



ELKH ARP projekt - Adatok górcső alatt 2.

Kihívások és megoldási lehetőségek egy interdiszciplináris adatrepositórium létrehozásában

Egyed-Gergely Júlia, Meiszterics Enikő
Társadalomtudományi Kutatóközpont
Kutatási Dokumentációs Központ

Interdiszciplináris adatrepositóriumot fejlesztünk

Az ELKH ARP projekt intézményi interdiszciplináris tudományos adatrepositóriumot fejleszt a magyar kutatói közösség számára

- feltöltés: kutatási hálózat valamennyi kutatója + egyéb kutatók
- elérés: hazai és nemzetközi tudományos közösség

Részletek: science-research-data.hu

➤ **Kihívások**

- tudományos környezet
- intézményi környezet
- emberi oldal
- szabályozási oldal

➤ **Megoldási ötletek, tervek**

Kihívások - tudományos környezet

"A legtöbb helyen most már ez nagyobb lapokban elvárás, hogy az adatokat is csatolni kell a publikációkhoz."



- Adatrepozitórium-használat
 - ismer adatrepozitóriumot, amit igénybe vehetne: 44% (ARP igényfelérés - kérdőíves kutatás, 2022.)
 - használt már adatrepozitóriumot: 17% rendszeresen, 23% alkalmanként (ARP igényfelérés - kérdőíves kutatás, 2022.)
 - elvárásbeli különbségek / célok
 - publikációs elvárásnak való megfelelés
 - mindig
 - folyóiratfüggő
 - egyszer találkozott már vele
 - kutatói körrel való megosztás
 - folyamatosan bővülő adatbázis építése
 - citizen science adatok gyűjtése
 - helyi tárhely felszabadítása

Kihívások - intézményi környezet I.

[Adatmenedzsment terv] "Az micsoda? De jól hangzik!"

- **Adatmenedzsment, szabályozás, policy**

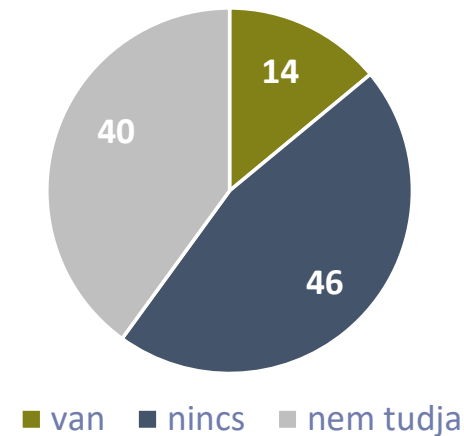
- alig van intézmény, ahol lenne szabályozás vagy ajánlás az adatok tárolásával, megosztásával kapcsolatban
 - szigorú szabályozás
 - erős ajánlás
 - esetenként tiltás
 - tilos online módon megosztani, tárolni
 - kutatószoba
 - pen drive-os, külső winchesteres megosztás

- **Intézményi segítség**

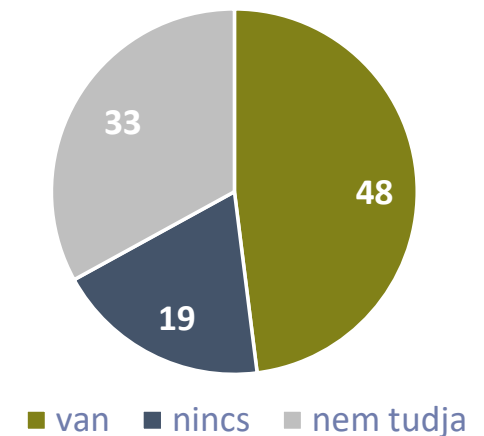
- kevés helyen van kutatástámogatásra személyzet
- van, ahol kialakult, hogy kit lehet kérdezni
- igény lenne rá

ARP igényfelmérés -
kérdőíves felmérés, N=129
(17 intézmény), 2022.

