutatók korrektúrázzák az év folyamán összegyűjtött rekordokat. Ekkor nem csak javításra, de hiánypótlásra is sor kerül.

A *Lekérdezés* az adatbevitel során többször történik:

- Annak megállapítására, hogy a szóban forgó bejelentés benne van-e már az adatbázisban,
- Kutatók kérésére, listák készítésére,
- Ellenőrző listák készítésére.

A folyamat ugyan külső szemlélő számára többszörösen túlbiztosítottnak tűnhet, a tapasztalat azonban azt mutatja, hogy egy, számunkra elégségesen pontos adatbázis feltöltéséhez ezekre a lépésekre (keresztellenőrzések és egyéb ellenőrzések) szükség van. Különösen igaz ez akkor, ha az adatokból tudományometriai mutatókat kívánunk előállítani illetve egyéb összefüggéseket feltárni, nem csak egyes kutatók publikációs listáira vagyunk kíváncsiak. Gondoljunk bele, egy konferencia nevének (az egyik legjellemzőbb bejelentési hiba, gyakorlatilag ahány bejelentő, annyi konferencianév) többféle nyilvántartása pl. azt

eredményezheti, hogy nem tudjuk teljességében lekérdezni, intézetünkben kik tartottak rajta előadást, így jelentős tévedésekre vezethet.

A *Tükrözés* a lekérdezéshez hasonló folyamat, az adatbázis tartalmának az interneten olvasható formában való leképezését jelenti. Gyakorlatilag egy másik számítógépen futó FORTRAN program elkészíti a leggyakrabban használt rendezett lekérdezések (intézeti év szerinti listák, egyéni publikációs listák, téma szerinti publikációs listák) eredményeinek megfelelő Web-lapokat. Ugyanez a programrendszer gondoskodik a web-lapok globális kereshetőségét biztosító Dublin Core szabvány szerinti metaadatokkal való ellátásáról, valamint a tudományometriai statisztikai jellemzők kiszámításáról (3. Függelék), illetve azok idősorainak lekérdezésre való előkészítéséről. Ebben a munkafolyamatban történik számos, nem a könyvtárban rögzített adat (a szerzők intézeti hovatartozásának, a társszerzők számának, az esetleges együttműködési neveknek, „-ig” oldalak, füzetszám, oldalt jelölő betűjelzések, impakt faktorok (hatástényezők) feltüntetése stb.) csatolására és a szükséges tudományometriai jellemzők kiszámítására, továbbá a listákon található számlálók működtetésével az egyes lapok látogatottságának figyelésére. A listák elkészítését a már említett fizikus kolléga végzi, kb. heti rendszerességgel, de szükség esetén ennél gyakrabban is. Az elkészített listákon túlmenő lokális kereshetőség csak belső használatra biztosított. A Web-es anyag megtekinthető a <http://www.atomki.hu/p2/> web-lapról kiindulva.

Az adatbázis a napi használaton kívül arra is szolgál, hogy különböző, egyéb programok felhasználásával adatokat szolgáltat, az erre a célra Perl programozási nyelven írt programok segítségével:

- Az intézet vezetésének különböző vezetési döntések meghozatalához,
- Az MTA vezetésének az intézet előző évi teljesítményéről (februárban),
- Részletes publikációs-hivatkozási adatokat szolgáltat az MTA központi Tudományos Publikációs Adattárának (<http://www.mtatpa.hu/>) (március-május).

Az adatbázis egyik előnye – aminek fenntartására nagy erőfeszítések történnek – annak viszonylagos naprakészsége (ez többnyire a bejelentéstől számított egy hetet jelenti). Ugyancsak a naprakészség biztosítására a könyvtár munkatársai napi rendszerességgel figyelik az intézet publikációinak megjelenési helyéül szolgáló irodalmat, ill. a munkatársak

konferenciákon való részvételét. Adott esetben felhívják a munkatársak figyelmét a bejelentésre.

Másik előnye a pontossága. Az írásbeli publikációk esetében a bejelentéssel egy időben a publikációk első oldaláról fénymásolatot kell benyújtani (könyvek, konferencia kiadványok esetén a fedőlapról, illetve mindazokról az oldalakról is, ahol a műre és a szerzőkre vonatkozó adatok találhatóak) abból a célból, hogy a feltöltött adatok minél pontosabbak és visszakereshetőek legyenek. (fentebb említettük, gyakran pontatlanok a bejelentések) Év végén az előző évben begyűjtött adatok még egyszer ellenőrzésre kerülnek. A publikációk esetében ilyenkor kb ~10 % az elmaradt publikációk – amelyeket valamilyen okból nem sikerült az év folyamán fellelni – aránya. Ugyanez a hivatkozásokra vonatkozólag márciusban történik, ilyenkor sikerül az előző évben megjelent független hivatkozásoknak mintegy 2/3-át begyűjteni a WoK, Scopus, SPIRES (ez a nagyenergiájú fizika Stanford Linear Accelerator National Accelerator Laboratory által létrehozott publikációs adatbázisa) és az Ovid segítségével.

Kiemelendő, hogy az adatbázis az evaluatív tudományometriai eszközökön túlmenőleg számos, a dinamikus és strukturális tudományometriai vizsgálatokat elősegítő eszközt szolgáltat. 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”). A számolt tudományometriai statisztikai jellemzőkről további tájékoztatás található a 3. függelékben.

4.0 A dolgozatban használt fontosabb fogalmak definíciói, alkalmazott vizsgálati módszerek

4.1 Definíciók

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.

Átlagos idézettség Egy cikkhalmaz ~e egyenlő a cikkhalmazra történt hivatkozások száma osztva a halmazt alkotó cikkek számával. Beszélhetünk egy cikk idézettségéről, vagy egy bizonyos szerzőkörhöz tartozó cikkek ~éről.

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

Elsődleges/másodlagos adatforrás A publikációs adattár feltöltése közvetlenül a megjelent művek alapján történik. Amennyiben nem így van, másodlagos adatforrásról beszélünk.

Feltöltöttség (szerzői, idő) Annak százalékos becslése, hogy a szerzői, vagy időbeli gyűjtőkörben levő publikációk és/vagy hivatkozások hányadrészét tartalmazza a szóbanforgó publikációs adattár egy adott időpontban.

Független hivatkozás Olyan hivatkozás ahol a hivatkozott és a hivatkozó cikkeknek nincs közös társszerzője

Független hivatkozási szám Összes idézet mínusz önidézet

"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.

A frissességi index hasznos a folyóiratok idézettségi gyorsaságának összehasonlításában. Mivel egy cikkre vonatkozó átlagérték, a nagy folyóiratokat nem részesíti különösebb előnyben. A gyakran megjelenő folyóiratok viszont előnyt élveznek, mert az év elején megjelenő cikkeknek nagyobb az esélyük, hogy az év hátralévő részében

idézzék őket. Különösen a kutatás frontvonalában tevékenykedő folyóiratok összehasonlításában lehet a frissességi indexnek hasznos szerepe.

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

Gyűjtőkör (szerzői, időbeli) A szerzők azon halmaza, akiknek a publikációit és/vagy hivatkozásait a szóbanforgó publikációs adattár gyűjti. Az időbeli gyűjtőkör azon megjelenési évek halmaza, amelyekben megjelent publikációkat a publikációs adattár gyűjti.

Hirsch-index, h-index Egy szerző Hirsch-indexe az a h szám, amelyre vonatkozólag a szerzőnek van legalább h darab publikációja, amelyekre legalább h darab hivatkozás történt. A különböző publikációs és hivatkozási gyakorlat alapján a h -index értéke egyes tudományterületeken lényegesen különböző lehet. Jelentős tudományos teljesítményre dolgozták ki, kevésbé kiemelkedő kutatóknál befolyásolhatja az önidézés mértéke pl. Korlátozottan alkalmas kevés de jelentős publikációval rendelkező kutatók teljesítményének mérésére.

Idegen rész A saját résznek a felőlelt publikáció, illetve hivatkozási számból való kivonásával adódik.

Idézettségi félidő Az idézettségi félidő a tárgyévtől visszafelé számított annyi év, amennyiben megjelent cikkeire a folyóirat a tárgyévben kapott összes idézeteinek felét kapta. Ez az érték segítséget nyújt annak megítélésében, hogy többségükben milyen korú cikkeit idézik a folyóiratnak. Az idézettségi félidő kisebb vagy nagyobb számértéke nem jelez semmiféle tényleges értékítéletet. Egy elsőközlő tudományos folyóirat idézettségi félideje pl. általában nagyobb, mint egy kurrens információkat tartalmazó gyorsközlő folyóiraté. Az idézettségi félidő hasznos lehet a gyűjteményépítési és archiválási döntésekben. Az idézettségi félidő drámai változásai a folyóirat formai változását jelezhetik. Az idézettségi félidő összehasonlító vizsgálatával a formai és publikációtörténeti különbségeket lehet feltárni.

