



# Kutatói igényfelmérés a repozitóriumfejlesztés támogatására

ELK ARP PROJEKT - WP3

# Célkitűzés

A felmérés célkitűzése az volt, hogy a leendő felhasználók kutatási adatkezeléssel és repozitóriumhasználattal kapcsolatos

- ismereteit,
  - tapasztalatait,
  - gyakorlatait
- alaposan megismerjük,

**annak érdekében**

hogy a készülő adatrepozitóriumot minél inkább ezen igényekre szabhassuk.

# Célcsoport

Az igényfelmérés célcsoportját az Eötvös Loránd Kutatási Hálózat kutatói jelentették, akik

- 11 kutatási központban,
- 7 kutatóintézetben és
- 116 támogatott kutatócsoportban

- az agrártudományok
- a bölcsészettudományok
- a matematikai tudományok
- a műszaki tudományok
- az orvostudományok
- a társadalomtudományok
- a természettudományok területén tevékenykednek.



# Alkalmazott módszerek

Az adatgyűjtést a Számítástechnikai és Automatizálási Kutatóintézet (SZTAKI) és a Társadalomtudományi Kutatóközpont (TK) munkatársai közösen koordinálták és végezték

- online kérdőíves megkérdezés és
- (többnyire) online interjú adatfelvételek megvalósításával.



A cél az volt, hogy mindegyik ELKH intézményt érjük el legalább az egyik módszerrel, ez teljesült is.

# Kérdéskörök



# Információk feldolgozása

Online kérdőíves  
felmérés adatai



A két adatfelvétel  
releváns  
információi  
együttesen

szoftverspecifikáció készítése



Interjúk leiratai

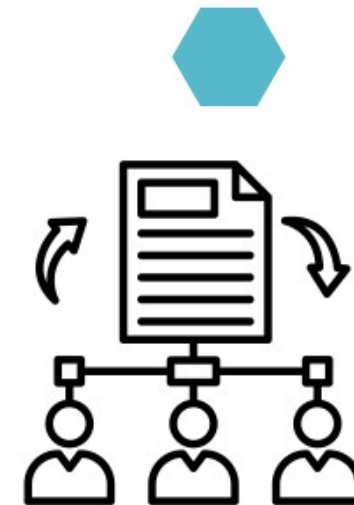
tartalomlemezés



# Eredmények bemutatása, megosztása

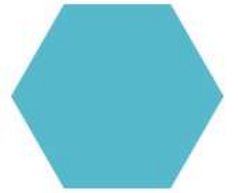
A kapcsolódó előadások, prezentációk folyamatosan kerülnek fel a projekt honlapjára, ahol nyíltan elérhetők

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





# Az igényfelmérésből a „motiváció” előadásban felmerülő kérdésekre adott válaszok





# Mi a kutatási adat?



A kutatásban az [opscience.hu](https://opscience.hu) definícióját használtuk:

„A kutatási adatok a tudományos közösség által létrehozott, rögzített, elfogadott és megőrzött tényadatok, amelyek a kutatási eredmények hitelességét támasztják alá. Létrejöhetnek megfigyelések, kísérletek, szimulációk eredményeképpen – egy konkrét kutatás számára előállítva – vagy korábban gyűjtött adatok összegyűjtésével, válogatásával, feldolgozásával.”

Tehát nem kizárólag digitális adatokat értettünk kutatási adat alatt – és a tapasztalatok szerint a kutatási hálózatban máig keletkeznek nem digitális kutatási adatok (pl. terepen felvett jegyzőkönyvek), illetve régebről is maradt meg még sok nem digitális adat.

Kérdés, hogy szükséges-e, illetve lehet-e, érdemes-e ezeket digitalizálni.



