

# ADATOK, LEHETŐSÉGEK, FELADATOK A KUTATÁSI ADATOK HOZZÁFÉRHETŐSÉGÉRŐL

Holl András

csillagász, informatikus,

MTA Csillagászati és Földtudományi Kutatóközpont

Konkoly Thege Miklós Csillagászati Intézet, MTA Könyvtár és Információs Központ

holl.andras@konyvtar.mta.hu

A tudomány fejlődése megköveteli, a technika fejlődése pedig lehetővé teszi a tudományos kommunikáció átalakítását, ezen belül a tudományos adatok hozzáférhetőségének javítását (Holl, 2013). A tudományos kommunikáció jövőjével kapcsolatban érdemes megismerni a FORCE II szervezet tevékenységét (URL1). A digitális adatmenedzsment aktuális téma, cikkünkben megpróbálunk ízelítőt adni az ez irányú eredményekből, kezdeményezésekből. Áttekintjük a motivációt és a nehézségeket, bemutatjuk a világszerte szaporodó szabályozási kezdeményezéseket, a tudományos adatok megőrzését és közzétételét vizsgáló különböző tanulmányokat, felméréseket, végül a már létező megoldásokat. Nem törekszünk a kérdés részletes tárgyalására, inkább csak fogódzókat adunk azoknak, akiket a téma érint vagy érdekel.

A tudományos adatok nyilvánossá tétele és hosszú távú megőrzésének számos előnye lehet. A közvetlenebb hasznát a publikációkhoz felhasznált adatok nyilvánosságra hozatala szolgáltatja: ezek lényegesek az eredmények ellenőrzése és a vizsgálatok megismételhetősége szempontjából. További haszon lehet az adatok többszöri felhasználása: gyakran a mérés, az adatgyűjtés eredeti célján

túlmenően más vizsgálatokra is használhatóak lennének, ha rendelkezésre állnának. Végül számos olyan eset van, amikor az adatok sokkal jobban használhatóak, amennyiben az adatgyűjtés hosszú távú, megfelelő lefedettséget biztosít, s ez gyakorta különböző adatgyűjtők adatainak használatát, esetleg különböző adatgyűjtési módszerekkel és eszközökkel nyert adatok egységesítését igényli, és megköveteli az adatok hosszú távú megőrzését és hozzáférhetővé tételét. Az utóbbi esetekben, ahol az adatgyűjtést különböző kutatók, esetleg különböző módon végezték, vagy a felhasználó(k) nem azonos(ak) a megfigyelővel, mérést végzővel, az eljárások és az adatok pontos dokumentálása, a megfelelő minőség biztosítása, a szabványos módszerek és formátumok alkalmazása kulcsfontosságú.

A tudományos adatok hozzáférhetőségének követelménye megjelenik az EU 2020 Horizont programjában, az Egyesült Államok kormányzatának Open Access kezdeményezésében (Holdren-memorandum),<sup>1</sup> de az OTKA támogatási szabályaiban is. Bár ennek széles körűen eleget tenni valószínűleg

<sup>1</sup> Az Open Access bevezetéséről az Egyesült Államok kutatásfinanszíró szervezeteinél (URL2)

csak hosszú távon lehet, a tudománypolitika által megjelölt irány egyértelmű.

A tudományos adatok megosztását és megőrzését a technika: az informatika és az internet jelentősen megkönnyíti - és egyben meg is nehezíti. Ha számadást készítünk majd a ránk maradt és továbbörökítendő adatvagyonról, könnyen lehet, hogy a hetvenes-nyolcvanas-kilencvenes években keletkezett adatok tekintetében hiányt fogunk tapasztalni: a korábbi adatok papírra nyomtatva, üveglemezre felvitt fotografikus emulziókon, mikrofilmeken rendelkezésre állnak majd, a korai digitális adatok lyukszalagjai, mágnesszalagjai viszont már olvashatatlanok lesznek, az adathordozók romlása vagy az olvasóberendezések hiánya miatt. És ahol a biteket sikerül is megmenteni, ott a régi formátumok olvashatóvá tételére kell majd jelentős munkát fordítani.

