

# Az ELKH Adatrepozitórium Platform (ARP) projekt bemutatása

Kacsuk Péter

SZTAKI

ELKH ARP projekt vezetője



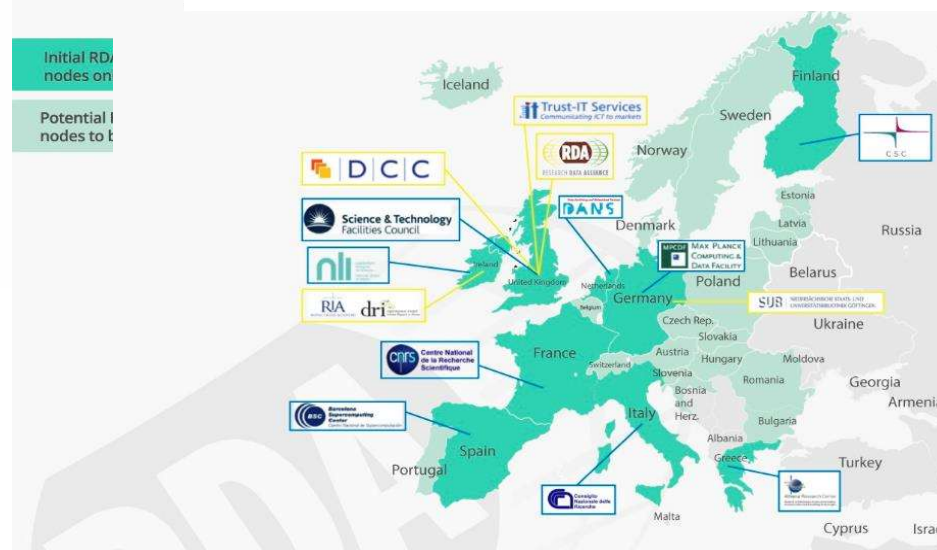
# Motivációk és előzmények

- EU-ban egyre nagyobb figyelem és támogatás a FAIR adatkezelés felé
  - EOSC (European Open Science Cloud)
  - RDA (Research Data Alliance) -> HRDA
- Magyarországon is növekvő figyelem a FAIR adatkezelés iránt
  - CONCORDA Adatrepozitórium a SZTAKI-ban (2020)
  - ELKH-HRDA Adatrepozitórium projektek támogatási programja (2021)
- Az ELKH Cloud projekt, mint **számítási** felhő sikere
- Ugyanakkor az ELKH Cloud, nem adattárolási felhőként működik
- ELKH intézmények adatkezelési workflowja sok helyen nagyon elavult

# RDA (Research Data Alliance) szervezet

- Egy **globális nemzetközi fórum**, amely segíti a kutatási adatok kölcsönös és akadályoktól mentes megosztását és felhasználását.
- Egy **találkozóhely a 13000+ nemzetközi kutatási adat szakértő számára** – bottom-up alulról építkező megközelítés
- Egy **probléma megoldó struktúra** – a tagok segítséget kapnak az RDA közösségtől a megoldások megtalálásához
- **61 Interest groups** és **45 Working groups**, cél: adatkezelési és adatelérési standardok kidolgozása
- **RDA recommendations & outputs & RDA ICT Technical specs** – elismerten alkalmazható megoldások