# A kutatási hálózatban keletkező adatok

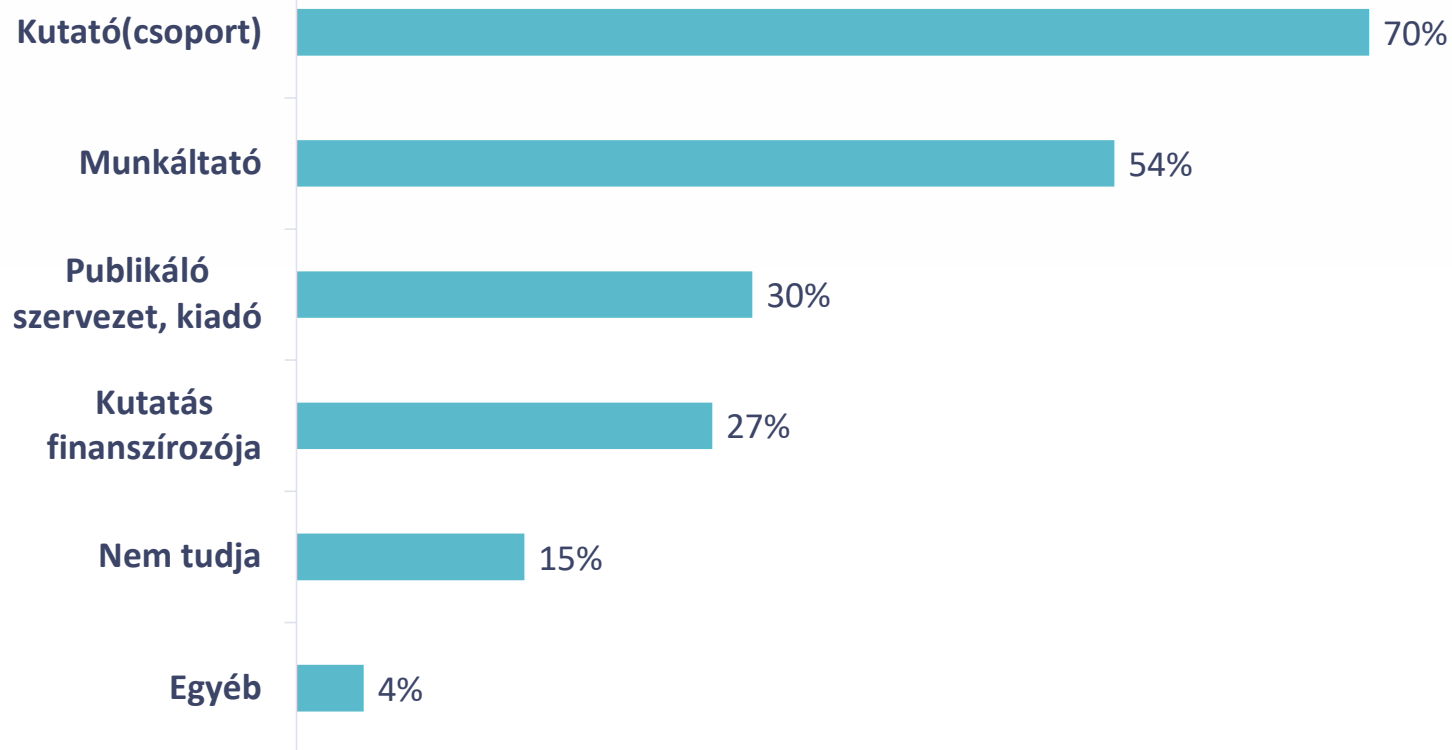
- A kutatások során jellemzően keletkeznek új adatok
- Az adatok **mérete** (a 10 KB-ostól a több terabájtosig) és **formátuma** (jegyzettől a programkódig) meglehetősen **heterogén**
- Jellemző a **többféle fájlformátum együttes** használata, esetenként tömörített fájlalba (pl. zip, tar.gz) csomagolva.
- A **leggyakrabban előforduló formátumok** a szöveges dokumentumok (pl. doc, txt), az adatsorok (pl. CSV, TSV, Excel) és a képek (pl. PNG, JPG).
- Ezeket követik a forráskódok/programkódok, videók (pl. MPEG, WMV) és az adatbázisok (pl. SQL, SPSS).
- Sok egyéb, jóval ritkábban említett formátum is felmerült (cha, xml, gbk, fasta, shp, kml, kmz, dxf, dwg, psd, ai, axml, spec, mca, spc, fastq, sam, bam, bai, fai, bed, plink, geno, FITS, gexf, TEI XML, HDF5, HEPMC, lmf, RAW, WIFF, prb, prm, Rdata, tex, json, zip, TextGrid, trs, TIF, ND2, AVI, MOV, mgf, dat, baf)