\*

Az adatok megőrzése és közzététele sokkal összetettebb problémát jelent, mint a publikációké. Számos nemzetközi projekt próbálja a szükséges követelményrendszert, politikákat, szabványokat kialakítani. Az Alliance for Permanent Access (URL3) és a kapcsolódó APARSEN-projekt weblapjai és konferenciái jó tájékoztató pontot kínálnak. Az európai politikák felméréseivel foglalkozik a SIM4RDM (URL4), amelyben hazai részről az Nemzeti Információs Infrastruktúra Fejlesztési (NIIF) Intézet is részt vesz. A projekt egyik fontos eredménye az az ajánlás, amelyet a kutatási adatok menedzselésére vonatkozóan tettek közzé.<sup>2</sup> Ezek közül a következőt emeljük ki: „A kutatási alapok bátorítsák a

<sup>2</sup> SIM4RDM Recommendations for support of research data management (URL5)

kutatókat arra, hogy adatkezelési tervet nyújtsonak be a pályázat részeként, valamint irányozzanak elő forrásokat az adatok kezelésére.” Az adatarhívumok felállításának már szabványa is van: az OAIS (URL6). A publikációs repozitáriumokkal foglalkozó európai OpenAIRE-projekt folytatása, az OpenAIRE-plus (URL7) már az adatokkal is foglalkozik. Finn, dán, német, egyesült királyságbeli és holland intézmények együttműködéséhez ad keretet a Knowledge Exchange (URL8). A projekt eredményei közül az egyik tanulmány a tudományos adatok scientometriai szempontjaival foglalkozik (*The Value of Research Data*, URL9). A kutatási adatokra való hivatkozást a DataCite (URL10) által szolgáltatott DOI-azonosítók teszik lehetővé, a hivatkozásokat pedig már gyűjti a Thomson Reuters (Data Citation Index, URL11).

A csillagászat területén a Virtuális Observatórium kezdeményezés foglalkozik az adatok újrafelhasználhatóvá tételével. A szervezeti kereteket az International Virtual Observatory Alliance (IVOA, URL12) biztosítja. A kibővített folyóiratok (*enhanced journals*) tekintetében érdemes a holland EJME-projektet (*Enhanced Journals...Made Easy!* URL13), és a GreyNet EPP (URL14) projektjét tanulmányozni. Egyes kutatóhelyek – mint az Edinburgh-i Egyetem – már eljutottak arra a fokra, hogy úttervet készítettek a kutatási adatok kezelésének bevezetésére (University of Edinburgh Research Data Management Roadmap, URL15).

\*

A tudományos adatok hozzáférhetővé tételére és megőrzésére sokféle megoldás lehetséges. A nagy, kifejezetten felmérés jellegű, alapvető adatok begyűjtésére szervezett programok esetében szokásos az adatok nyilvánossá té-

tele. A Hubble Űrtávcső segítségével gyűjtött adatok csak meghatározott ideig maradnak a távcsőidőt elnyert pályázó „birtokában”, azután nyilvános archívumba kerülnek. A „Big Science” esetében nem jelent gondot az adatok megfelelő előkészítése, archiválása és nyilvánossá tétele – erre bőven futja a program költségvetéséből. A finanszírozó, ha egyszer sok pénzt költött egy ilyen programra, általában elő is írja az adatok nyilvánosságra hozását, ezzel a befektetésből származó tudományos haszon maximalizálását. Az említett Hubble Űrtávcső esetében mára több publikáció keletkezik a nyilvános archívum adataiból, mint amennyit az eredeti megfigyelési adatokból írnak maguk a megfigyelők. A részecskefizika területén az LHC (nagy hadronütköztető) adatainak megőrzésére szóba került rendszer a DASPOS (URL16).

Nem csupán a „Big Science” keretében keletkeznek adatok. A kis projektek keretében gyűjtött adatok tárolására kínál megoldást a Harvard Egyetemhez köthető Dataverse (URL17). Egy-egy Dataverse repozitórium számos projekt adatainak adhat helyet, és a repozitóriumok hálózatba szerveződnek (Dataverse Network). A projekt az adatok idézhetővé tételét a DOI alkalmazásán túl az adattárolás pontos verzióját azonosító Univerzális Numerikus Ujjlenyomat (*Universal Numerical Fingerprint*; Altman – King, 2007) alkalmazásával segíti elő.