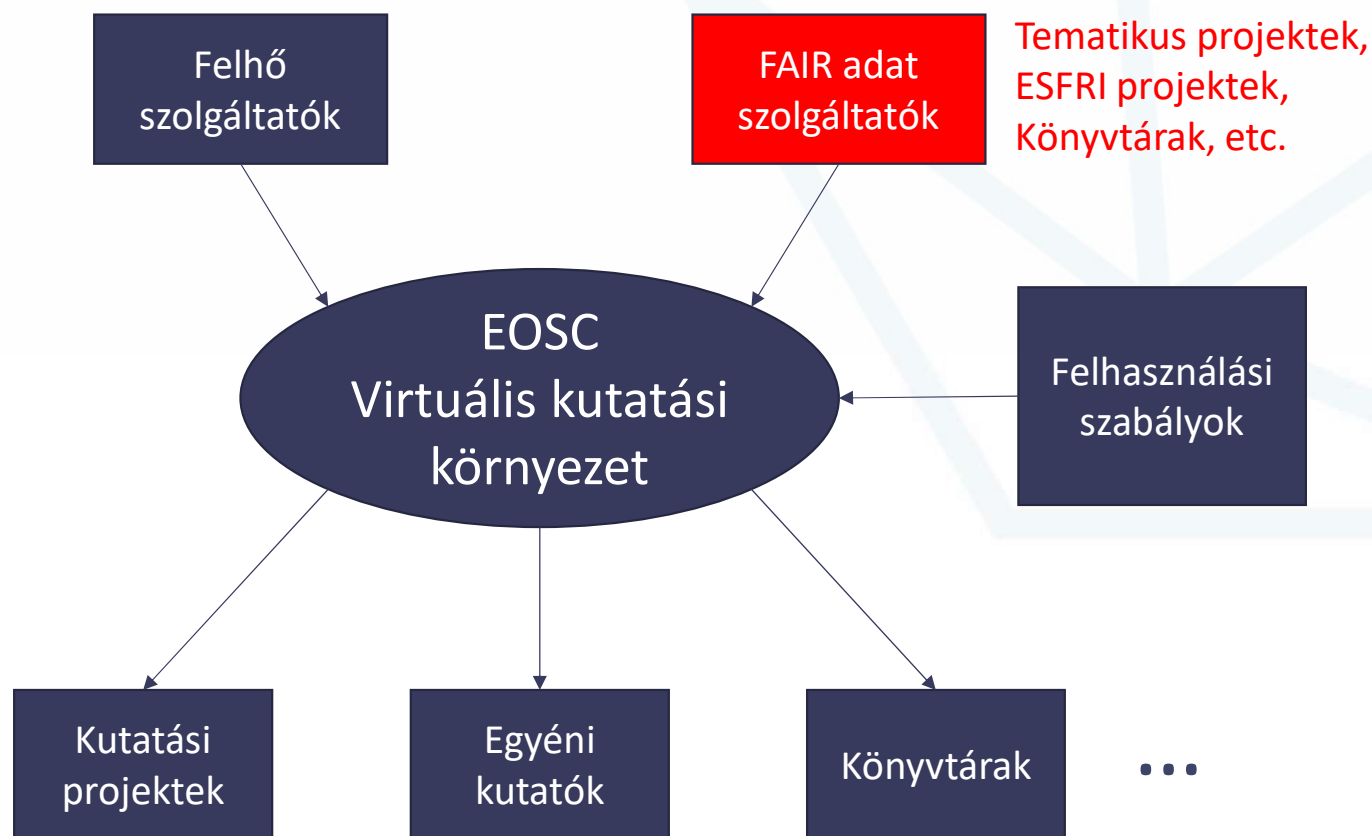
9 Nemzeti Node az RDA Europe 4.0 indulásától  
13 további alapítása a projekt végrehajtása során



<https://www.rd-alliance.org/>

# EOSC (European Open Science Cloud)

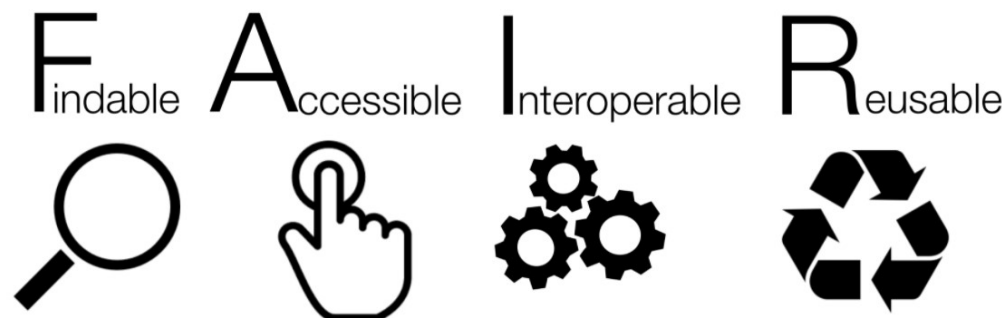
- Az RDA-ban kidolgozott standardok és ajánlások alapján
  - egy olyan európai infrastruktúra létrehozása, amiben
  - a különböző európai (kutatási és ipari) felhőkből
  - a különböző európai adatrepozitóriumok adatai elérhetők és feldolgozhatók
- **Virtuális elérés és használat**



