

## KUTATÁSI ADAT-KEZELÉS A CSILLAGÁSZAT TERÜLETÉN

Holl András

ORCID: [0000-0002-6873-3425](https://orcid.org/0000-0002-6873-3425)

MTA Könyvtár és Információs Központ

A cikk bemutatja egy tudományterület kutatási adat-kezelési gyakorlatát. A csillagászat élen jár a kutatási adatok megosztásában – érdemes tanulmányozni az évtizedek alatt elért eredményeket.

### 1.) Fejlett adatkezelés – okok, történet

A csillagászat elöl jár a kutatási adatok kezelésének területén.<sup>1</sup> Ennek a ténynek az okát részben a tudomány erősen haszon-talan (non-profit) jellegében kell keresnünk. Részint abban, hogy az égi jelenségeket nemhogy kisajátítani nem lehet, de gyakorta szükség van egymástól távoli földrajzi helyű megfigyelők együttműködésére, vagy különböző megfigyelési eszközök adatainak egyesítésére.

Az obszervatóriumok közötti adatcsere igénye hozta létre a Flexible Image Transport System (FITS) nevű adatszabványt 1981-ben (Greisen, 2003). A szabványos adatformátum léte, de még inkább azon tulajdonsága, hogy a rugalmasan bővíthető, de mégis szabványos metaadatok magában az adatfájlban kaptak helyet, kínálta a lehetőséget arra, hogy az (amúgy nyílt forráskódú, szabad) feldolgozó szoftverek is támogassák. Mindennek az lett az eredménye, hogy a teljes adatfeldolgozási folyamatba beépült a részletes dokumentáció (Holl, 2020).

Az adatok megosztása, újrafelhasználása egyre nagyobb jelentőségre tett szert a megaprojektek korában. Fél évszázaddal ezelőtt a kutatási adatokból kevés volt, az adatok összegyűjtése (megfigyelés) sok időt és energiát emésztett fel. A megosztás praktikus lehetősége az adatpublikálás volt.

---

1 „Astronomy has been a pioneer of Open data sharing, and remains at the forefront.” Turning FAIR into reality. European Union, 2018. doi: [10.2777/1524](https://doi.org/10.2777/1524)

Publikálás előtt a megfigyelési adatokat gondosan őrizték, és amikor publikálták (intézeti kiadványsorozatban, katalógusban), akkor az hivatkozhatóvá is vált. Az űrobszervatóriumok (csillagászati megfigyelést végző szatelliták, mint a *Hubble Űrteleszkóp*, *Kepler*, *Gaia*) és a földfelszíni nagyprojektek (pl. *Sloane Digital Sky Survey*) térnyerése adatbőséget teremtett a csillagászatban. Immár több az adat, mint amennyit a pályán lévő csillagászok belátható idő alatt fel tudnak dolgozni. Az adatbőség teremtette meg a szükségességét annak, hogy az adatok dokumentáltan, mind könnyebben hozzáférhetően – természetesen digitális formában – adatbázisokba kerüljenek. A Hubble Űrteleszkóp adatain alapuló publikációk többsége immár az archívum felhasználásával készül (Novacescu et al., 2018, 1. ábra).

## 2.) Adatmegosztás: a Virtuális Obszervatórium

Virtuális obszervatóriumként működhetnek a megfelelően dokumentált és hozzáférhetővé tett kutatási adatok: esetenként nem kell új (drága) megfigyeléseket végezni, mivel a meglévő, kutatható adatok is képezhetik a kutatás alapját. Az Egyesült Államok *National Virtual Observatory* kezdeményezésének alapidokumentumát 2002-ben adták ki.<sup>2</sup> Ugyanebben az évben hozták létre a Nemzetközi Virtuális Obszervatórium Szövetséget (*International Virtual Observatory Alliance*).