Idézési félidő Az idézési félidő a tárgyévtől visszafelé számított annyi év, amennyiben megjelent cikkekre a folyóirat tárgyévi hivatkozásainak fele vonatkozott. Ez az érték segítséget nyújt annak megítélésében, hogy többségükben milyen korú cikkekre

hivatkozik a folyóirat. Az idézési félidő drámai változásai a folyóirat formai változását jelezhetik. Az idézési félidő összehasonlító vizsgálatával formai különbségeket lehet feltárni.

Impakt faktor (impact factor, hatástényező) Egy folyóirat n -dik évi \sim a egyenlő az illető folyóirat $n-1$ -dik és $n-2$ -dik évfolyamában megjelent cikkekre való $n-1$ és $n-2$ évben megjelent (minden, az adatbázisban levő) hivatkozások száma osztva az ezekben az években az illető folyóiratban megjelent összes cikk számával. Valamilyen mértékben információt ad arra nézve, hogy mennyire használják a folyóiratban közölt információkat. (Nem minden magas impaktfaktorú periodikában megjelent cikk kap hivatkozást, a 32 körüli impaktfaktorú Nature-ben pl. a cikkek ötödének nincs hivatkozása.) Az \sim arra ad becslést, hogy egy folyóiratban megjelent átlagos publikációra a megjelenést követő két évben átlagosan hány hivatkozás várható. Az \sim értékének finomítására több lehetőséget is kifejlesztettek. (ilyen pl. az Index Copernikus, amely orvosi folyóiratokat rangsorol)

Az alábbi négy tényező befolyásolhatja a folyóiratok impakt faktorát:

- **Publikációtípus.** A publikált cikkeket az ISI dokumentumtípus szerint egyenként kódolja, de a hivatkozások évente feldolgozott millióit nem lehetséges egyenként kódolni. Ezért a JCR-ben található idézetszámok nem tesznek különbséget az eredeti közleményekre, összefoglaló cikkekre, kivonatokra vagy levelekre kapott idézetek között, jóllehet a JCR publikációs számai csak az eredeti közleményekre és összefoglaló cikkekre vonatkoznak. Ha egy folyóirat egy évben nagyszámú előadáskivonatot, levelet, stb. közöl, ez átmenetileg megnövelheti a kapott idézetek számát. Ez a növekedés a cikkek számában nem jelentkezik. Az ilyen jelenség felismerése és értékelése csak részletes cikkenkénti idéztelemzés segítségével lehetséges.
- **Formátumváltás.** Egy folyóirat hirtelen terjedelmváltozása befolyásolhatja az impakt faktort. A cikkenkénti átlagos idézetszám csökken, ha több egyéves cikk van, mint kétéves, ugyanis az idézettség általában a második évben tetőzik. Hasonlóképpen hirtelen terjedelmcsökkenés átmeneti impakt faktor növekedést eredményezhet. Az impakt faktor kiszámításához használt publikációk száma a JCR-ben megtalálható, ennek figyelésével ezek a hatások ellenőrizhetők.
- **Címváltozások.** A címváltozás utáni első évben a JCR az új címet impakt faktor nélkül tünteti fel, hiszen az impakt faktor kiszámításához szükséges előző két éves össz cikkszám nulla. A régi cím a szabályszerűen kiszámított impakt faktoralal található

meg. Egy évvel később a JCR-ben külön impakt faktora lesz a régi és az új címnek. Ebben a második évben az új címnek a várhatónál kisebb, a réginek a várhatónál nagyobb impakt faktora lesz, mert az új cím impakt faktora újabb, a régié régebbi cikkek idézettségén alapul. Az egyesített impakt faktort úgy számíthatjuk ki, hogy a két évre vonatkozó idézetek összegét elosztjuk a két év publikáció számainak összegével.

- **Csak idézett folyóiratként nyilvántartott folyóiratok.** Vannak olyan folyóiratok, amelyeket a JCR idéző folyóiratként nem, csak idézett folyóiratként tart nyilván. Ezt a folyóiratok összehasonlítása során figyelembe kell venni, mert az ilyen folyóiratok önidézeteit a JCR nem tartalmazza. Az önidézetek pedig a folyóiratok által kapott idézetek jelentős százalékát tehetik ki. A csak idézett folyóiratok egy része megszűnt, átmenetileg szünetelő vagy megváltozott című folyóirat.

JCR

Journal Citation Report

Lokális kereshetőség

Az adatbázis tartalmának általában dinamikus web-lapok segítségével történő lekérdezhetősége.

Kiemelkedő idézettségű publikáció

Olyan publikáció, amelynek idézettsége a megjelenést követően egy adott T időintervallumban egy bizonyos idézettségi számértéket N -szeresen meghaladta. Szokásosan: $T=2$ év; $N=10$.

Kooperativitás I.

A publikációknak az idegen része osztva a saját résszel.

Kooperativitás II.

Egy hivatkozáshalmaz száma osztva a halmaz valamilyen szempontból definiált saját részével. Azt adja meg, hogy a hivatkozáshalmaz lényeges részét adó publikációkat átlagosan hány társzerzővel publikálták.

Önhivatkozás I.

Olyan hivatkozás, amely olyan publikációra történik, ami ugyanabban a folyóiratban jelent meg, vagy egyidejűleg jelenik meg. Ilyen a hivatkozások mintegy 20 %-a. Ezt a definíciót a JCR használja.

Önhivatkozás II.

Olyan publikációban szereplő hivatkozás amely esetében a hivatkozott és hivatkozó cikk szerzői halmaza átfed.

Peer-review

Az adott publikációs anyagnak a szakterületen jártas, független szakértők által való áttekintése, értékelése

Produktivitás	Valamilyen szempontból definiált saját rész osztva az eredmény eléréséhez felhasznált évek számával. Ennek használatban van az egy szerzőre eső része is.
Repozitórium	A dokumentumok (publikációk) teljes szövegű változatát tartalmazó adatbázis.
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. Lásd még: idegen rész.
Referált folyóirat I.	Amelyben a közlési folyamatban szakértők (referre-k) működnek közre.
Referált folyóirat II.	Amelyik cikkeit a referáló folyóiratok (Pl. Physics Abstracts) indexelik.
Referált folyóirat III.	Amelynek van impakt faktora.
Relatív idézettség	Egy cikkhalmaz átlagos idézettségének aránya egy másik cikkhalmaz átlagos idézettségéhez képest.
RIS formátum	A Thomson Reuters cég <i>Reference Manager</i> nevű bibliográfiakezelő szoftvere által használt, a publikációk adatai import/export-jára szolgáló formátum szabvány [RIS].
SCI	Science Citation Index, az amerikai Institute of Scientific Information (ma a Thomson cég) által összeállított Science Citation Index (SCI) (a weben ma elérhető formáját Web of Science-nek, később Web of Knowledge-nek nevezik) [WoK].
SCI publikáció, SCI hivatkozás	Olyan folyóiratban megjelent cikk, vagy hivatkozás, amelyet az SCI impakt faktor táblázata tartalmazza a megjelenési évre vonatkozólag. Az Atomki az SCI hivatkozást úgy definiálja, hogy a hivatkozás folyóiratának valaha is volt impakt faktora.
Statisztikai jellemző	Science indicator
Teljes impakt	Egy publikációhalmaz n -dik évi ~ja az egyes folyóiratokban megjelent cikkek számának a folyóiratok n -dik évi impakt faktorával súlyozott

összege. Pl: egy intézmény ~ja. A ~ arra ad becslést, hogy egy bizonyos cikkhalmazra a megjelenést követő két évben hány hivatkozás várható.

WOK: Az SCI weben elérhető formája.

4.2 Vizsgálati eszközök és módszerek

Vizsgálataim során az előző alfejezetben definiált fogalmakat használtam.

A vizsgálat eszközei/módszerei a következők voltak:

- A leggyakrabban alkalmazott módszer a vizsgált publikációs adattárak webes felületein való tájékozódás, kipróbálás volt. Ennek elmélyítésére szolgált az a módszer, hogy az egyes adattárakban felhasználói azonosítót/jelszót kértem. Így, az egyébként nyilvánosság előtt elzárt tulajdonságokról is sikerült képet kapni.
- Az adattárak nem mindegyike rendelkezik friss adatokkal az adattárak állapotának számszerű jellemzőiről. Ilyen esetekben általában 10 próbálkozásból álló mintavétel alapján becsültem meg az 5. fejezetben közölt adatokat.
- Sokat segített munkámban a vonatkozó hazai szakirodalom tanulmányozása.
- Végül, a még homályos kérdéseket tisztázandó, de a működtetéssel – üzemeltetéssel kapcsolatosan is, személyesen felvettem a kapcsolatot az egyes adattárak kezelőivel. Ugyanezen a módon szereztem – igaz hézagos – információt az egyes adattárak jövőbeli terveit illetően.

