



# Adatok a górcső alatt 1. *Helyzetkép*

2023.04.13.

Horváth Anna / Kutatási Dokumentációs Központ /  
Társadalomtudományi Kutatóközpont

[horvath.anna@tk.hu](mailto:horvath.anna@tk.hu)

# Kutatói igényfelmérés

- Cél: az adatrepozitórium leendő használói jelenlegi adatkezelési; - tárolási; és –megosztási gyakorlatainak és igényeinek megismerése
- 2022. február-május
- Két ütemben:
  - Február 1.-március 31.: kérdőíves felmérés
    - Online kérdőíves felmérés, önkéntes válaszadás, anonim kérdőív
    - 129 kitöltés (17 intézmény)
  - Március-május: interjú felmérés
  - A két módszerrel teljes lefedettség



# Interjú igényfelmérés: áttekintés

- 17 ELKH-s intézmény (kutatóközpont és intézet), 51 interjú
- Összesen 55 interjúalany, legnagyobb részt kutatók, néhány kivétel
- Minél nagyobb tudományterületi és szakterületi lefedettség elérése
- 2022. március-május



# Az interjúk intézmények szerinti megoszlása

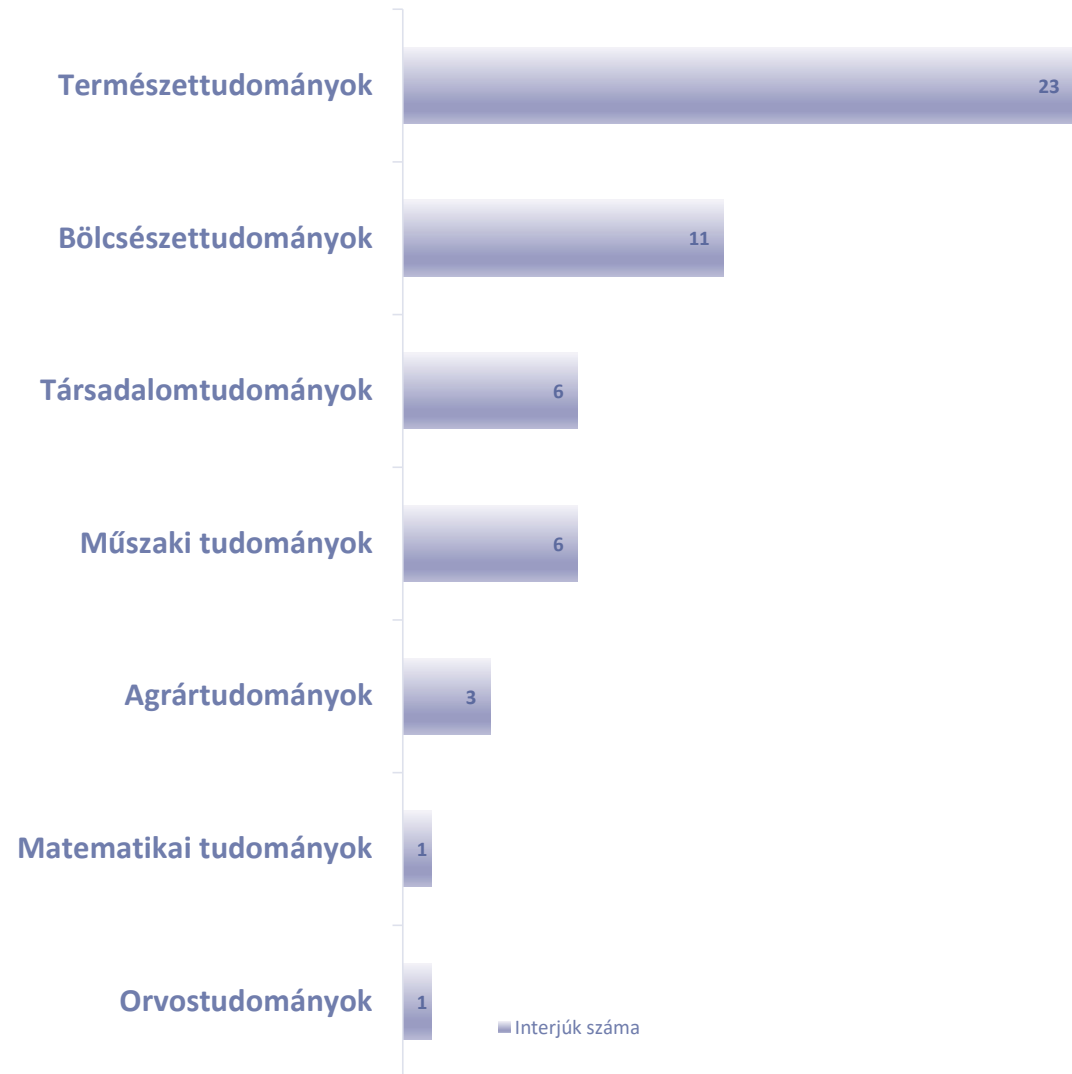
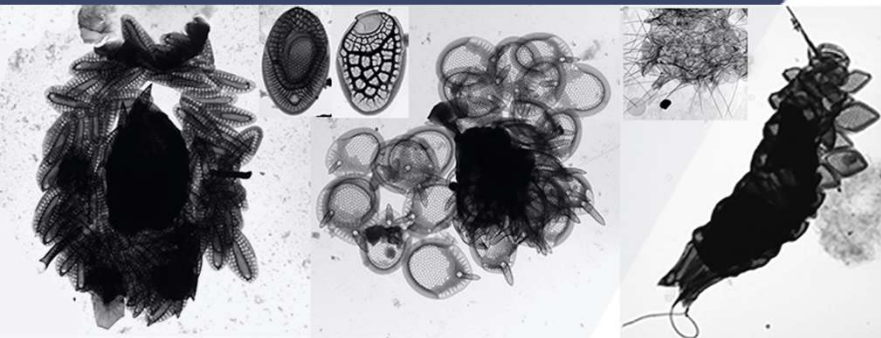
Interjúk száma ~  
szakterületek száma