Nem kellett sokat várni az NVO első eredményeire. Egy 2003-as demonstráció során, amikor két nagy katalógus (SDSS és 2MASS) adatait párosították egy szoftverprototípus segítségével, felfedeztek egy új barna törpecsillagot. Az új technikával a két hatalmas adathalmazban való keresés (és találat) mindössze két percet vett igénybe – ez korábban több hétnyi vagy hónapnyi emberi munkát igényelt volna.<sup>3</sup>

A VO húsz éve alatt elsősorban szabványokat teremtett, de létrejöttek szoftverek és adatközpontok által nyújtott VO szolgáltatások is. Ezek közül mutatunk be néhányat.

2 Towards the National Virtual Observatory. <http://www.virtualobservatory.org/documents/sdt-final.pdf>

3 Johns Hopkins University. „Virtual Observatory Prototype Produces Surprise Discovery; Early Demo Project Identifies New Brown Dwarf.” *ScienceDaily*. 12 March 2003. <http://www.sciencedaily.com/releases/2003/03/030312071232.htm>

A strasbourgri CDS-ben készült az *Aladin* égbolt-térkép szoftver, melynek van webes szolgáltatásként működő, személyi számítógépre telepíthető és böngészőben futó javascript változata is. A *Google Maps*-hez hasonlítható, csak nem a Földet, hanem az eget ábrázolja. Számos nagy, teljes égboltot lefedő felmérés képeit vagy katalógusait meg lehet vele jeleníteni, de ki-ki saját képeit is megnézheti a „beépített” térképekre vetítve. A hazai *Information Bulletin on Variable Stars* (IBVS) folyóirat is használta megjelenítő eszközként: mind egy égbolterület több színben fotometrált csillagainak adatait tartalmazó táblázatot, mind egy kis égterület fényképét meg lehetett jeleníteni a segítségével. Az olvasó csak kattintott egyet, és a folyóiratcikkhez tartozó adatok megjelentek a felbukkanó térképi alkalmazásban (Holl, 2022).

Ahhoz, hogy az *Aladin* segítségével „privát” adatokat fel lehessen tenni a térképre, szabványokra van szükség. Az egyik ilyen a táblázatos adatokban található mennyiségek szabványos leírására szolgáló Universal Column Descriptor, a másik maga a szabványos táblázat formátum: a VOTable.

A gyors lefutású égi események felfedezése esetén a közösség riasztására szolgáló protokoll a VOEvent. Például a gravitációshullám-obszervatóriumok ilyen híradásokat adnak ki azért, hogy az általuk megfigyelt események elektromágneses sugárzását a készenléletben álló teleszkópok percek múlva megkísérelhessék detektálni (Williams és Seaman, 2006).

A csillagászatban korán létrejött tudományterületi szabványok egyike volt a közlemények (de akár publikált adatok) azonosítására szolgáló BIBCODE. Viszont ennek az azonosítónak a használata esetenként késleltette a DOI bevezetését.

Az International Virtual Observatory Alliance weboldala jelenleg 58 szabványjellegű dokumentumot tartalmaz.<sup>4</sup> A szervezet tevékenységét Berriman et al. (2020) foglalja össze.

---

4 <https://www.ivoa.net/documents/>

### 3.) IBVS

A már említett IBVS adatközlő folyóiratként is működött. Közölt adat-cikkeket, melyekben az adatok a cikkben foglalt táblázatokban a nyomtatott/PDF cikk részét képezték, más esetekben pedig a cikk digitális mellékleteként voltak az adatok a webről letölthetőek (1995 után). Ez utóbbi adatfájlok egyszerű szerkezetűek voltak, gépi feldolgozhatóságra alkalmas formában. A fentebb említett, CDS Aladinban való megjelenítésre szánt adatfájloknak elérhető volt egy VOTable változata is.

Az IBVS már nem aktív folyóirat – az utolsó cikkek 2019-ben jelentek meg. A cikkek linkekkel gazdagított HTML változatai sem érhetőek már el, csak a PDF-ek. A többnyire egyszerű, szöveges állományok formájában elhelyezett adatfájlok viszont elérhetőek továbbra is a folyóirat honlapján, és külső archívum(ok)ban is.