5.0 Táblázatos összehasonlítás

A következő táblázatokban összehasonlítjuk a dolgozatban vizsgált adattárak tulajdonságait. A használt fogalmakat a 4. fejezetben írtuk le részletesen. A vizsgálat a 2009. szeptember-október havi időszakban történt.

Teljes név	Magyar Tudományos Akadémia Tudományos Publikációs Adattár	Magyar Tudományos Akadémia 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 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?cik=602	http://www.atomki.hu/p2/
Létrehozás dátuma	2001	2003	2005	2002	1986
Platform	mySQL, PHP	mySQL, PHP	mySQL, PHP	T-Series HTML	FORTAN-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
Céldőszak	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	~14000 [To09]	~1500	~250 (eddig 212 szerző)	~500

használt rövidítés	(MTA TPA, ATPA)	(MTA KPA)	(MEPA)	(PPKE BTK OPK)	(APHA)
Gyűjtőkör	Publikációk és hivatkozások Az MTA kutatóhelyek tudományos és ismeretterjesztő 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, tankönyve	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.
repozitóriumba való törekvés	igen	igen	igen	igen	igen
Angol nyelvű felület	részben	részben	nincs	nincs	igen

használt rövidítés	(MTA TPA, ATPA)	(MTA KPA)	(MEPA)	(PPKE BTK OPK)	(APHA)
Feltöltő személye	Kutatók, oktatók Önálló adatbázissal rendelkező intézmények.	Kutatók, oktatók	Bejelentés alapján a könyvtár dolgozói, az egyetemi oktatók sk.	Bejelentés alapján a könyvtár dolgozói	Bejelentés alapján a könyvtár dolgozói
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	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.

használt rövidítés	(MTA TPA, ATPA)	(MTA KPA)	(MEPA)	(PPKE BTK OPK)	(APHA)
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
Tudományometriai jellemzők	Számos	Számos	Alig	Néhány	Számos Idősoros is
látogatottsági statisztika	nincs	nincs	Van (böngészés, gyorskeresés, összetett keresés)	Időnként, manuálisan készítik	Van (honlapon elhelyezett számláló)

6.0 Szöveges összehasonlítás, következtetések

Az előző fejezetbeli táblázat sokféle szempontból módot nyújt a dolgozatban vizsgált adatbázisok összehasonlítására. Ezek teljességére való törekvés nélkül én itt csak a számomra, a továbblépés szempontjából kulcsfontosságú kérdésekre térek ki. A kérdések egy része túlmegy az egyszerű „adattári” problematikán, fontos voltak akkor válna jelentőssé, ha az adattárak anyaga és/vagy a belőlük származtatott tudományometriai statisztikai adatok alkalmazásra kerülnének az egyes szervezetek által végrehajtott értékelésekben.

6.1 Az adatfolyamok összehasonlító elemzése

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

A nemzetközileg elsősorban számon tartott publikációs adatbázisoknál (WoK, 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, köztisztviselők 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 köztisztviselők 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 köztisztviselők anyagának felvitelkor 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 régebről vezetett publikációs-hivatkozási listája, a feltöltöttség pedig sokszor hiányos. A MEPA (egy elvetélt próbálkozást – RESEDA – követően) felhagyott a hivatkozások gyűjtésével. 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:

- Adatbevitel pontossága. Nyilvánvaló követelmény az adatbázisokat illetően, az azokban tárolt adatok pontossága. Ebből a szempontból legjobb az elsődleges adatforrások használata, az autopszián alapuló adatbevitel, 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. Az adatbevitel különböző módokon történik: manuálisan az autopszia elve alapján, vagy a bejelentő másodlagos adatai alapján vagy pedig automatikusan, más adatbázisokból nyert adatok alapján. A következő táblázatban összegyűjtöttük a legjellemzőbb bejelentési hibákat. A táblázat kialakításakor Dudás Andrea [2008] cikkére támaszkodtam. Látható hogy két eltérő tudományterület bejelentési gyakorlata, bölcsész és természettudomány, sok hasonló problémát vet fel. A két adatbázis különböző utat követ a hibák kiküszöbölésére. A PPKE BTK Adatbázisa adatbázisok használatával és a szerzők bevonásával korrigálja az adatokat, az APHA pedig a publikációk első oldalát (oldalait) bekérve, azok segítségével végzi a pontatlan adatok kiszűrését. Mindkét rendszer hangsúlyt fektet a bejelentőkkel való szoros kommunikációra. A KPA a nyilvánvaló ellentmondásokat szűri, a MEPA nem végez ellenőrzést a feltöltött adatokkal kapcsolatban, de a hiányos adatokat pótolni igyekszik. [Kiss – Vitéz 2006]

Adatelem	Adatközlők (a beérkező jegyzékeken szereplő közlésmód)	PPKE	APHA
Szerző	kifelejtik a társszerzőket	X	X
	rövidítve adják meg a neveket	X	
	hibásan adják meg a nevet, keresztnév betűje lemarad	X	X
	fontosabb közreműködőkről nem nyújtanak adatot (pl. a gyűjteményes kötet szerkesztője)	X	
Cím	írásmű címe pontatlanul szerepel	X	X
	rövid gerinc cím (vagy önkényesen rövidített cím) a címdalon szereplő, hosszabb főcím helyett	X	X
	kézirat cím vagy munkacím szerepel, amely különbözik a későbbi véglegesített, a publikáción megjelenő címtől	X	X
	az alcímet nem tüntetik fel, holott az kifejezőbben utal a tanulmány tartalmára, mint például egy "irodalmias" főcím	X	
	elmarad a bibliográfiai gyakorlatban szokásos konferenciaadatok megadása az alcímben	X	X
	hibás vagy hiányos a konferencia címe		X
	műcím, kötet cím, alcím, sorozat cím, folyóirat cím keveredik; nem különböztethető meg a közölt adatsor alapján a rész-egész vagy a médiatípus	X	
	a címfordítás van megadva, nem pedig a valódi cím az eredeti közlés nyelvén (műcím, folyóirat cím stb.), a publikáció így csak nagy nehézségek árán azonosítható	X	
Egyéb adatok	nincs feltüntetve a kiadó	X	X
	az idegen nyelvű, illetve a külföldön kiadott könyv kiadójának neve a magyar változattal van megadva	X	
	idegen nyelvű címmel párosuló szerzőségi közlés az "ed.", "red.", "hrsg.", "transl.", "trad." stb. közreműködői minősítést vagy más partikulák magyarra fordítva "szerk.", "ford." stb. formában szerepel, a bibliográfiai gyakorlatban honos tükrözhető módszer helyett	X	
	kötetszám, füzettség, oldalszám adatok hiányoznak		X
	kötetszám, füzettség, oldalszám adatok keverednek		X
	terjedelem megadása pontatlan, gyakran füzet számmal keveredik		X

6.1 ábra. A publikáció bejelentések problémái a Pázmány Péter Katolikus Egyetem Bölcsészettudományi Kar Oktatói Publikációs Adatbázisának és az APHA adatbázisának tapasztalatai alapján.

- Adatbeviteli felület. 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 felületéről is. A legbonyolultabb adatbeviteli felülettel (valószínűleg a fejlesztésre fordított nagyobb befektetés miatt) a KPA rendelkezik, ami 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. Utóbbi úgy oldja meg a közelséget, hogy intézeti szinten közbeiktat olyan egységeket (rendszerint a könyvtárakat), akik 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.

- Humán tényezők – motiváltság. A teljességet, 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 – kiemelendő hogy nem nagy szerzőszámot felelő esetben – az APHÁ-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.

- Szerzői jog. 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.

6.2 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 publikációk területén jó példa 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.

Az intézményi érdekek felől tekintve a problémát, könnyen 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 az esetek egy részében valamelyest enyhíti, hogy a társszerzői struktúra abba az irányba tolódik el, hogy egyre többen vesznek részt a publikációkban azonos munkahelyről.

A jelenség érzékeltetéséül 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 ezekben az esetekben – a publikációk létezésének ismeretében – az adatbázis kezelőinek felszólítására jelentették be.

6.3 Az azonos nevű szerzők problémája

A publikációs adattáraknak nem csak 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 szerzők között. Ellenkező esetben téves bibliográfiai tételek kerülhetnek az elemzésbe. 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 kódolnak (a címek és a megjelenési helyek is kapnak kódszámot, ami több feladat egyszerű megoldását – például a folyóiratok impakt faktorról való ellátását – megkönnyíti). 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.