A csillagászatban nem csupán a nagy programoknak van lehetőségük az adatok szabványos megosztására. A már említett Virtuális Obszervatórium projekt keretében kifejlesztett, szabadon alkalmazható, nyílt forráskódú szoftverek – mint például a Strasbourgban kifejlesztett Saada – minden obszervatóriumnak lehetőséget adnak adataik megosztására. A szoftver által alkalmazott protokoll

lok mellett az adatformátumok is szabványosak, és rendelkezésre állnak azok a listák – *registry* a VO terminológiájában –, amelyek felsorolják a világszerte elérhető hasonló szolgáltatásokat. Rendelkezésre állnak olyan információs rendszerek is, mint a szintén strasbourgi Aladin, amelyek képesek adatokat – képeket – átvenni ezektől a szolgáltatóktól. Bár a csillagászatban rendelkezésre állnak ezek a lehetőségek, sokan mégsem használják őket. Az adatok megfelelő tisztítása, dokumentálása ugyanis jelentős teher; a kutatók gyakorlatilag csak annyit végeznek el, amennyi a saját publikációjuk megírásához szükséges, az általános felhasználhatósághoz szükséges többletmunkát már nem vállalják.

A kutatók jelentős része nem szívesen osztja meg adatait, ha ezt nem követelik meg tőle. Aki megkövetelhetné, az a munkáltatója vagy a munkáját támogató kutatási alap lehet. Megkövetelhetik még azok a folyóiratok, ahol publikál – nem annyira az esetleges újrafelhasználást, mint inkább az eredmények igazolását, reprodukálhatóságát szem előtt tartva. Ritkán fordul elő, hogy maga a tudományos folyóirat vállalkozna a cikkhez tartozó adatok tárolására és nyilvánosságra hozatalára. De az internet által kínált lehetőségeket úttörő módon használó kibővített folyóiratok ezt megteszik. Az MTA Csillagászati és Földtudományi Kutatóközpontjának Konkoly Thege Miklós Csillagászati Intézetében (CsFK CsI) kiadott IBVS (*Information Bulletin on Variable Stars*) tárolja a cikkekhez tartozó adatokat, és kereshetővé is teszi ezeket.

A hagyományosabb folyóiratok esetében is előfordul egy némileg szerényebb igényű megoldás. Általában a lapok nem hajlandóak annyi ábrát, táblázatot közölni, mint amennyit a szerzők publikálni szeretnének. Régebben egyes folyóiratok kiegészítő folyamokat

(*Supplement series, Ergänzungshefte*) indítottak a nagyobb mennyiségű adat közlésére, manapság pedig előfordul, hogy a nyomtatott változat lehetőségeit már meghaladó táblázatokat csak elektronikus formában közlik. Ugyanez kevésbé fontos ábrákkal is előfordulhat.

Abban az esetben, amikor a folyóirat nem tud helyet adni nagyszámú táblázat vagy ábra elektronikus elhelyezésének, a táblázatos vagy ábra (kép) jellegű adatok elhelyezésére szóba jöhetnek a publikációs repozitóriumok is. Ezek a repozitóriumok többnyire könyvtárakban működnek, és többnyire nem áll rendelkezésre se szakértelem, se technika bonyolult (mondjuk bináris, sokdimenziós, egyedi formátumú) adatok kezelésére. Nagy mennyiségű (gigabyte [GB] -nyinál nagyobb) adat elhelyezésére nem igazán alkalmasak a publikációs repozitóriumok. Viszont egyszerű, ismert, szabványos formátumokban (PDF/A, JPEG) elhelyezett, alapvetően szöveges, táblázat jellegű vagy az értelmezéshez feldolgozó szoftvert nem igénylő képi adatok elhelyezése megfelelnek. A kívánatos munkamegosztás és munkamenet az lehet, hogy a szerző (talán a beküldött kézirattal együtt) a repozitóriumban elhelyezi a táblázatokat, ábrákat, és az URL-t, az azonosítót közli a szerkesztővel, akik továbbadják ezt az információt a cikk bírálóinak. A bírálók a cikkkel együtt vizsgálják az adatokat. A cikk elfogadása esetén az újság az olvasókkal is közli az adatok elérhetőségét. Hasonló funkció implementálásán gondolkodunk a REAL (az MTA

Könyvtár és Információs Központ az Akadémia kutatóhálózatát is kiszolgáló repozitórium) esetében is.