### 4.) Az élenjáró és a kevésbé fejlett területek

A kutatási adatok FAIR követelmények szerinti kezelése munkaigényes. A megfelelő szintű dokumentálás legkönnyebben a nagy projektek (például űrobszervatóriumok vagy nagy költségvetésű földfelszíni projektek) esetében oldható meg. Ezekben az esetekben az adatfeldolgozás automatikus folyamatban történik, így a megfelelő dokumentálás, gazdag metaadatkészlet biztosítható.

Megoldható a FAIR adatkezelés a nagy adatközpontokban is – ezeknél rendelkezésre áll a költségvetés, és az adatokkal képzett adatgazdászok (data steward) foglalkoznak. A csillagászatban ilyen adatközpont például a *Centre Données de Strasbourg*, az *ESAC Science Data Centre*, a *Mikulski Archive for Space Telescopes* és az *ESO Science Archive Facility*. Külön említendő a NASA támogatásával működő *Astrophysics Data System*, amely ugyan bibliográfiai adatbázis, de nyilvántartja a cikkekhez kapcsolódó adatjellegű forrásokat is.

Ha a kis költségvetésű projektek adatait vizsgáljuk, az adatok nyilvánosságának, dokumentáltságának helyzete már korántsem ilyen jó. Bár a kutatóknak adódnak lehetőségei az adatok látható közzétételére – ritkábban a folyóiratoknál digitális mellékletekként, gyakrabban a generalista repozitóriumokban mint a *Figshare* vagy a *Zenodo*, alkalmanként

intézményi adatrepozitóriumban, de ezekkel a lehetőségekkel ritkán élnek. Talán azért, mert munkaiigényes az adatokat közzétehető formába hozni, talán mert hiányzik a motiváció, nincs meg a megfelelő ösztönző rendszer. A szakterület legfontosabb folyóiratainak egyike, az európai *Astronomy & Astrophysics* szerzőknek szóló útmutatójában leírja, hogy a cikkekhez kapcsolódó elsődleges adatok elhelyezését a CDS végzi, és az archiválható állományok létrehozásában szükség esetén segítenek is. Mindazonáltal adatelhelyezési kötelezettséget az útmutató nem említ.

Egy hazai kutatóhely éves publikációit végignézve az ADS-ben arra a következtetésre juthatunk, hogy bár a cikkek többségénél az adatbázis jelöl valamiféle adatkapcsolatot, ezek többségét a CDS által a cikkekből kigyűjtött objektumnevek, esetleg táblázatok adják. A hazai csillagászati kutatóhelyek esetében egyre nő a nagy projektek adatait használó cikkek száma – ezeknél a megfigyelési adatok elérhetősége és újrafelhasználhatósága megoldott, így a cikkhez használt adatok is elérhetőek. A hazai teleszkópokkal végzett észlelések adatai viszont kevésbé felelnek meg a FAIR kritériumoknak, de ez gyaníthatóan így van világszerte más kisebb obszervatóriumok esetében is.

## 5.) A könyvtárak szerepe

Kis mennyiségű adat a közleményekben megjelentethető. Konkoly Thege Miklós rendszeresen közölte az ógyallai csillagvizsgálójában végzett megfigyeléseinek adatait az Akadémia kiadásában, például az „Értekezések a matematikai tudományok köréből” sorozatban. Az – akkori mércével – sok adatközlés úgy tűnik, terheket jelentett az Akadémiának: Konkoly levélben panasolta el Eötvös Lorándnak az őt ért kritikákat (Vargha 2001).

Konkoly obszervatóriuma a későbbiekben is adott ki kötetekben megfigyelési anyagot (Tass 1925). A publikált megfigyelési anyagok papíron megtalálhatóak a könyvtárakban, digitalizálva pedig a könyvtárak által üzemeltetett repozitóriumban.