A módszer hátrányai:

- csak viszonylag kis létszámú célcsoport esetén alkalmazható könnyen,
- munkaigényes,
- csak a szerzői célcsoportot tekintve megbízható, külföldi – főleg távol-keleti – szerzők esetén nem pontos. (Utóbbiak vezetékneve gyakran egyszótagos, számos homonim nevű szerző létezik viszonylag kis szakterületen is)

Az ATPA régebben a születési évvel próbálta azonosítani a szerzőket, jelenleg – nyilván a KPA igényeinek megfelelően a köztestületi azonosító megadásával kísérletezik. Mivel a köztestületi azonosító egy újabb adatbázis (MTA Tudós Adattár) használatát is szükségessé teszi, ez némiképp megnehezíti a megállapítását.

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-tulajdonost”, 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. A szerzők rendelhetnek a nevükhöz egyedi azonosítót (jellemzően tudományterületet és munkahelyet) 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.

Véleményem szerint azonban figyelembe kell azonban venni azok igényét is, akik az adott szakterületen laikusként próbálnak egy adatbázisban eligazodni és a kutatók munkájáról képet nyerni. Számukra talán legjobb lenne a születési év és szakterület meghatározásának kombinációja. Ésszerűnek tűnhet a munkahely megjelölése is, azonban annak változása az adatbázisban is meg kell jelenjen. A köztestületi azonosítót nehézkesnek érzem, az APHA gyakorlata pedig viszonylag kisszámú kutató publikációjának elemzésekor működik jól.

6.4 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 társszerzőszám [Schubert, 2007].

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 szerzői névsort, a szerzők 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 az előnye, hogy a későbbi 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 (pl. MEPA) megelégszenek az et. al. megjelölés használatával. Az alábbiakban néhány, a sok társszerzős publikációk megjelenésével kapcsolatos problémát veszek 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 bízzák. Egyes adattárak (ATPA és KPA) nagy erőfeszítéseket tettek a függetlenség gépi detektálására, a WoK-ból megszerezve a hivatkozott és hivatkozó művek teljes szerző névsorát. 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. 6.7 fejezet)
- A probléma akut voltát az APHA-ból nyert következő grafikonon szemléltetem:



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

6.5 Önhivatkozások

Egy adott, esetleg kisebb szakterületen elkerülhetetlen az önidézetek sora. Ez egyszemélyes cikkek esetén könnyen kiküszöbölhető. Más a helyzet azonban az előző pontban taglalt problémánál, mikor is számos, esetenként ezres nagyságrendű szerzőgárda jegyzi a cikket. Amennyiben nem tartjuk számon a teljes szerzőlistát, az önhivatkozások kiküszöbölése igen esetleges, gyakran maga a nyilvántartott szerző sem tud erről megbízható adatot szolgáltatni.

Egyszerűbb esetben természetesen az önidézetek kiküszöbölhetőek. Ezt jelezve került bevezetésre a független hivatkozási szám.

Másrészt azonban, vannak olyan, gyakran a tudományos kutatás élvonalába tartozó területek, melyeknél már-már a teljes, témával foglalkozó közösség dolgozik egy cikken, következésképpen a későbbiekben az önidézetség nagyon valószínű, mivel nem marad olyan, témával foglalkozó, aki függetlenül idézne. Amennyiben ez esetekben kizárjuk a nem független idézeteket, gyakorlatilag a tudományos közlemény idézettségének jó részéről – elméletileg akár az összes citációról – lemondunk. Ez a gyakorlat pont azokat a kutatókat hozza kedvezőtlenebb helyzetbe, akik nagyszámú szakemberrel a kutatások élvonalában

dolgoznak együtt. Könnyű belátni, hogy ezekben az esetekben éppen hogy célszerű nyilvántartani a nem független idézeteket is.

Más eset az mikor egy csoport, team tagjai rendszeresen hivatkoznak egymásra. Ez az úgynevezett citációs gyűrű. Jelenleg erre Magyarországon korrekciót vagy mérőszámot nem alkalmaznak.

6.6 ISI/nem ISI, SCI/nem SCI

A tudományos kutatókra ráömlő rengeteg publikáció a tudományometriai értékelő eljárásokat is hatalmas feladat elé állította (ezekről részletesebben ld. pl. [Marton, 1999] és az ott idézett művek). 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 között (SCI: Publikáció: Ha a megjelenés évében a megjelenési helynek van impakt faktora (OTKA definíció); SCI Hivatkozás: Ha az illető megjelenési helynek valaha volt impakt faktora)

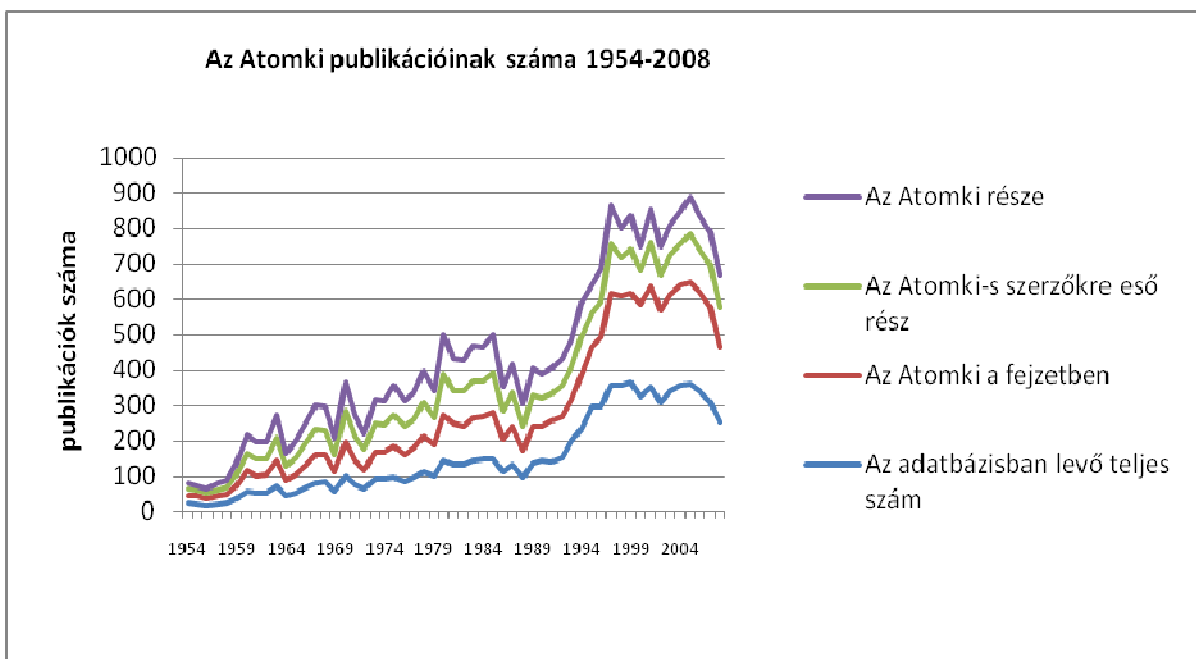
6.7 A teljesítmények additivitásának kérdése – az egyéni/intézményi teljesítmény elkülönítése

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. Néha felmerül a gyanú, hogy a szerzők egy részének szerepeltetése vitatható, a közreműködésük nem indokolja szerzői státusukat. A szerzőség egyes esetekben nagyon liberálisan, szubjektíven értelmezett. Valószínűleg ezeken a problémákon csak az segít, ha magukat a szerzőket tesszük érdekeltté

abban hogy csak azokat tüntessék fel akiknek valós szerepük volt a közlemény megszületésében. Ennek lehetséges eszköze a teljesítmény szétosztása.

Amennyiben más támpont nincs – mint a szabadalmak esetén -, célszerű úgy felosztani a szóban forgó „teljesítményt” (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. Ennek a problémának hazai felvetése [Zolnai, 1989]-ben történt, további taglalása megtalálható [Zolnai, 1998], [Zolnai, 2008]-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. Véleményem szerint számosabb publikációt tartalmazó lista esetén az „érdem” kiátlagolódik, ugyanis a több résztvevős együttműködések nem tűrik meg a „lazsálást”, ha valaki rajta van a szerzői 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 részarány $1/n$, míg az intézetének a részaránya $1/(n*(m+1))$. Az itt vázolt megoldást alkalmazza, és viszi végig az APHA.

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



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

Az 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 – véleményem szerint ez a felfogás túlságosan rugalmas. A harmadik görbe az intézeti szerzőkre eső, az intézetre összegzett publikációk számát 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. [Bencze, 1995], [Braun, 2002], [Tolnai, 2008b], [Marton, 2009]) általában a WoK adataira alapozódnak. Ezekben 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” módszer [Schubert, 2007]). Így egyes országok és/vagy intézetek a valóságos helyzetnél jóval kedvezőbb színben tűnhetnek fel, amint az már más szempontból is felvetődött [Török, 2000]. Természetesen a fentiek csak akkor igazak, ha mint az evaluatív tudománymetria általában teszi, a publikációkat azonos statisztikai értékűnek fogadjuk el.