© Can Stock Photo

# Kié a kutatási adat?

A kutatásokban keletkező adatok tulajdonosa (a kérdőíves felmérésből)



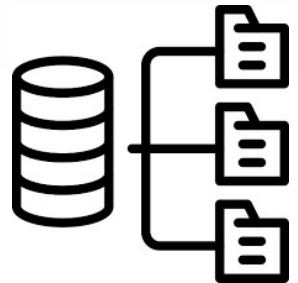
# Kinek a feladata az adatkezelés? – kötelezettségek és lehetőségek

- Az ELKH kutatóinak többsége viszonylag szabad kezet kap intézményétől adatainak kezelésére, kutatás alatti és utáni tárolására.
- Jellemző az általános szabályozás hiánya, kisebbségben vannak azon intézmények, amelyekben intézményi előírással vagy ajánlással látnák el a kutatókat a kérdésben.
- Néhány intézményben ugyanakkor vannak külső jogi szabályozásból fakadó nagyon szigorú adatkezelési előírások.



# Adatarchiválás

- A feldolgozott adatok hosszútávú tárolására a legtöbb kutatónak saját kialakult gyakorlata van, a leginkább jellemző megoldás a külső tárolóra és/vagy saját gépre történő kézi mentés.
- Bár egyre több intézmény biztosít intézményi tárhelyet is ehhez, az inkább lehetőség, kevésbé jellemző az előírt archiválási kötelezettség.
- Egyelőre viszonylag ritka a repozitóriumi elhelyezés és jellemzően nem automatikusan történik, hanem alkalmasszerűen, konkrét helyzetekben (pl. finanszírozói vagy publikációs előírások).



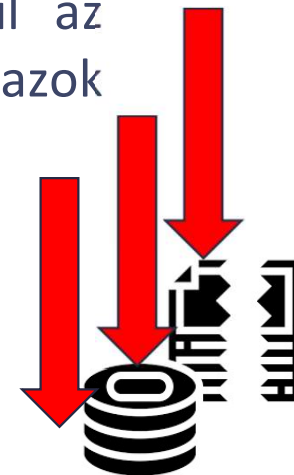
# Mekkora az adatvesztés kockázata?

- A kutatási hálózatban a technikai okból – belső/külső merevlemez, pendrive meghibásodása – vagy emberi hiba miatt bekövetkező adatvesztés előfordulásának esélye mára jelentősen lecsökkent, részben a technikai fejlődésnek, részben az esetleges korábbi eset(ek) miatt megváltoz(tat)ott felhasználói szokásoknak köszönhetően.
- Ám a kutatások lezárása után adatok „vesznek el” úgy is, hogy fizikailag ugyan megvannak, de a felhasználásukat vagy az értelmezésüket lehetővé tevő szoftver/hardver, esetleg információ – akár arról, hogy az adott intézmény más kutatásában vagy korábbi kutatásából létezik ilyen adat –, vagy a kereshetőséget lehetővé tevő metaadatok, tárolási struktúrák híján ma már nem, vagy csak plusz munka árán lehet hozzájuk férni, őket felhasználni.



# Repozitóriumi elhelyezés hatása az adatvesztés lehetőségére

- A technikai okból vagy emberi hiba miatt bekövetkező adatvesztés előfordulásának lehetőségét a repositóriumi elhelyezés gyakorlatilag kizárja.
- Az avulós adatvesztés lehetőségét két úton is jelentősen csökkenteni tudja. Egyrészt a repositóriumok által ajánlott fájlformátumok nem szoftverspecifikusak, másrészt egy jól gondozott repositórium felkészül az esetleges technikai változásokra, és megteszi a szükséges lépéseket azok hatásának kivédésére.