**Személy, szerepkör adatok
archiválásra az intézményben
(%)**

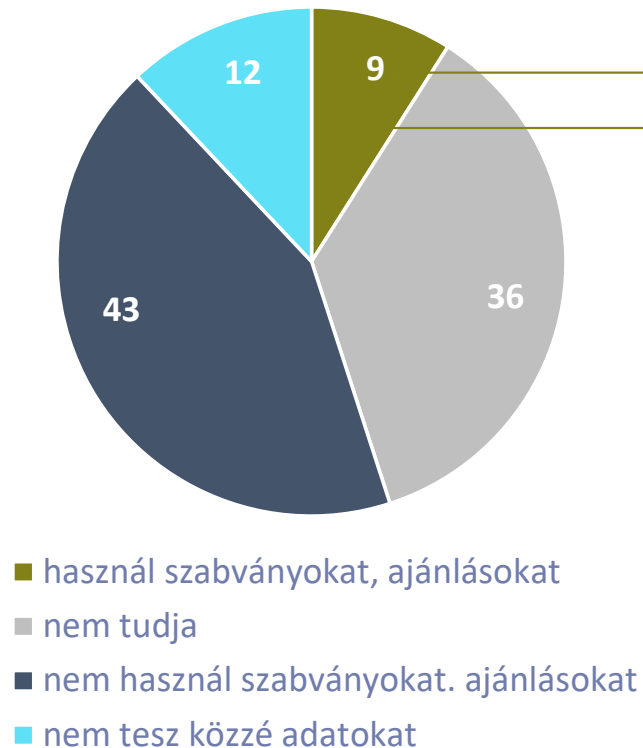


**Adatkezelésben,
adatarchiválásban támogatást
nyújtó munkatárs (%)**



Kihívások - intézményi környezet II.

Adatok, dokumentumok közzétételekor
használt szabványok, ajánlások (%)



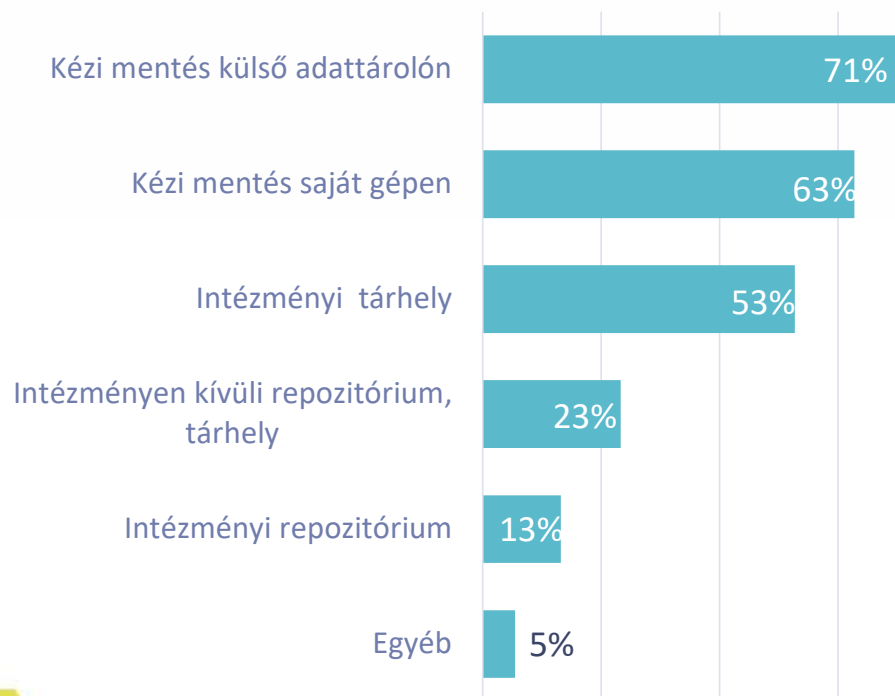
Használ szabványokat, ajánlásokat (9%):

- Adott újság ajánlását követve
- Az arXiv előírásait kell követnem
- DataCite
- FAIR Guidelines
- FITS, CDS
- FITS, SQL, JSON, CSV/TSV
- Folyóiratfüggő
- GDPR, adatgazdával történő megállapodás szerint
- GRPD, Concorda
- Helyzetfüggő, ENA és GitHub szabályozások alapján
- Intézményi vonatkozó szabályzat
- TEI, ISO, UTF8, DC