6.8 Tárgyszavazás lehetőségei

Az áttekintett adatbázisok egy részében megteremtették a lehetőséget tárgyszavak felvitelére. Könnyen belátható: valamilyen tárgyi visszakereshetőség megteremtése kívánatos és előnyös lenne, de ez újabb problémákat vet fel. Egységes tárgyszavazás akkor biztosítható, ha azt

- az adatbázis kezelője végzi – ebben az esetben viszont szükséges a dokumentum tartalmának ismerete, a címből ezek általában nem eldönthetőek
- a szerző mint adatfeltöltő végzi – ekkor egy tárgyszórendszert kell kialakítani és széles körű felhasználói oktatásra is szükség van, máskülönben még közelítően sem lesz egységes a tárgyszavazás. Kérdéses, hogy hány szerző vállalkozna erre.
- Mivel számos folyóirat már a cikk leadásakor tárgyszavakat kér a szerzőtől, átvehetőek lennének ezek a tárgyszavak, bár ebben az esetben sem lennének egységesek, és az ezt nélkülöző cikkek ugyanúgy problémákat vetnek fel.

6.9 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 publikációs politikái (ld.: <http://www.sherpa.ac.uk/romeo.php>) 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 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. 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 dönteni, 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.

6.10 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.

Ebből a tényből azonban a hazai publikációs adattárak egy sajátos és nagyon lényeges tulajdonsága eredeztethető. Míg pl. 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 költségvetési pénzek korlátozott voltából ered az is, hogy az adattárak nagy és látványos célokat tűznek ki – amelyekhez szükséges eszközök fejlesztéséhez még, úgy ahogy meg vannak az eszközök – a legmunkaigényesebb munkafolyamatot, a feltöltést már az érintett szerzőkre bízják rá. Véleményem szerint, mivel egy adatbázis fenntartása hosszú távú folyamat, rendszeres karbantartást igényel, ennek költségeit az intézmény állandó költségvetésébe be kell építeni továbbá a humán erőforrásokat is biztosítani kell. Mint a 6.1 pontban vázoltam, nem tartom reálisnak azt az elképzelést, hogy alacsony motivációval nagyszámú, inhomogén csoportot rá lehet venni adatszolgáltatásra és adatrögzítésre. Hasonlóan fontos kérdés a rugalmasság, előre nem látható szempontok merülhetnek fel egy

adatbázis életében – ilyen volt pl. a cikkszám bevezetése egyes folyóiratokban füzet és oldalszám helyett – az adatbázist ekkor viszonylag gyorsan alkalmassá kell tenni ezen adatok tárolására is. Ezeknek a fejlesztéseknek a költségeire úgyszintén számítani kell.

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. Kivételesnek mondható az az eset, amelyben sikerül a munkafolyamat nagy részét egy személynek (APHA) átfogni, erre több tudományterületet átfogó adatbázisnál a lehetőség csekély.

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, további két adatbázis, az egyetemi fenntartású MEPA és PPKE BTK OPK néhány adatra korlátozza a statisztikát. 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 időt álló. Félő, hogy a hazai fejlesztésű adattárak ki vannak szolgáltatva az illető fejlesztőcégeknek.

6.11 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 hosszútávú tudománypolitikai döntések segítőjeként, vagy az egyes tudománybeli érdekcsoportok taktikai bunkósbotjaként használja. Ez egyrészt a fenntartókra, másrészt a felhasználókra gyakorol nyomást. Ettől 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 RIS formátumszabvány [RIS] használata.
 - Egységesíteni kellene az egyes adattárak által minimálisan szolgáltatott tudományometriai statisztikai jellemzőket és azok pontos definícióját.
 - Célszerű lenne az intézmények egyes szervezeti egységeire vonatkozó statisztikai idősorok bemutatása.
2. Ésszerű 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 tágitása nem jelentene jelentős kapacitásnövelési igényt, viszont jelentős hozadéka lehetne a dinamikus tudományometriai szempontokra vonatkozóan.

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 hasznosulnak hatékonyan.
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. A kapacitásbővítést mindenképpen egyszerűbbnek, gyorsabbnak és hatékonyabbnak látom, mint a kétséges sikerű motivációnövelést.
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 web-lapként újra kell generálni.
9. Az előző ponthoz kapcsolódik a kérdés, kívánjuk-e nyilvánossá tenni a hazai kutatási eredményeket a szélesebb közösség számára? Amennyiben igen, ésszerű lenne az angol nyelvű változatok elkészítése illetve az adatbázisok – regisztráció nélkül is – nyilvánossá tétele is. Vitatható az a gyakorlat miszerint az egyik legfontosabbnak szánt adatbázis, az MTA TPA csak korlátozottan megtekinthető [Rózsa – Papp 2009].

10. Meg kellene oldani néhány minőségbiztosítási kérdést, lehetőséget kellene biztosítani a felhasználóktól a fejlesztőig/kezelőig, sőt a fenntartóig irányuló visszacsatolásra. 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 6. fejezet előző részeiben ismertetett problémák konszenzuson alapuló tisztázása, annak eldöntése egyáltalán, hogy a jövőbeli teljesítményértékelési szempontok kiterjednek-e ezekre a tényezőkre.
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.

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 egyelőre csak minimális mértékben kapnak teret.

7.0 Összefoglalás

Dolgozatom néhány hazai fejlesztésű tudományos publikációs adattár összehasonlító elemzésével foglalkozik.

Rövid bevezetés, és a célkitűzés vázolója után a 2. fejezetben áttekintettem a publikációs adattárak kialakulásához vezető előzményeket, részletesebben a tudományometriai értékelések időbeli fejlődésének alakulását. Megpróbáltam érzékeltetni, hogy milyen érdekek motiválták a hazai publikációs adattárak kialakítását, és mennyiben befolyásolják azok az adattárak fejlesztését. A 3. fejezetben egy kivétellel röviden ismertettem a vizsgálatba bevont adattárakat. A 4. fejezetben a vizsgálatban felhasznált fogalmak definícióit, ill. a vizsgálatban alkalmazott módszereket tekintetem át. Az 5. fejezetben a vizsgált adattárak fontosabb jellemzőit hasonlítottam össze egy táblázatban.

A 6. fejezetben néhány, az adattárak működtetésével kapcsolatos fontosabb szempontból tárgyaltam az egyes adattárakat:

- Összehasonlítottam az egyes adattárak által használt feltöltési sémákat. Megállapítottam, hogy a működés szempontjából az esetek többségében döntő fontosságú a feltöltésre felkért szerzők motiváltsága.
- Elemeztem az „inaktív” szerzők problémáját. Arra a következtetésre jutottam, hogy a dolgozatban vizsgált adatbázisok teljes körű működtetése elképzelhetetlen a kezelői szinten beépített, az inaktív szerzők tevékenységét elvégző pótlólagos erőforrások nélkül.
- Ráműtattam az azonos nevű szerzők problémájának részleges – de nem megnyugtató – megoldására a vizsgált adattárak esetében.
- Megvizsgáltam a sok társszerzős publikációk adatainak tárolási és teljesítmény-értékelési problémáját.
- Analizáltam annak a jelentőségét, hogy az egyes adattárak hogyan kezelik az „ISI” publikációk kérdését.
- Megvizsgáltam a tárolt publikáció adatokkal kapcsolatosan az adatokból adódóan számolható teljesítmények additivitásának kérdését. Ráműtattam, hogy az APHA által alkalmazott megoldás lehetőséget ad az egyéni/intézményi teljesítmények algebrailag helyes összeszámlálására.

- Áttekintettem, hogy a vizsgált publikációs adattárak milyen esetekben döntöttek a tárolt publikációk teljes szövegű változatának tárolása mellett. Megállapítottam, hogy az ilyen irányú fejlesztések csak kevés esetben vezettek sikerre.
- A rendelkezésemre álló szűkös adatok alapján elemeztem a vizsgált adattárak működtetésének néhány gazdasági kérdését.

A 6.11 fejezetben következtetéseket és javaslatokat fogalmaztam meg az adattárak jövőbeli fejlesztésével kapcsolatosan, abban a reményben, hogy javaslataim felkeltik az illetékes fenntartók, fejlesztők, kezelők és felhasználók figyelmét.

A dolgozat anyaggyűjtési időszakának lezárulta után értesültem 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 kívánja hozni a Magyar Tudományos Művek Tárát (MTMT). Ezen adatbázis távlatilag egyesíti a publikációk és hivatkozások nyilvántartására vonatkozó igények kiszolgálását (http://mta.hu/index.php?id=634&no_cache=1&backPid=390&tt_news=11266&cHash=873f77c3c7).