# HRDA (Hungarian Node of RDA)

- Az RDA magyarországi szervezete, 2019-ben alakul
- Célok:
  - Részvétel az RDA munkájában
  - Az RDA eredmények hazai bevezetése
- HRDA adat repo pályázat (2021)
  - ELKH 50 mFt támogatásának felhasználásával
  - 7 elfogadott pályázat, amik többsége a CONCORDA-t használta
  - Projektenként 1.5 – 5 mFt támogatás
  - Értékelés: rendkívül sikeres program volt

# A projekt célkitűzései

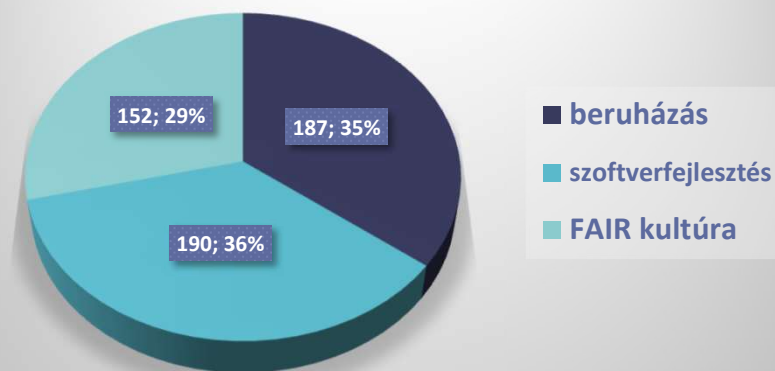


- FAIR adatrepozitórium **infrastruktúra** megteremtése
  - Olyan új adatrepozitórium infrastruktúra-szolgáltatás alapjainak megteremtése, ami a teljes ELKH intézményhálózat folyamatos, hosszútávú kutatási adatkezelését képes lesz támogatni
    - ELKH Cloud projekt -> számítási infrastruktúra megteremtése
    - ELKH ARP projekt -> FAIR adatkezelési infrastruktúra megteremtése
- A FAIR adatkezelési **kultúra** megteremtése az ELKH intézményhálózatában
  - A meglévő **kutatási workflow**-k kibővítése a FAIR adatrepozitóriumok használatával
- Bekapcsolódás hasonló célú **európai kezdeményezésekbe**