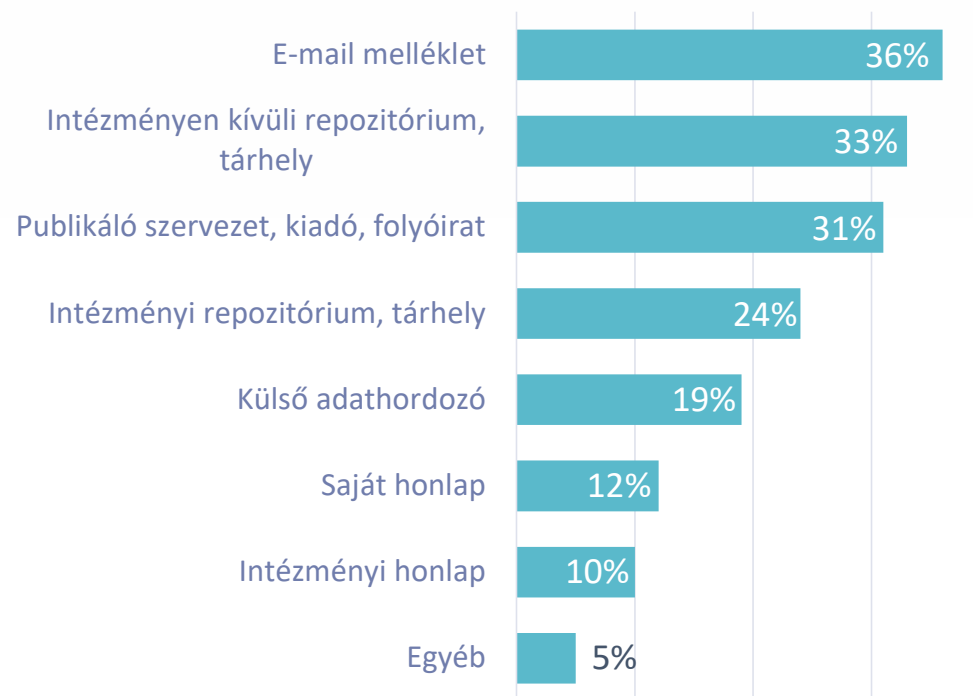
ARP igényfelmérés - kérdőíves felmérés,
N=129 (17 intézmény), 2022.

Kihívások - intézményi környezet III.