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

8.0 Köszönetnyilvánítás

Ezúton szeretnék köszönetet mondani témavezetőmnek, Eszenyiné dr. Borbély Máriának a dolgozat elkészítése során nyújtott segítő tanácsaiért, a dolgozat végleges változatának gondos átolvasásáért.

Köszönöm dr. Nagy Józsefnének támogatását továbbá a számos hasznos tapasztalatot, amit megosztott velem a publikációs adatbázisok gyakorlati működéséről, és munkahelyi felettesemnek dr. Sulik Bélának segítő megértését.

9.0 Irodalomjegyzék

Bencze Gyula [1995]: Helyünk Európában. Fizikai Szemle, 45. évf., pp. 18-20.

Bencze Gyula [2006]: H-index: Egy új javaslat az egyéni tudományos teljesítmény értékelésére. Magyar Tudomány, 167. évf., pp. 88-91.

Braun Tibor – Glänzel, Wolfgang – Némethné Kovács Éva – Pereszteginé Szabadi Zsuzsa [2002]: Magyarország helyzete a természettudományi alap kutatás világában – tudományometriai tájkép a második évezred végén. Magyar Tudomány, 163. évf., pp. 935-945.

Braun Tibor – Glänzel, Wolfgang – Schubert András [2006]: A Hirsch-type Index for Journals. Scientometrics. 69. sz. pp. 169–173.

Braun Tibor [2008]: Szellem a palackból, tudományometriai értékelések. Magyar Tudomány, 169. évf. pp. 1366-1371.

Dobrov, G. M. [1973]: A tudomány tudománya. Gondolat-Kossuth, Budapest

Drótos László – Kőrös Kata – Somogyi Tamás: Adattármustra. Publikációs adatbázisok. Tudományos és Műszaki Tájékoztatás, 55. évf. 3. sz. pp. 127-145.

Dudás Anikó [2008]: Tudományos önismeret: az egyetemi bibliográfiák szükségességéről és egy bölcsészettudományi oktatói publikációs adatbázisról. Tudományos és műszaki tájékoztatás. 55. évf. 3. sz. pp. 107-125.

Eger, Arthur [2009]: Hogyan írjunk világszínvonalú tanulmányt? Praktikus tanácsok publikáláshoz. Tudományos és Műszaki Tájékoztatás, 56. évf. 7. sz. pp. 307-319.

Garfield E., Sher I. H., [1963]: New Factors in Evaluation of Scientific Literature through Citation Indexing. American Documentation, 14. vol. pp. 195-201.

Garfield E., Sher I. H., Torpie R. J. [1964]: The use of citation data in writing the history of Science. Institute of Scientific Information Inc. Philadelphia, Pennsylvania, USA

Garfield E. [1972]: Citation analysis as a tool in journal evaluation. Science, 178. vol. pp. 471-479.

Gyergyai Albert [1984]: „Az el nem ért bizonyosság” In: Védelem az esszé ügyében. Szépirodalmi könyvkiadó, Budapest

Hirsch J. E. [2005]: 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. pp. 16569-16572.

Heti Válasz [2008a]: A száz legidézettebb természetkutató. 2008. jan. 17.

Heti Válasz [2008b]: Tudósok a mérlegen. 2008. febr. 21.

Jehoda Imola [2006]: Európai tudományometriai modellek az orvos- és egészségügyi publikációs tevékenység értékeléséhez. Tudományos és Műszaki Tájékoztatás, 53. évf. 5. sz. pp. 224-237.

Kiss Andrea – Vitéz Gáborné [2006]: Miskolci Egyetemi Publikációs Adatbázis. Tudományos és Műszaki Tájékoztatás, 53. évf. 7-8. sz. pp. 321-327.

Kollár István – Michelberger Pál [2008]: Hozzászólás Tolnai Márton *Tudósaink mérhető teljesítménye az MTA Köztestületi Publikációs Adattár adatainak tükrében* című cikkéhez. Magyar Tudomány, 169. évf. pp.1262-1273.

Korn András – Schubert András – Telcs András: Lobby index in networks. [2009] Physica A, 388. vol. pp. 2221-2226.

Szakterületi statisztikák magyar kutatók Hirsch-féle h-indexeire az MTA Köztestületi Publikációs Adattárának adatai alapján. MTA-KSzI, 2005 november.

Marton János [1999]: Bibliometria. In: *Könyvtárosok kézikönyve 1. kötet* Osiris, Budapest pp. 83-151.

Marton János [2009]: Tizenkét futam az európai tudományometriai mezőnyben. Élettudományok. Tudományos és Műszaki Tájékoztatás, 56. évf. 3. sz. pp. 118-122.

Nalimov V. V., Mulcsenko Z. M. [1980]: Tudománymetria. Bp., Akadémiai Kiadó, 1980.

[RIS] http://www.refman.com/support/risformat_intro.asp

Rózsa Lajos – Papp László [2009]: A csoportmunka, mint egyéni érdem és egyéb megjegyzések a Köztestületi Publikációs Adattár használatáról. Magyar Tudomány, 170. évf. pp. 95-102.

Solla Price, Derek de [1979]: Kis tudomány – Nagy tudomány. Akadémiai Kiadó, Budapest

Schubert András [2007]: Tudománykommunikáció képzés. Előadás. 2007. okt. 3.

- Tolnai Márton** [2008a]: Tudósaink mérhető teljesítménye az MTA Köztudományi Közlöny Publikációs Adattár adatainak tükrében. Magyar Tudomány, 169. évf. pp. 976-988.
- Tolnai Márton – Schubert András – Wolf György** [2008b]: Tudományunk mérhető teljesítménye az Essential Science Indicators mutatószámainak tükrében. Magyar Tudomány, 169 (2008) 989-997.
- Tolnai Márton – Mosoniné Fried Judit – Soós Sándor** [2009]: F fiatal kutatók az MTA köztestületeiben. Magyar Tudomány, 170. évf. pp. 334-344.
- Török Ádám** [2000]: Reális-e a magyar tudomány 20. helye a (képzeltbeli) világranglistán? Magyar Tudomány (45. (108.). évf. pp. 1307–1328.
- Vinkler Péter** [2008]: Tudományometriai kutatások Magyarországon. Magyar Tudomány, 169. évf. pp. 1372-1380.
- Zádori Zsolt** [2006a]: Mit ér a tudós, ha magyar. Heti Világgazdaság, 28. évf. .20. sz. Melléklet
- Zádori Zsolt** [2006b]: Hódít a teljesítménymérés. Heti Világgazdaság, 28. évf. 41. sz. pp. 68-71.
- Zolnai László – Berényi Dénes** [1989]: Kutatóintézeti tudománymetria – ATOMKI 1954–1989. Fizikai Szemle. 39. évf. pp. 285–291.
- Zolnai László – Gácsi Zoltán** [1998]: Mérünk, de mit? Egy formula margójára. Magyar Tudomány, 43. (105.) évf. pp. 988-993.
- Zolnai László** [2001]: Mennyire magyar a magyar? Magyar Tudomány, 48. (108.) évf. pp. 1497-1498.
- Zolnai László** [2008]: A „sötét teljesítmény” nyomában. Magyar Tudomány, 169. évf. pp. 870-874.
- Zolnai László** [2001]: Tudománymetria és kollaboráció. Fizikai Szemle. 51. évf., pp. 264–265.

10.0 Függelék

1. Függelék

Az akadémiai kutatóintézetek 1996-os értékelésére használt formula

1996-ban az MTA természettudományi intézeteit több formulát alkalmazva is megpróbálták értékelni. Ezek a formulák lényegileg egyformák, mind egy pontszámban próbálta összesíteni az intézetek "teljesítményét". Ezek közül egyet ismertetünk, amelynek nagy hatása volt az intézetek besorolásában. Ezt a formulát – a többivel együtt – az MTA Matematikai és Természettudományi Kuratóriumának Kutatáselemzési Munkabizottsága alakította ki.

Az intézetektől az 1992-95-os időszakra vonatkozó különböző teljesítménymutatókat kértek. Ezek közül a formulába végül is szereplők^{1,2} definíciói a következők voltak:

$a_{1,int}$ az adott időszakban egy intézeti kutatóra évenként eső publikációs szám átlaga

$a_{2,int}$ az adott időszakban egy intézeti kutatóra évenként eső idegen nyelven megjelent publikációk számának átlaga

$a_{3,int}$ az adott időszakban egy intézeti kutatóra évenként eső SCI³-folyóiratokban megjelent publikációk számának átlaga

$a_{4,int}$ az adott időszakban évenként megjelent publikációk hatástényezőinek⁴ (impakt faktorainak) összege osztva az SCI-cikkek illető évbeli számával, átlagolva az adott időszakra

¹ Az intézetektől más teljesítménymutatókat is kértek, pl. az egyetemi oktatásban tartott órák számát, az intézetekben készült PhD disszertációk számát, stb. Ezeket a rossz kérdésfeltevésből eredő "értékelhetetlen" válaszok miatt nem vonták be az értékelésbe.