A XX. század második felére az obszervatóriumi kiadványsorozatok szerepe csökkent (bár a Konkolyról elnevezett hazai obszervatórium kiadványsorozatában még ez időszakban is sok adatot közöltek). Az adatok jobbára az ekkorra uralkodóvá vált folyóiratok mellékleteibe

(*Supplement Series, Ergänzungshefte*) szorultak. 1993 és 1998 között az *American Astronomical Society* folyóirataihoz CD-ROM mellékletet adtak ki, ezután már az adatok a hálózatra kerültek. CD-ROM-os formában indult a *Journal of Astronomical Data*, majd ez a kiadvány is a webre került.<sup>5</sup>

A Gutenberg-korszakban keletkezett, ránk maradt megfigyelési eredmények jelentős része a könyvtárak polcain, és papírról digitalizálva a repozitóriumokban találhatóak, ám ezek az adatok gépi feldolgozásra nem alkalmasak (átalakítás nélkül). Mint fentebb láthattuk, a kisebb volumenű, modern adatok is legtöbbször a folyóiratok közvetítésével kerülnek a nagy adatközpontok adatbázisaiba.

Az adatok „könyvtári útjának” fontosságát mutatja, hogy az adatokat (szakirodalmi kapcsolataikkal együtt) kereshetővé tévő adatbázisok is szoros kapcsolatban állnak a könyvtárakkal. A NASA/SAO ADS átfogó tudományterületi bibliográfiaként teszi láthatóvá a cikkekhez kapcsolódó adatokat, míg a CDS erős könyvtáros csoportot fenntartva, az adatok oldaláról mutatja meg az adatokat említő tudományos cikkeket.

– *Library and Information Services in Astronomy (LISA)*

A csillagász könyvtárosok konferenciasorozata a LISA 1988-ban indult, a legutóbbi, kilencedik összejövetel 2021-ben volt (a járványhelyzet miatt online). A konferenciák visszatérő témája a kutatási adatok kezelése, a nyílt tudomány. A konferenciák kiadványköteteiben a kutatási adatok kezelésével foglalkozó cikkek száma az elmúlt két évtizedben a 2002-es hétről tizenötre nőtt.

A legtöbb esetben külön szekció foglalkozott a kutatásiadat-kezeléssel, 2017-ben annyira hangsúlyos volt a téma, hogy három szekciót töltöttek meg az adatkezeléssel kapcsolatos előadások.

– *Tudománymetria – adathasználat követése az ESO-ban*

Az Európai Déli Obszervatórium könyvtára vállalta fel az intézményben végzett megfigyelések publikálásának szakirodalmi követését, mérését. A *telbib* projekt szakirodalmi könyvtárosi ellenőrzéssel megerősített

---

5 JAD: <http://journalofastronomicaldata.be/>

adatbányászattal követi a távcsőidő-pályázati azonosítók szereplését a publikációkban (Grothkopf, 2018).

## 6.) Közösségi tudomány

Ez a terület az, ahol a kívülálló leginkább felmérheti a csillagászati kutatási adat-kezelés helyzetét. A csillagászati közösségi tudomány hazai előképe Konkoly Thege Miklós meteorészlelő hálózata 1875-ből (Bartha, 1988). 1911-ben alakult az *American Association of Variable Star Observers*, többségében amatőr csillagászok megfigyeléseinek szervezésére, az adatok összegyűjtésére.

A modern közösségi tudomány egyik legfontosabb platformja a *Zooniverse* portál. A névadó első alkalmazás, a *Galaxy Zoo* önkénteseinek száma jelenleg közel 87 ezer. Nem valószínű volt meg a projekt az SDSS és a többi időközben bevont égbolttelmerés adatainak nyilvánosságra hozatala nélkül.

Még korábbi az emberi agy jelfeldolgozó képességei helyett az otthoni számítógépek szabad processzorkapacitását felhasználó *SETI@home*.