Mentés a kutatás lezárulta után



Hozzáférés a kutatás lezárulta után



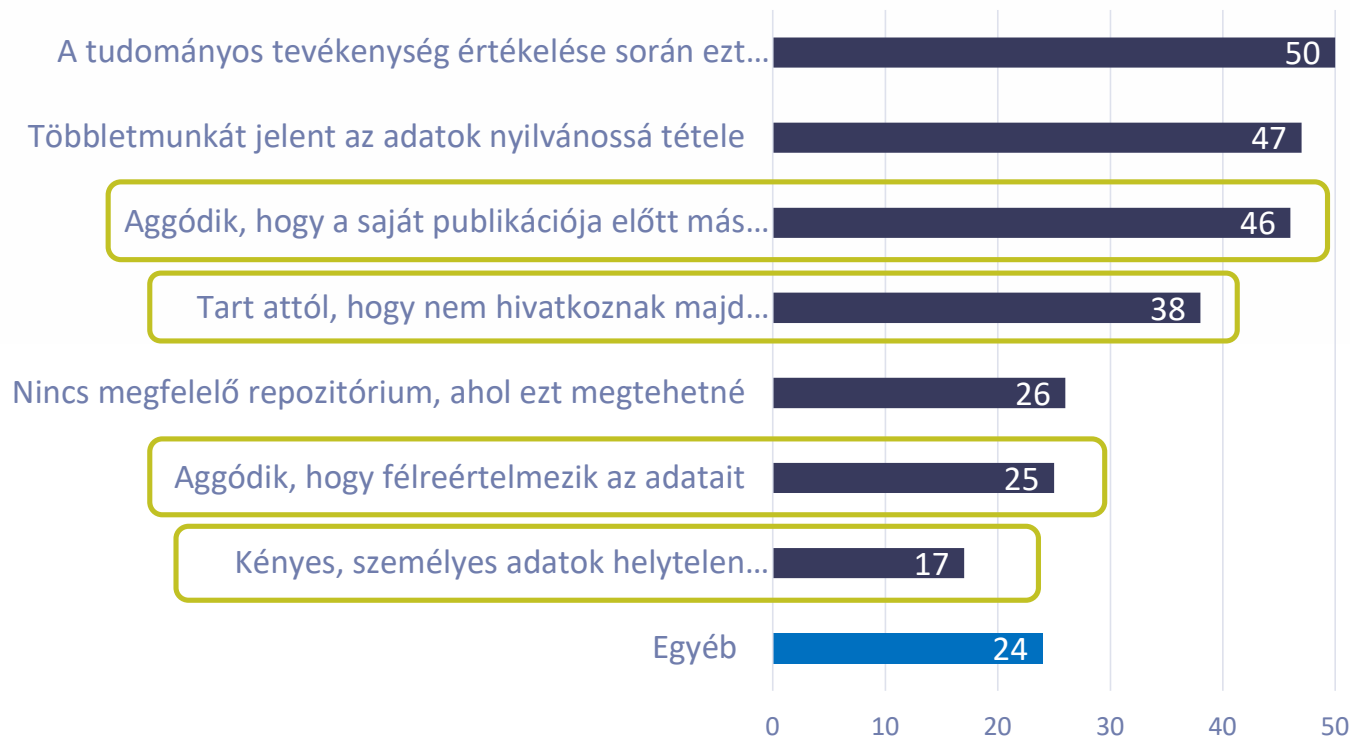
Kihívások - emberi oldal I.

„Egy kísérletező mindig egy kicsit fél attól, hogy egyrészt ellopják az adatait, másrészt, hogy mások esetleg máshogy interpretálják, és akkor neki abban, hogyha már egyszer kiadta a kezéből az adatot, nincsen beleszólása esetleg, hogy mások hogyan interpretálják esetleg félre az ő adatait.”

- Jellemző problémák

- adatféltség
- bizalomhiány - csak az a biztos, ami saját kézben van
- megszokások - egyes humánkutatások: még nem bevett szokás előre beleegyezést kérni az adatok későbbi felhasználásával kapcsolatban
- újítkészség (hiánya)
 - új fogalmak, új terület megismerése (adatrepozitórium, nyílt tudomány, metaadat...)
- szkepticizmus

Kihívások - emberi oldal II.



Egyéb:
Nincsenek ilyen akadályok

Nincsenek keletkező adatok
Nincs jogosultsága megosztani
Az adatok önmagukban nem értelmezhetőek
Nincs szükség rá, elég a publikáció
Tartalmazhat hibákat / nem publikálásra érett

Túl nagyok a fájlok (10GB-néhány 100GB)
Fizetősek az ilyen szolgáltatások
Nincs kultúrája

Melyek azok a legfőbb akadályok, amelyek meggátolják abban, hogy nyers adatait nyilvánosan hozzáférhetővé tegye (több válasz is megadható)? – Kutatási adatok kezelése, kérdőíves felmérés, N=173, HRDA-ELKH Titk., 2020.

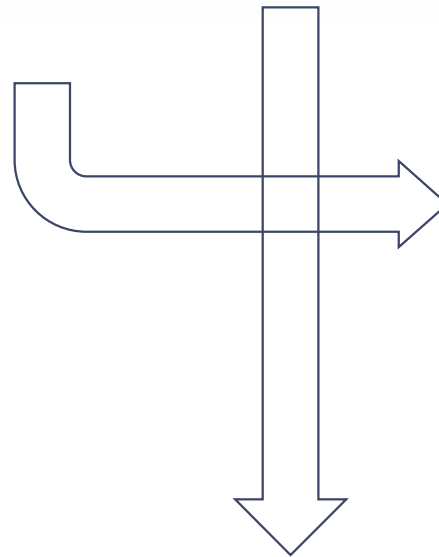
Kihívások – technikai, szabályozási oldal I.