² Véleményünk szerint az értékelés talán legnagyobb hiányossága, hogy nem veszi figyelembe a különböző tudományterületek eltérő relatív tökeellátottságát.

³ Science Citation Index, az Institute for Scientific Information, Philadelphia által kiadott mutató, amely a világ legjelentősebb természettudományos folyóirataiban megjelent hivatkozások adatait tartalmazza. SCI-folyóiratok mindazok a folyóiratok, amelyeket az SCI nyomon követ.

⁴ A hatástényezőt az SCI évente számolja ki a benne szereplő folyóiratokra. Egy folyóirat n -edik évi hatástényezője egyenlő a folyóiratban $n-1$ -edik és $n-2$ -edik évben megjelent közleményekre az SCI-ben az $n-1$ és $n-2$ -dik évben fellelhető hivatkozások száma osztva a folyóiratban $n-1$ -edik és $n-2$ -edik évben megjelent közlemények számával. Ily módon a hatástényező az adott folyóiratban megjelent egy közleményre várható SCI hivatkozások számára ad relatív becslést. Mivel a hatástényező egy olyan sokaságon vett átlag, amely lényegesen szélesebb mint az egy intézetből publikált cikkek halmaza, ezért véleményünk szerint csak a közelmúltban megjelent közlemények hatásának figyelembevételére célszerű használni.

$a_{5,int}$ az $a_{4,int}$ osztva h -val ahol h az tudományterület folyóiratainak átlagára jellemző átlagos hatástényező (az intézetek által megadott "tevékenységi spektrum"-ból az MTA Kutatásszervezési Intézete számolta ki)

$a_{6,int}$ az intézeti kutatóévenkénti idézetszám, (a 1982-95 között megjelent publikációkra vonatkozó, 1992-95-ben évenként megjelent független SCI-idézetszám, osztva az illető intézet kutatóinak évenkénti átlaglétszámával, átlagolva az 1992-95-ös időszakra)

$a_{7,int}$ az adott időszakban nemzetközi együttműködésben készült publikációk száma osztva az összes publikációk számával

$a_{8,int}$ az adott időszakban bejelentett szabadalmak száma osztva az intézeti kutatók számával

$a_{9,int}$ az adott időszakban idegen nyelven megjelent publikációk száma osztva az MTA támogatás összegével (MFt-ban)

$a_{10,int}$ az adott időszakban találmányok értékesítéséből befolyt összeg osztva az intézeti kutatók számával

$a_{11,int}$ az intézetben tudományos minősítéssel rendelkezők száma osztva az intézeti kutatók számával.

1996 februárjában a fentiekben publikációnak nevezett kategória egyenlő volt folyóiratokban megjelent közlemények, konferenciaközlemények, könyvek és kiadványrészletek összegével. Az intézet közleményének számított az, amelyben az intézet neve bármilyen formában fel volt tüntetve. Nemzetközi együttműködésben született publikáció volt az, amelyben külföldi társszerzők szerepeltek. SCI-hivatkozás volt az, melyben az idéző cikk folyóiratának a megjelenési évre vonatkozólag létezett hatástényezője. Független hivatkozásnak számított az a hivatkozás, amelyben a hivatkozó és hivatkozott publikációban nincs közös társszerző.

1996 októberében a definíciók a Természettudományi Főosztály kezdeményezésére annyiban módosultak, hogy 1. Intézeti tudományos publikációnak csak az számít, amelyben az intézet neve a publikáció fejezetében szerepel, 2. SCI hivatkozás az, amely az SCI-ben fellelhető.

1997 februárjában a Természettudományi Főosztály által kért, 1996-os évre vonatkozó beszámolóban csak olyan tudományos közlemény számított elfogadott publikációnak, amely ISSN/ISBN számmal rendelkező kiadványban jelent meg.

A fentebb említett munkabizottság 1997 márciusi "továbbfejlesztett" értékelésében mégis az összes idézetszámot (tekintet nélkül azok leőhelyére) vette figyelembe⁵.

A k -dik intézet P_k teljesítményét, a kuratórium által elfogadott w_i súlyfaktorok⁶ segítségével a következő képlet alapján számolták ki:

$$P_k = \sum_{i=1}^n w_i \frac{a_{i,k}}{\sum_{j=1}^m a_{i,j}}$$
 ahol n a szempontok, m az értékelésbe bevont intézetek száma, $a_{i,j}$ az

egy-egy intézetnek a különböző szempontok szerinti teljesítménymutatói.

⁵ Ennek valószínű oka az volt, hogy nem sikerült az intézetek által beküldött hivatkozási listákat ellenőrizni. Amennyiben ez így van kérdéses, hogy a hivatkozások tekintetében érdemes-e az SCI-hivatkozásokra szorítkozni, annál is inkább mivel néhány tudományterület az SCI-ben különböző okoknál fogva alulreprezentált. Ebben az esetben szükség lenne a hivatkozás pontos definíciójára is.

⁶ A súlyfaktorok jelen számításban alkalmazott értékei a teljesítménymutatók fenti sorrendjében a következők voltak: 2, 3, 5, 7, 11, 12, 10, 3, 5, 7, 10.

2. Függelék

Az APHA publikációs és hivatkozást bejelentő lapja

P **Publikációt bejelentő lap** Kód:**P** (_ _ _ _)

_ _ _ _ _Téma (több is megjelölhető!): magfiz.①; atomfiz.②; anyagtud. és analitika③; föld- kozmikus és környezettud.④; műszerek és módszerek fejl.⑤; részecskefiz. és egyéb⑥; Biológiai és orvosi kutatás⑦

_ _Jelleg: tud. közl.:**A**; disszertáció:**B**; dipl. munka:**C**; kiadvány- könyv, könyvrészlet:**D**; szerk. mű:**E**; szabadalom:**F**; kutatási jelentés:**G**; konf. kivonat, poszter, előadás:**H**; egyéb poszter, előadás:**I**; ismeretterj. közl.:**J**; ismeretterj. előadás:**K**; egyéb:**L**; évkönyv jellegű kiadvány:**M**. A jelleg **H**, és a proc.-ben is megjelenik, igen (1), nem (0)☞ _ _

_ _Nyelv: angol:**A**; magyar:**B**; orosz:**C**; német:**D**; francia:**E**; spanyol:**F**; egyéb:**G**.

(_ _ _ _)Cím (eredeti nyelven):.....

.....

(_ _ _ _)Megj. hely:.....

Kötet: _ _ _ _ _ Megj.év: _ _ _ _ _ Füzet: _ _ _ Oldal: _ _ _ _ _-tól, _ _ _ _ _-ig

AmgjhnykdSSNISBNszámvn①, rncs②: _ _ _ Impkt(_ _ . _ _ _)

Szerzők: ☞

Szerzők száma: _ _ _ _ _

(_ _ _ _) 1.....(_ _ _ _) 16.....

(_ _ _ _) 2.....(_ _ _ _) 17.....

(_ _ _ _) 3.....(_ _ _ _) 18.....

(_ _ _ _) 4.....(_ _ _ _) 19.....

(_ _ _ _) 5.....(_ _ _ _) 20.....

(_ _ _ _) 6.....(_ _ _ _) 21.....

(_ _ _ _) 7.....(_ _ _ _) 22.....

(_ _ _ _) 8.....(_ _ _ _) 23.....

(_ _ _ _) 9.....(_ _ _ _) 24.....

(_ _ _ _) 10.....(_ _ _ _) 25.....

(_ _ _ _) 11.....(_ _ _ _) 26.....

(_ _ _ _) 12.....(_ _ _ _) 27.....

(_ _ _ _) 13.....(_ _ _ _) 28.....

(_ _ _ _) 14.....(_ _ _ _) 29.....

(_ _ _ _) 15.....(_ _ _ _) 30.....

_ _Hozó: "ATOMKI" fejlécben:① "On leave...":② "Permanent address", stb.:③ "ATOMKI" nem szerepel:④

A bejelentés dátuma: _ _ _ _ _ hó..... -án

_____ a bejelentő olvasható aláírása

H

Hivatkozást bejelentő lap

A hivatkozott mű adatai:

Kód:**P** (_ _ _ _)

Első szerző:.....

Megj. hely:.....

.....
.....

Kötet:|_|_|_| Megj. év:|_|_|_| Oldal:|_|_|_|_|

A hivatkozó mű adatai:

Első szerző:.....(_ _ _ _)

Megj. hely:.....