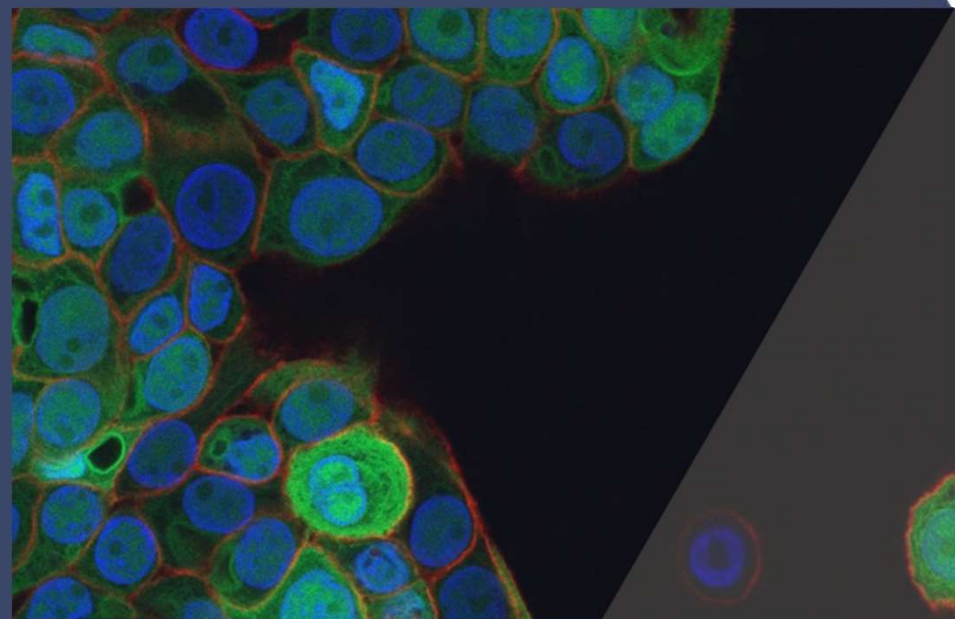
Intézmény	N
Agrártudományi Kutatóközpont (ATK)	3
Atommagkutató Intézet (ATOMKI)	1
Baltoni Limnológiai Kutatóintézet (BLKI)	1
Bölcsészettudományi Kutatóközpont (BTK)	7
Csillagászati és Földtudományi Kutatóközpont (CSFK)	5
Energiatudományi Kutatóközpont (EK)	5
Földfizikai és Űrtudományi Kutatóintézet (FI)	1
Kísérleti Orvostudományi Kutatóintézet (KOKI)	1
Közgazdaság- és Regionális Tudományi Kutatóközpont (KRTK)	4
Nyelvtudományi Kutatóközpont (NYTK)	5
Ökológiai Kutatóközpont (ÖK)	6
Rényi Alfréd Matematikai Kutatóintézet (RÉNYI, RI)	1
Számítástechnikai és Automatizálási Kutatóintézet (SZTAKI)	1
Szegedi Biológiai Kutatóközpont (SZBK)	1
Társadalomtudományi Kutatóközpont (TK)	1
Természettudományi Kutatóközpont (TTK)	5
Wigner Fizikai Kutatóközpont (WIGNER FK)	3
Összesen	51