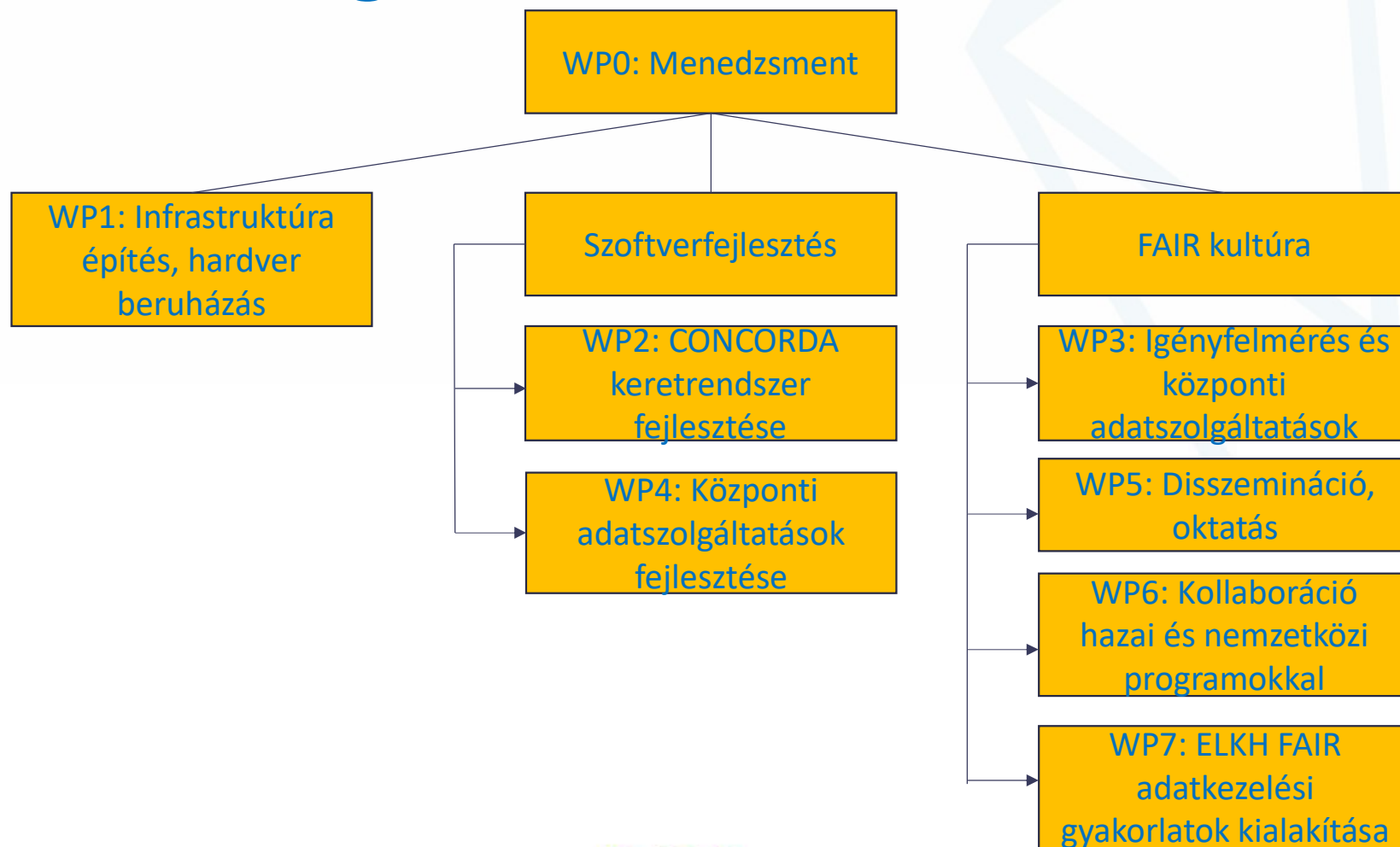
# A projekt megvalósítása

- A projekt megvalósításában résztvevő intézmények:
  - SZTAKI, TK, WIGNER FK
- A projekt megvalósításának tervezett időtartama:
  - 18 hónap: 2021. nov. 1. – 2023. ápr. 30.
- A projekt költségvetése: 529 millió Ft
  - Beruházás: 187 mFt
  - Szoftver fejlesztési költség: 190 mFt
  - FAIR adatrepó kultúra kialakítása, terjesztése: 152 mFt
- **A projekt támogatója: ELKH**