- Adattípusok

- hogyan lehet sok és egymástól teljesen különböző típusú adatot egy helyen tárolni
 - keresés, metaadatolás, megjelenítés
- mit lehet a teljesen egyéni adattípusokkal kezdeni

- Jogi, etikai kérdések

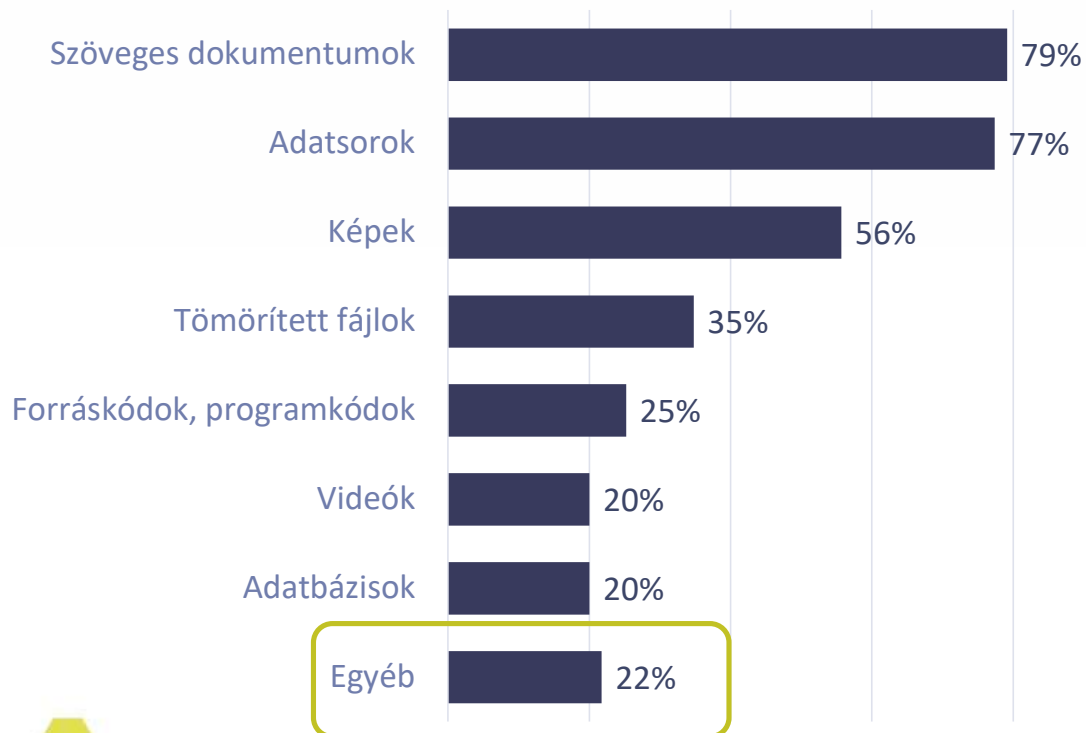
- szenzitív adatok
- tulajdonjogok - kié az adat
- jogi, szerzői jogi stb. akadályok
- etikai kérdések



„Utána ez az egész lesz elmentve egy olyan formátumba, amit senki nem tud használni, csak az, aki rendelkezik ezzel a programmal, ez egy nagyon szép, házi fejlesztésű program, egyik kollégám fejlesztette”

Kihívások - technikai, szabályozási oldal II.

Az adatcsomagban található fájlok típusa



- Szöveges dokumentumok: pl. doc, txt
- Adatsorok: pl. CSV, TSV, Excel
- Képek: pl. PNG, JPG
- Tömörített fájlok: pl. zip, tar.gz
- Forráskódok/programkódok
- Videók: pl. MPEG, WMV
- Adatbázisok: pl. SQL, SPSS
- Egyéb említések (22%!!): 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

ARP igényfelmérés - kérdőíves felmérés,
N=129 (17 intézmény), 2022.