# Az interjúk (nagy) tudományterületek szerinti megoszlása



# Tapasztalatok : lefedettség



- Adatrepozitóriumot már használt/használó/használni tervező, érdeklődő interjúalanyok
- Tapasztalatokkal kevésbé rendelkezők meglátásai kevésbé jelentek meg
- **Nem reprezentatív**
- Az elhangzottak elsősorban a megkérdezett kutatók és/vagy kutatócsoportjuk gyakorlataira világítanak rá



# Kérdéskörök

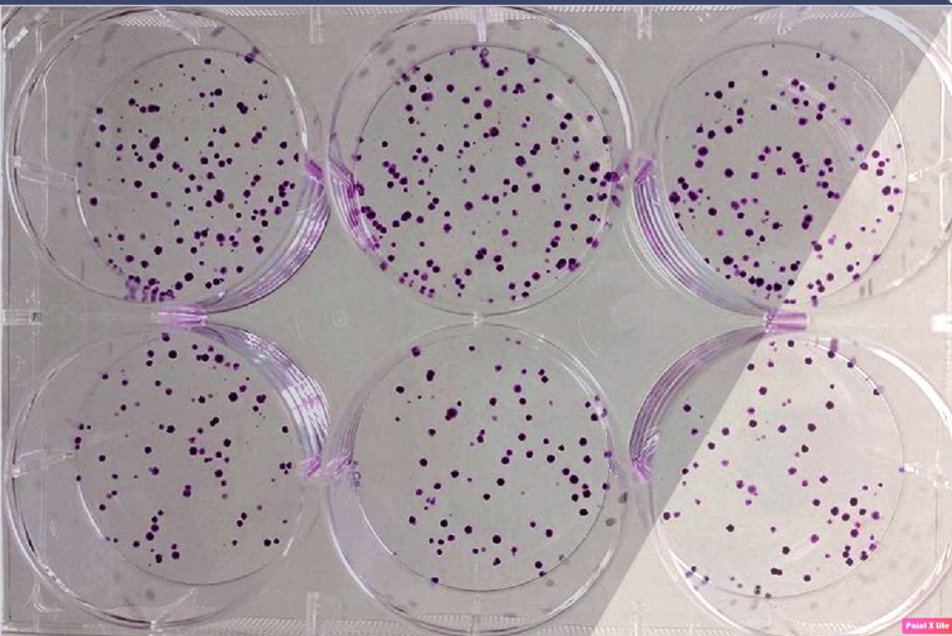
- Saját kutatási témák, területek
- Kutatások indulása
- Tipikus adat-életciklus
- Felmerülő problémák
- Az ARP lehetséges belépési pontjai
- Tendenciák, jövőkép