Az adatok megosztását felhő alapú alkalmazások is segíthetik, mint a figshare (URL18). A feltöltött állományok DOI-t kapnak, hosszú távú megőrzésüket a CLOCKSS Archívum (URL19) garantálja. Publikációs repozitóriumokhoz hasonlít a Dryad (URL20), és csak publikációkhoz kapcsolódó adatokat tárol. A holland DANS, a felsőoktatás és kutatás hálózati és digitális archiválási feladatait szolgáló szervezet, adatrepozitóriumuk webes felülete az EASY (URL21). A tárolt adattárolások jelenleg a történelemtudomány, régészet, társadalomtudományok köréből kerülnek ki, valamint tartalmaz térképészeti-kataszteri adatokat is. Az Egyesült Királyságban már évek óta működik, gyűjt és szolgáltat régészeti adatokat az Archaeology Data Service (URL22).

A tudományos publikációk digitális formában való hozzáférhetővé tétele és hosszú távú megőrzése sem könnyen megoldható probléma – de ebben jó úton haladunk a megoldás felé. A tudományos adatok esetében az akadályok nagyobbak, a megoldások jelenleg még kevésbé beláthatóak. De mint fentebb is bemutattuk, egyes területeken már vannak működő rendszerek. Nem szabad a nehézségektől visszarettenni, el kell indulni a tudományos adatmenedzsment útján, előre kell lépni, ahol lehet.

Kulcsszavak: *tudományos adatok, Open Access*

## IRODALOM

- Altman, M. – King, G. (2007): A Proposed Standard for the Scholarly Citation of Quantitative Data. *D-Lib Magazine*. 13 March, 3–4, • <http://www.dlib.org/dlib/march07/altman/03altman.html>
- Holl András (2013): Információáradat és hullámvaglás. *Magyar Tudomány*. 4, 473–478. • <http://www.matud.iif.hu/2013/04/13.htm>
- URL1: FORCE II <http://www.forceii.org/>
- URL2: <http://www.whitehouse.gov/sites/default/files/microsites/ostp/scientific-integrity-memo-12172010.pdf>
- URL3: Alliance for Permanent Access <http://www.alliancepermanentaccess.org/>
- URL4: SIM4RDM <http://www.sim4rdm.eu/> <http://videotorium.hu/hu/recordings/details/4158>, Tudományos\_Celu\_Adatmenedzsment\_Europaban\_-\_SIM4RDM\_projekt
- URL5: SIM4RDM Recommendations for support of research data management <http://www.sim4rdm.eu/sites/default/files/uploads/documents/Recommendations%20for%20support%20of%20research%20data%20management.pdf>
- URL6: OAIS ISO 14721:2003 [http://www.iso.org/iso/iso\\_catalogue/catalogue\\_tc/catalogue\\_detail.htm?csnumber=24683](http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=24683)
- URL7: OpenAireplus <http://www.openaire.eu/en/component/content/article/76-highlights/326-openaireplus-press-release>
- URL8: Knowledge Exchange <http://www.knowledge-exchange.info/>
- URL9: The Value of Research Data <http://www.knowledge-exchange.info/datametrics>
- URL10: DataCite <http://www.datacite.org/>
- URL11: Data Citation Index [http://wokinfo.com/products\\_tools/multidisciplinary/dci/](http://wokinfo.com/products_tools/multidisciplinary/dci/)
- URL12: IVOA <http://www.ivoa.net/>
- URL13: EJME - Enhanced Journals...Made Easy! <http://www.surf.nl/en/projecten/Pages/EJME.aspx>
- URL14: GreyNet EPP [http://www.greynet.org/open\\_greyrepository/enhancedpublications.html](http://www.greynet.org/open_greyrepository/enhancedpublications.html)
- URL15: University of Edinburgh Research Data Management Roadmap [http://www.ed.ac.uk/polopoly\\_fs/1.101223!/fileManager/UoE-RDM-Roadmap\\_20112102.pdf](http://www.ed.ac.uk/polopoly_fs/1.101223!/fileManager/UoE-RDM-Roadmap_20112102.pdf)
- URL16: DASPOS <https://daspos.crc.nd.edu/> <https://indico.cern.ch/getFile.py/access?contribId=2&resId=0&materialId=2&confId=191826>
- URL17: Dataverse Network Project <http://thedata.org/>
- URL18: figshare <http://figshare.com/>
- URL19: CLOCKSS <http://www.clockss.org/clockss/Home>
- URL20: Dryad Digital Repository <http://datadryad.org/>
- URL21: DANS EASY <https://easy.dans.knaw.nl/ui/home>
- URL22: ADS <http://archaeologydataservice.ac.uk/>