.....
.....(_ _ _ _)

Kötet:|_|_|_| Megj. év:|_|_|_| Oldal:|_|_|_|_|

A bejelentés dátuma:|_|_|_|.hó _____

a

bejelentő

olvasható aláírása

Kód:**H** (_ _ _ _)

3. Függelék

Az APHA által számolt tudományometriai statisztikai jellemzők

Az Atomki publikációs-hivatkozási adatbázisában (APHA)

használt elsődleges táblázatok leírása

Listed\	Total	Headed by the ATOMKI	Part of the authors from the ATOMKI	Part of the ATOMKI
Publications:	323.000	261.000	95.526	70.939
SCI Publications:	216.000	172.000	44.136	29.623
Cited publications:	186.000	156.000	39.997	28.791
Non-locally cited publications:	184.000	154.000	38.552	27.403
Cited ^{SCI} publications:	164.000	140.000	31.508	21.990
Non-locally cited ^{SCI} publications:	163.000	139.000	31.063	21.601
SCI Cited publications:	158.000	136.000	29.439	19.795
Non-locally ^{SCI} cited publications:	158.000	136.000	29.439	19.795
SCI Cited ^{SCI} publications:	150.000	128.000	26.862	17.622
Non-locally ^{SCI} cited ^{SCI} publications:	150.000	128.000	26.862	17.622
Citations:	2472.000+(0.000)	2017.000	358.428	235.372
Non-local citations:	2424.000	1977.000	334.222	214.218
SCI Citations:	1828.000+(0.000)	1484.000	273.950	182.533
Non-local ^{SCI} citations:	1808.000	1466.000	262.646	172.334
Citations found:	2678.000	1571.000	573.496	483.197
Non-local citations found:	2603.000	1502.000	523.553	434.226
SCI Citations found:	2145.000	1130.000	461.409	385.443
Non-local ^{SCI} citations found:	2102.000	1092.000	434.925	359.767
Citations appeared:	2656.000	1705.000	433.971	323.224
Non-local citations appeared:	2625.000	1677.000	410.771	301.608
SCI Citations appeared:	1802.000	1167.000	346.461	257.721
Non-local ^{SCI} citations appeared:	1777.000	1143.000	327.345	239.605
Impact ²⁰⁰⁷ :	547.106	446.712	71.634	40.054
Citational effectivity:	7.653	7.728	3.752	3.318
Non-local citational effectivity:	7.505	7.575	3.499	3.020
SCI citational effectivity:	8.463	8.628	6.207	6.162
Non-local ^{SCI} citational effectivity:	8.370	8.523	5.951	5.818
Averaged impact:	1.694	1.712	0.750	0.565
SCI averaged impact:	2.533	2.597	1.623	1.352
Publications <n>:	172.693	208.226	*****	*****
SCI Publications <n>:	188.199	231.140	*****	*****
Cited publications <n>:	149.409	173.628	*****	*****
Non-locally cited publications <n>:	150.978	175.818	*****	*****
Cited ^{SCI} publications <n>:	168.012	192.700	*****	*****
Non-locally cited ^{SCI} publications <n>:	168.988	194.022	*****	*****
SCI Cited publications <n>:	173.639	197.794	*****	*****
Non-locally ^{SCI} cited publications <n>:	173.639	197.794	*****	*****
SCI Cited ^{SCI} publications <n>:	182.620	209.828	*****	*****
Non-locally ^{SCI} cited ^{SCI} publications <n>:	182.620	209.828	*****	*****
Effective citations (INR):	541.800	407.352	187.047	128.163
Equivalent citations (HAS):	1146.750	914.500	317.429	211.615
z-index:	25.000	24.000	6.916	6.083

Footprints in science				
Papers\	Total	Headed by the ATOMKI	Part of the authors from the ATOMKI	Part of the ATOMKI
Unknown ($n_{cit}=0$)	138.000	106.000	56.078	42.575
Less-known ($0 < n_{cit} < 10$)	112.000	90.000	29.740	21.459
Known ($9 < n_{cit} < 50$)	65.000	61.000	7.914	5.983
Well-known ($49 < n_{cit} < 100$)	7.000	3.000	1.685	0.917
Famous ($99 < n_{cit} < 500$)	1.000	1.000	0.005	0.002
Renowned ($499 < n_{cit}$)	0.000	0.000	0.000	0.000
SCI ¹ Unknown ($n_{cit}=0$)	165.000	125.000	66.065	51.143
SCI ¹ Less-known ($0 < n_{cit} < 10$)	97.000	82.000	21.972	14.549
SCI ¹ Known ($9 < n_{cit} < 50$)	55.000	51.000	6.447	4.574
SCI ¹ Well-known ($49 < n_{cit} < 100$)	5.000	2.000	0.935	0.667
SCI ¹ Famous ($99 < n_{cit} < 500$)	1.000	1.000	0.005	0.002
SCI ¹ Renowned ($499 < n_{cit}$)	0.000	0.000	0.000	0.000

Az APHA-ban elsődleges produktumok az intézet éves publikációs listái (Pl.:

<http://www.atomki.hu/p2/years/yea01981.htm#Table>)

ill. a kutatók publikációs listái (Pl.:

<http://www.atomki.hu/p2/authors/aut00125.htm#Table>)

A szóban forgó táblázatok ezeknek a listáknak a végén találhatóak, és adataik mindig az előtük lévő lista adataiból származnak. A fent említett két lista kevéssé tér el egymástól. Az alábbiakban a listákban használt nem triviális kifejezések jelentését adjuk meg:

Total A listában valamilyen okból szereplő publikációk és az azokra vonatkozó hivatkozások teljes száma (pl.: Total publications)

Headed by the ATOMKI Azok a publikációk, amelyek fejezetében szerepel az Atomki megnevezése, valamint az ilyen publikációkra történt hivatkozások.

Part of the authors from the ATOMKI: A publikációk és a rájuk vonatkozó hivatkozásoknak az Atomkival kapcsolatban lévők kutató(k)ra eső része (a szerzők számával beosztott rész annyiszor véve ahány intézeti szerzőről van szó – egyéni listánál egyszer)

Part of the ATOMKI A publikációk és a rájuk vonatkozó hivatkozásoknak az Atomkival kapcsolatban lévők kutató(k)ra eső része, korrigálva az egyes szerzőknél szereplő intézetek számával. Pl.: ha egy publikáció hatszerzős, és ebből egy Atomki-s, és az 3 intézetet írt a nevéhez, amiből egyik az Atomki, akkor az publikációs érték $1/(6*3)$, a hivatkozási érték (hivatkozások száma)/(6*3)

SCI Az SCI-ben szereplő, pontosabban:

Publikáció: Ha a megjelenés évében van impakt faktora (OTKA definíció)

Hivatkozás: Ha az illető megjelenési helynek valaha volt impakt faktora (Atomki definíció)

Non-locally cited Olyan hivatkozás, aminek (első) szerzője nem intézeti (Ez csak tájékoztató adat, mivel az APHA-ban csak a hivatkozások első szerzőjét rögzítjük).

Citations Az adott lista publikációira kapott hivatkozások száma.

Citations found (Csak intézeti listánál) Az adott évben fellelt, a lista publikációira történt hivatkozások száma

Citations appeared (Csak intézeti listánál) Az adott évben megjelent, a lista publikációira történt hivatkozások száma

Impact^{YEAR} A listában szereplő publikációk összegzett impakt faktora. A felső index az utolsó figyelembe vett impakt faktor táblázat éve.

Citational effectivity Hivatkozási hatékonyság, a hivatkozások száma osztva a megfelelő publikációk számával.

Averaged impact A publikációk számára átlagolt összipakt

<n> Társszerzők száma

Effective citations A lista effektív idézettsége ($\sum_i h_i t_i^{-1/2}$, ahol t_i az i -edik publikáció társszerzőinek száma, h_i pedig az erre vonatkozó független hivatkozásoké, és az összegzés a lista összes publikációjára vonatkozik). Ezt a mutatót az Atomki használja habilitációs eljárásban.

Equivalent citations A lista ekvivalens idézettsége ($\sum_i h_i t_i$, ahol t_i az i -edik publikáció n társszerzőinek számától függő faktor ($t=1$, ha $n<6$, $t=0.75$ ha $5<n<11$, $t=0.5$ ha $10<n<21$ és $t=0.25$ ha $n>20$), h_i pedig az erre vonatkozó független hivatkozásoké, és az

összegzés a lista összes publikációjára vonatkozik). Ezt a mutatót az MTA Fizikai Osztálya használja az MTA doktora címre való pályázatban.

z-index A Hirsch-index analógiájára, az adatbázis adataiból számolt mennyiségek. A leglényegesebb eltérés az, hogy csak független hivatkozások vannak figyelembe véve.

A „Footprints in science” táblázat a nagyenergiájú fizikából kölcsönzött módon az adott számú hivatkozással bíró publikációk száma szerint osztályozza a listában szereplő publikációkat.