# Eredmények: Kutatási adatok



- Jellemzően keletkeznek új adatok a kutatások során
  - Az előállított új adatok főbb típusai:
    - megfigyelések / kísérletek / szimulációs eljárás során létrehozott
    - meglévő adatforrásokat felhasználó, feldolgozó
- Az adatok mérete:
  - Mérésenként 10 KB-ostól a több terabájtosig (nyers adatok)
  - Változó, hogy a feldolgozott vagy a nyers adatok mérete, mennyisége a nagyobb
- Az adatok formátuma
  - Kézi jegyzettől a szoftverprogramozásig
  - Legjellemzőbb a sokféle fájlformátum együttes használata





# Eredmények: Kutatási adatok kezelése a kutatás idején

- Tárolás:
  - Leggyakrabban nem egy helyen: saját gép/megosztásra alkalmas felület
  - Jellemzően saját, kevésbé intézményi megoldások
- Megosztás:
  - Kutatócsoport tagjai között
  - Ritka, de vannak kutatók, akik senkivel nem osztanak meg adatot



*„Az adatokat alapvetően mi tároljuk. Egy darabig magán a mérőgépen. Aztán amikor ugye mondjuk arról a mérés befejeződik, akkor azt föltöltjük egy laborszerverre, azt már ugye a kiértékelő gépekről el tudjuk érni. És a külsős partnerek, azok hát ugye egyedileg mondjuk letöltik úgymond az adatokat, tehát őket nem engedjük be a mi infrastruktúránkba, hanem kapnak mondjuk egy ilyen felhőben egy letöltési linket. [..] Onnan letöltik az adatokat, és aztán azzal akármit is csinálhatnak.”*

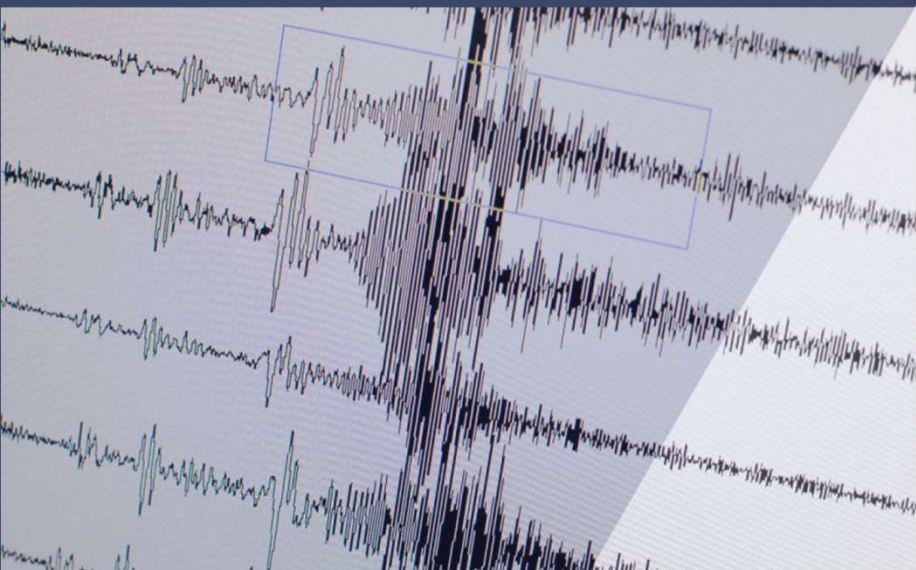
# Eredmények: Kutatási adatok kezelése a kutatás lezárását követően