Megoldási tervek – ismeretterjesztés

- Ismeretterjesztő előadássorozat, kerekasztal-beszélgetések
 - *Kincs, ami van* sorozat (április 20: citizen science)
 - *Szakmai nap* sorozat
- Önálló projekthonlap - www.science-research-data.hu
 - rendszeresen frissülő hírek
 - ismeretterjesztő- és oktatóanyagok
 - tudástár
 - információ a fejlesztés aktuális állapotáról
- ARP nagykövet pályázat - 2023. tavasz-ősz
 - 6 ARP nagykövet
 - 6 intézmény
 - ARP kipróbálása, „megtanulása”, adatfeltöltés, tesztelés
 - képzésen való részvétel
 - intézményi workflow kialakítása
 - ismeretátadás intézményen belül

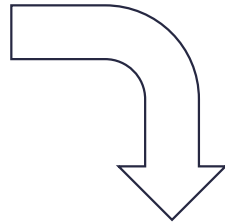


Megoldási tervek – ajánlások, sablonok

- Közös ajánlások, sablonok
 - közös policy
 - közös adatmenedzsment sablon
 - etikai ajánlás
 - Magyar Szociológiai Társasággal (MSZT) közös tervek
- Tudományterületi vagy intézményi ajánlások, sablonok
 - a közös ajánlások alapján, azok specifikálásával kialakíthatóak a tudományterületi vagy intézményi ajánlások
- Metaadatsémák
 - létezők összeszedése, felhasználása
 - nemzetközi metaadatsémák
 - tudományterületi metaadatsémák
 - többszintű metaadatolási lehetőség
 - alapadatok + választható adatok

Megoldási tervek – személyes kapcsolat I.

- Kutatók bevonása már az előkészületekbe
 - igényfelmérés: kérdőíves felmérés, interjúk
 - kapcsolattartás az interjúalanyokkal, érdeklődőkkel
 - intézeti kapcsolattartó-hálózat
- ARP nagykövet pályázat
 - mentorálás
 - adatgazda (data steward) minden intézménybe
 - adatgazdász hálózat kiépítése az ELKH intézményeiben, akik munkáját az ARP szakemberei fogják össze
 - tesztelés - visszacsatolás
- ARP szakértők rendelkezésre állása
 - technikai (informatikusok)
 - gyűjteménygondozói