## A GYÓGYÍTÓ KÉP

Hárdi István

a pszichológia-tudomány kandidátusa, c. főiskolai tanár  
ihardi@freemail.hu

*In herbis, verbis et lapidibus magna vis est.*  
(A növényekben, szavakban  
és a kövekben nagy erő van.)

Századunkat sokan a kép századának tartják. Elég, ha a televízióra, a digitális fényképezés elterjedésére gondolunk. A hétköznapi életben is ismerjük: „Ez a fénykép véletlenül nálam van...” – és a szülők, nagyszülők büszkén mutatják gyermekeik, unokáik, szeretteik képmását. Ismert mondás a bennünk élő képről: „képeddel alszom el, képeddel ébredek...” Kép a polcon – nemegyszer az elvesztett emléket és személyét pótolja. Ezt sokszor a magatartás is alátámaszthatja (például valaki csókolgatja a képet).

Hogy jutott el az ember a kép (vagy szobor) gyógyító hatásához? – érdemes ezt a lélek orvosának szemléletével áttekinteni.

Sigmund Freud szerint (1935, 1985) a kép (emlék-, álom-, fantáziakép stb.) a lélek legősibb tartalmi közé tartozik. A kép régibb, mint a szó, akár a barlangrajzokra gondolunk, amik sok mindent elárulnak a huszonöt-harmincezer évvel ezelőtt élt őseinkről, akár az írás történetére, amely képirással kezdődött (piktográfia). Az ember, amikor a keze nyomát hagyta a falon, embert és állatot, vagy vadászatot vésett, karcolt a falba – művei nem csupán a tényeket s az eseményt rögzítették. Az ősember – akiben az én és a külvilág ha-

tárai elmosódtak – az őt körülvevő környezetet „belelkesítette”, saját érzéseivel, vágyaival és félelmeivel töltötte meg. Ez az animisztikus szemlélet az ábrázolt állatot a valóssal azonosnak vélte, sőt ha ábrázolta („falra festette az ördögöt”) „meg is idézhette” – ez a mágia. Az ilyen *animisztikus-mágikus szemlélet* sokáig uralkodott, például az ókorban Apellész a nagy görög művész festményeinek lovait olyan élethűnek tartották, hogy azok nyerítését is hallani vélték. Ilyen természetű Pygmalion (Pügmalió) mitológiai története: Ovidius szerint a szobrász olyan szép leányt alkotott, hogy bele is szeretett. Aphrodité megkönyörült rajta, és életre keltette. A történetnek számos régi és új változata van (hogy csak a legújabbakra utalj: a *Pigmalion*-tól a *My Fair Lady*-ig).

A mágikus-animisztikus hit az alapja számos középkori varázslatnak: az elpusztítandó személyt ábrázoló bábút átszúrták egy tűvel. Ennek nemegy formájával ma is találkozunk a politikai ellenség *in effigie* megsemmisítésében, a képmás elégetésével őt is elpusztítják. Természetesen a mai modern világban a tömegek tudatos manipulálása, a népharag felkeltése is állhat mögötte, de sokakban ez tudattalanul mágikus-animisztikus hatást is kelt. Nehéz élethelyzetekben – főként nagyobb szorongásos állapotokban (például súlyos betegségben, halálveszedelemben) –