- Tárolás:
  - A legtöbb kutatónak saját, kialakult gyakorlata van
  - Sok helyen használnak valamilyen közös tárolást a kutatást követően is
  - Felhőalapú megoldások népszerűek (sokszor az intézmény biztosítja)
  - Néhány esetben nem digitális megoldások is születnek
- Megosztás:
  - Több tudományterületen tapasztalták, hogy kérik az adatokat a publikációkhoz
  - Általánosan elterjedt a megosztás, 40% az adatai egy részét vagy egészét globálisan is kereshetővé teszi
  - Vannak jogi akadályok, melyek gátat jelentenek
  - Repozitóriumhasználat nem elterjedt



# A kutatási adatok metaadatokkal történő ellátása



- A kutatók egy része nem ismeri a metaadat kifejezés jelentését
- Ennek ellenére sokan ellátják metaadatokkal a saját kutatási adataikat
- A repozitóriumhasználó kutatók ismerték általában a metaadat és metaadat-séma fogalmakat
- A legtöbb kutató adatai nem megtalálhatóak és kutatásuk nem reprodukálható a metaadatok alapján



# A kutatási adatok újrafelhasználása és adatvesztés



- A legtöbb kutató a saját adatait használja újra (amennyiben újrafelhasznál adatokat)
- Néhányan használnak mások által keletkeztetett adatot, repozitóriumból, vagy személyes kapcsolat révén
- Van olyan tudományterület, ahol a dinamikus fejlődés okafogyottá teszi a korábbi adatok újrafelhasználását
- Az adatvesztés általánosságban nem jellemző, a technikai fejlődés és ezzel karöltve technikai jártasság előrehaladtával jelentősen csökkenő tendencia



# Felmerült igények a repozitóriummal kapcsolatban

*"jelen állapotban az egyéni kutatói gárdák alakítják ki a fájlstruktúrát, a mappastruktúrát. Ők nevezik el a fájlokat és így kerülnek egy-egy szerverrészre jó esetben, vagy egy-egy külső tárolóra ezek az adatok. És jelenleg ott tartunk, hogy ki kell alakítanunk egy egységes rendszert, vagy valamiféle logikai rendszert arra, hogy ezeknek az állományoknak mi legyen a nyilvántartása, mi legyen a rendszere."*

- A legtöbb ELKH-s intézményben szükséges valamilyen előfeltétel teljesülése, hogy a kutatók optimális módon építhessék be a munkafolyamatukba a repozitórium használatát



- adatstruktúrák és metaadatok egységesítése
- adatgazda munkatárs, aki segítséget nyújt az adatkezelésben és repozitálásban
- oktatás, kultúrateremtő események és szabályozás fontossága

*„nem lesz kapacitásom, meg tekintélyem se, hogy az egyes emberekből kiverjem, hogy te még nem adtad le a fotóidat. Szóval ezért találtuk ki azt, hogy az éves jelentéshez, ahol ugye rendbe kell tenni az MTMT-t, hogy mit írt, azzal párhuzamosan akkor ezt is rendbe kell tenni, hogy amit vállalt, hogy kutat, annak mi a lenyomata, és azon belül aztán már szabadság van. [...] ehhez például a metaadatokat is nekik kell kitölteni, tehát az sem működik, hogy jó, nesze, itt vannak a képek.”*

# Felmerült igények a repozitóriummal kapcsolatban



- A legáltalánosabb a *hosszútávú és biztonságos tárolás* igénye
  - Van olyan kutatóhely, ahol a nyers adatok tárolása megoldott, de a feldolgozottaké kevésbé
  - Van, ahol hosszútávon egyáltalán nem látják biztosítottnak az adatok biztonságos tárolását
- Többek a *köztes tárolás és megosztás* fontosságára helyezték a hangsúlyt
- A *publikációs kötelezettségeknek való megfelelés és repozitálás* is kiemelten fontos igényként jelentkezett (DOI)
- Repozitóriumokban jártasabb kutatók említették a *metaadat-sémák* és az *összetett keresőfelület* hasznosságát is



**Köszönöm a figyelmet!**



<https://science-research-data.hu/>