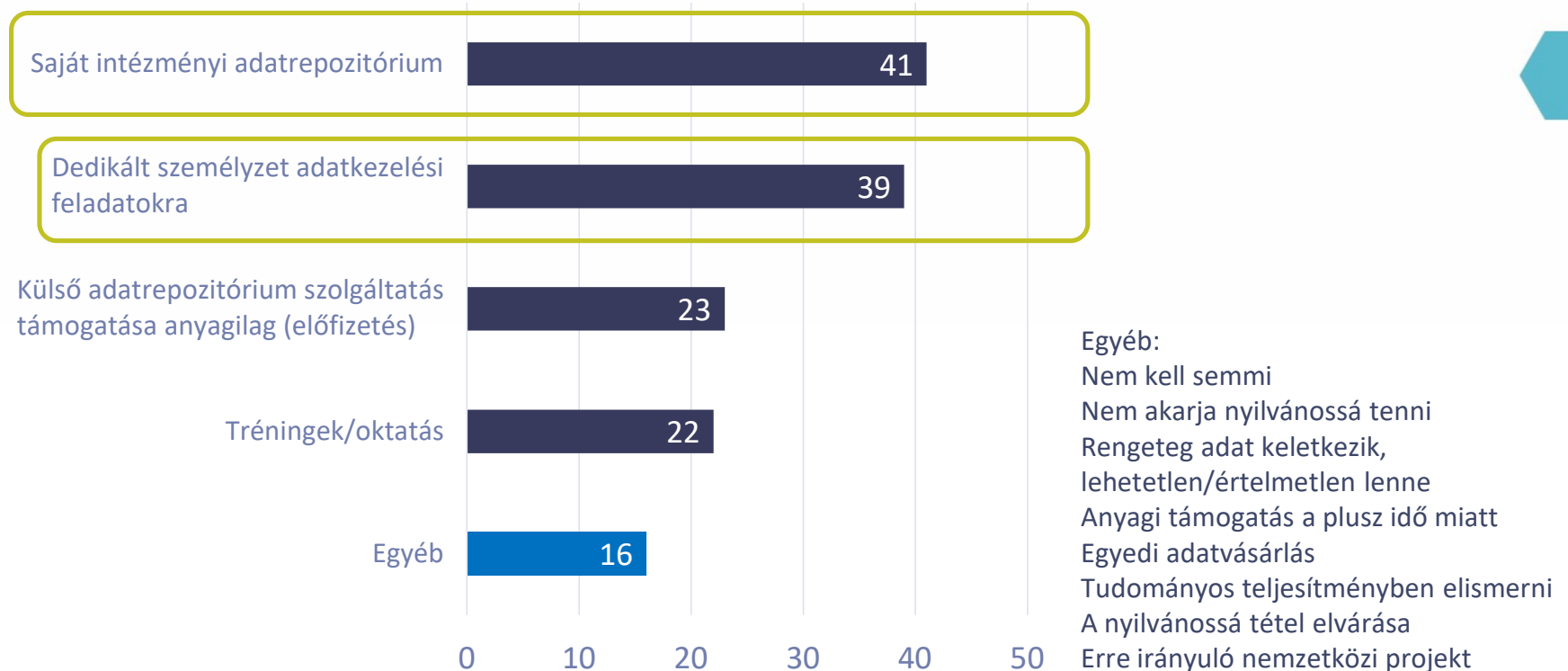


ARP nagyköveteket keresünk!

📅 2023. február 23.

Az ELKH ARP projekt pályázatot hirdet a kutatási hálózat dolgozói számára adatrepozitóriumi nagykövet pozíciók betöltésére.

Megoldási tervek – személyes kapcsolat II.



Milyen jellegű támogatás növelné annak esélyét, hogy nyers adatait nyilvánosan hozzáférhetővé tegye? (több válasz is megadható) – Kutatási adatok kezelése, kérdőíves felmérés, N=173, HRDA-ELKH Titk., 2020.

„The State of Open Data” kutatás

A Springer Nature kiadó és a Figshare repozitórium közös „The State of Open Data,, kutatása egy online kérdőíves adatfelvétel keretében 2016 óta gyűjt évente adatokat a világ minden tájáról a kutatók kutatási adatokkal, adatkezeléssel és -megosztással kapcsolatos véleményeiről, gyakorlatairól.

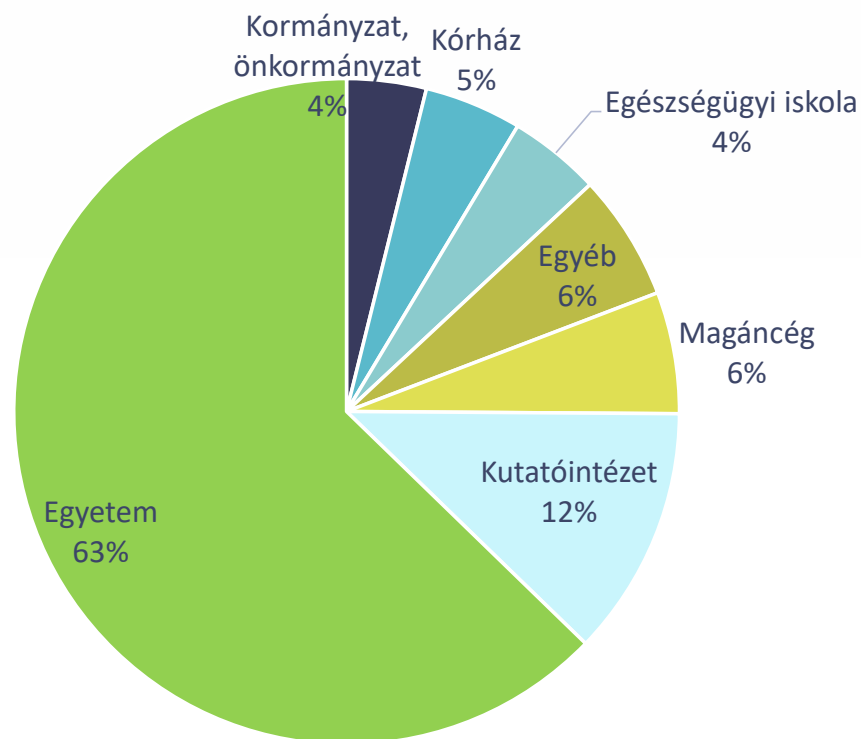
Az adatmegosztást hátráltató tényezőket vizsgáló kérdés 2018-ban került be a kérdőívbe.