## 7.) Összefoglalás

Negyven évnyi kooperatív kutatási adat-kezelési gyakorlat birtokában a csillagászat ma olyan tudomány, ahol saját megfigyelések nélkül, kizárólag publikusan elérhető adatokra alapozva lehet tudományos kutatást folytatni. A csillagászatban (és egyre több más tudományterületen is) sok a fókusz [az adat], és kevés az eszkimó [a kutató]. Chris Lintott nyilatkozta a *Time* magazinnak: „In many parts of science, we’re not constrained by what data we can get. We’re constrained by what we can do with the data we have.”<sup>6</sup> Ahhoz, hogy a digitális korszakban keletkező adatáradatot a feladatra vállalkozók feldolgozhassák, megfelelő dokumentáltságra, és nyilvánosságra van szükség.

---

6 <https://web.archive.org/web/20100331061938/http://www.time.com/time/health/article/0,8599,1975296,00.html>

## Irodalom

- Bartha L., 1988. Az első magyarországi észlelőhálózat, Meteor, 7/8. 7–11.
- Berriman G.B. et al., 2020. The International Virtual Observatory Alliance (IVOA) in 2020. [arXiv:2012.05988](https://arxiv.org/abs/2012.05988) [[astro-ph.IM](https://arxiv.org/abs/2012.05988)]
- Greisen, E.W., 2003. FITS: A Remarkable Achievement in Information Exchange. In: Heck, A. (eds) Information Handling in Astronomy - Historical Vistas. Astrophysics and Space Science Library, vol. 285. Springer, Dordrecht. [https://doi.org/10.1007/0-306-48080-8\\_5](https://doi.org/10.1007/0-306-48080-8_5)
- Grothkopf, U., Meakins, A., Bordelon, D., 2018. ESO telbib: learning from experience, preparing for the future. [ArXiv:1806.08746](https://arxiv.org/abs/1806.08746) [[astro-ph.IM](https://arxiv.org/abs/1806.08746)] <https://doi.org/10.1117/12.2311667>
- Holl, András, 2020. A kutatási adatok dokumentálását elősegítő szoftverek. In: Networkshop 2020. Országos Online Konferencia. 2020. szeptember 2–4. HUNGARNET. pp. 7–12. <http://real.mtak.hu/119187/> <https://doi.org/10.31915/NWS.2020.1>
- Holl, András, 2022. IBVS Data Files – Case Study of a Small Data Journal. Bulletin of the AAS, 54 (2). <https://doi.org/10.3847/25c2cfef.fl4d187b>
- Novacescu, J. et al., 2018. Elevating MAST-Data Publications with Digital Object Identifiers (DOIs). Jenny EPJ Web Conf., 186, 10003. <https://doi.org/10.1051/epjconf/201818610003>
- Tass, Antal, ed., 1925. Photometric Observations of Variable Stars = Photometrische Beobachtungen Veranderlicher Sterne = Változó csillagok fotometrikus megfigyelései. Publications of the Royal Hungarian Astrophysical Observatory, Foundation of Konkoly in Budapest, 2. Konkoly-Alapítványú Budapesti Magyar Királyi Csillagvizsgáló Intézet, Ogyalla – Budapest. <http://real-eod.mtak.hu/9417/>
- Vargha Domokosné: Konkoly Thege Miklós magyar nyelvű írásai. Magyar Tudomány, 2001/7. 867. o. <http://www.matud.iif.hu/01jul/vargha.html> [http://real-j.mtak.hu/158/1/MATUD\\_2001.pdf#Page=895](http://real-j.mtak.hu/158/1/MATUD_2001.pdf#Page=895)
- Williams, R. D., Seaman, R., 2006. VOEvent: Information Infrastructure for Real-Time Astronomy. ADASS XV. <https://ui.adsabs.harvard.edu/abs/2006ASPC..351..637W/abstract>