# Miért érdemes kutatási adatokat repozitóriumba helyezni?

Kutatói érvek a repozitóriumi elhelyezés mellett (egyben elvárások is):

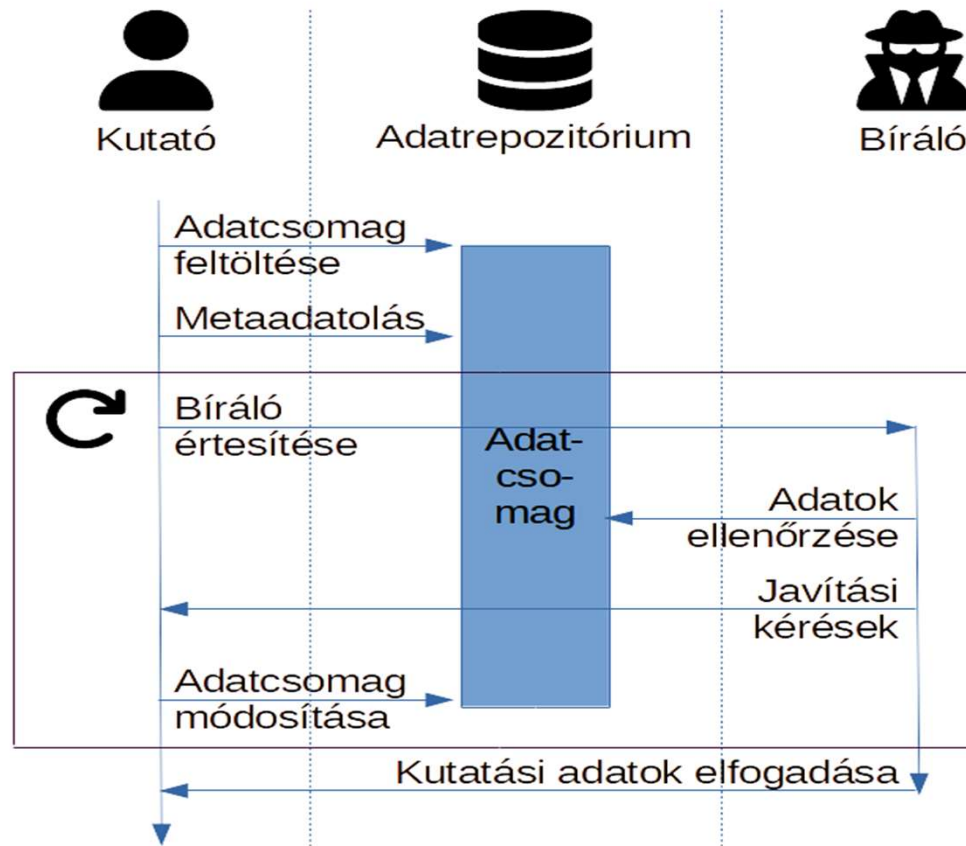
- A hosszútávú és biztonságos tárolás lehetősége
- A köztes (kutatás során történő) tárolás és megosztás lehetősége
- A publikációs kötelezettségeknek való megfelelés és repozitálás (DOI)
- Metaadat-sémákhoz történő hozzáférés, keresőfelületeken való megjelenés lehetősége, hasznossága





# Mire használhatják a kutatók a repositóriumot?

Használati esetleírás példa – adatcsomag feltöltése (sequence diagram)



# Hogyan lehet a repozitóriumi elhelyezést ösztönözni?

- LEGYEN KÖNNYEN KEZELHETŐ:
  - ✓ egyszerűség, egyértelműség, kereshetőség, minél kevesebb extra adminisztráció
- LEGYEN MELLÉ INTÉZMÉNYI TÁMOGATÁS:
  - ✓ az intézmény vezetése támogassa, segítse, jutalmazza a kutatási adatok archiválását és nyílt közzétételét
  - ✓ biztosítson hozzá dedikált személyzetet (intézménykönyvtáros / könyvtár-informatikus / kutatástámogató)
  - ✓ legyenek illeszkedő pályázatok
- LEGYEN HOZZÁ KÖVETHETŐ SZABÁLYOZÁS/AJÁNLÁS:
  - ✓ intézményi iránymutatás



**Köszönöm a figyelmet!**