Az éves jelentés elkészülte után a kérdőívet és az adatbázist is mindenki számára hozzáférhetővé teszik a Figshare repozitóriumban.



A 2022-es adatfelvétel

2022-ben minden korábbi évet meghaladta a válaszadók száma, 147 országból 6798 (21 magyar) kutató töltötte ki a kérdőívet, közülük 827 fő dolgozott kutatóintézetben.



Research, Nature; Goodey, Gregory (2022): State of Open Data Survey 2022 additional resources. figshare. Dataset. <https://doi.org/10.6084/m9.figshare.21295422.v2>

Adatmegosztási hajlandóságot hátráltató tényezők (2022)

	Összes válaszadó N = 6104	Kutatóintézetben dolgozó kutatók N = 776
Nincs problémája	22%	20%
Tart a téves értelmezéstől	35%	34%
Megfelelő elismerés hiánya	28%	31%
Érzékeny/hozzájárulást igénylő adatok	28%	26%
Szerzői jogokkal, licencekkel kapcsolatos bizonytalanság	27%	27%
Egy másik laboratórium megelőzheti a felfedezésben	20%	24%
Nem biztos, hogy finanszírozója/intézménye hozzájárul	20%	23%
Költségvonzat	18%	18%
Nem biztos, hogy teljesen kiaknázza a bennük rejlő lehetőséget	18%	19%
Rendszerezni kellene azokat, hogy másoknak is érthetőek és használhatóak legyenek	16%	16%
Időhiány	14%	14%
Egy másik laboratórium másképp értelmezi az adataimat	12%	14%
Nem tudja, melyik repo-t használja	11%	11%
Túl nagy adatok	9%	9%
Mások nem ugyarra a következtésre jutnak	6%	6%
Túl kicsi/érdektelen adatok	6%	5%
Mások hibát találnak az adatokban	5%	7%
Nem kívánja megosztani az adatait	3%	2%
Egyéb	3%	2%



Változások (2018-2022)

	2018 N = 1356	2022 N = 6104
Nincs problémája	2%	22%
Rendszerezni kellene azokat, hogy másoknak is érthetőek és használhatóak legyenek	34%	16%
Nem biztos, hogy finanszírozója/intézménye hozzájárul	34%	20%
Nem tudja, melyik repo-t használja	25%	11%
Túl kicsi/érdektelen adatok	18%	6%
Szerzői jogokkal, licencekkel kapcsolatos bizonytalanság	37%	27%
Időhiány	22%	14%
Megfelelő elismerés hiánya	35%	28%
Túl nagy adatok	14%	9%
Egyéb	7%	3%
Tart a téves értelmezéstől	38%	35%
Érzékeny/hozzájárulást igénylő adatok	31%	28%
Költségvonzat	20%	18%
Mások hibát találnak az adatokban*	7%	5%
Mások nem ugyanarra a következtésre jutnak*	7%	6%
Egy másik laboratórium megelőzheti a felfedezésben***	20%	20%
Nem biztos, hogy teljesen kiaknázza a bennük rejlő lehetőséget**	18%	18%
Egy másik laboratórium másképp értelmezi azokat*	12%	12%
Nem kívánja megosztani az adatait*	3%	3%

*2019-től, **2020-tól, ***2021-től kérdés



Köszönjük a figyelmet



egyed-gergely.julia@tk.hu
meiszterics.eniko@tk.hu