ARP költségek felosztása



# Munkacsomagok





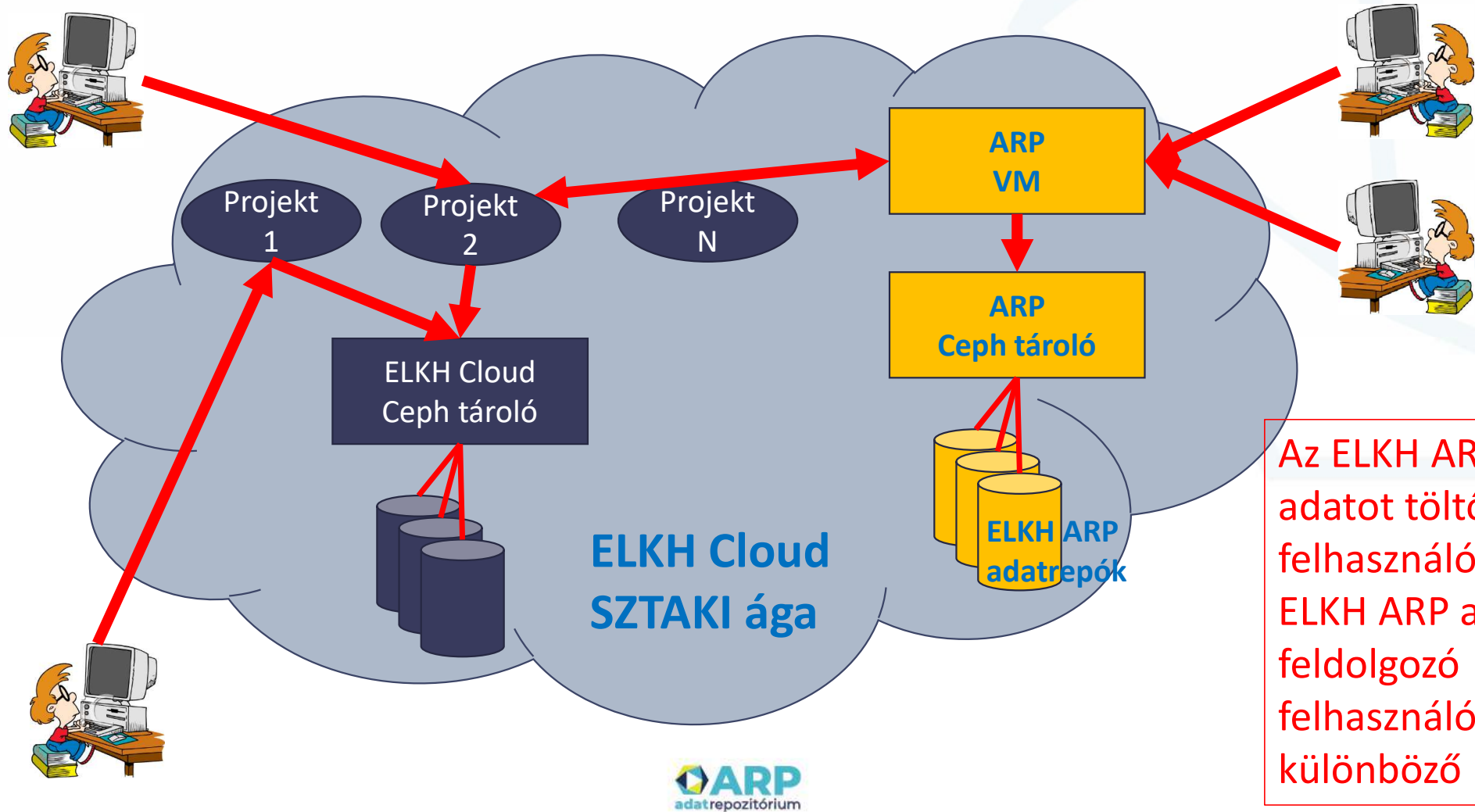
# Infrastruktúra építés, hardver beruházás

- Irányelvek
  - Kétszintű adattárolás létrehozása
    - Központi ARP tároló
    - Lokális, intézeti tárolók
  - A központi ARP tároló az ELKH Cloud kiterjesztése
  - Előnyök
    - Nem kell új infrastruktúrát kiépíteni
    - Olcsóbb és gyorsabban kivitelezhető megoldás
    - Az adat tárolása és feldolgozása ugyanabban az infrastruktúrában történhet
    - Nem kell az adatokat a feldolgozás helyére szállítani

# Adattárolási redundancia a SZTAKI és a Wigner felhő között

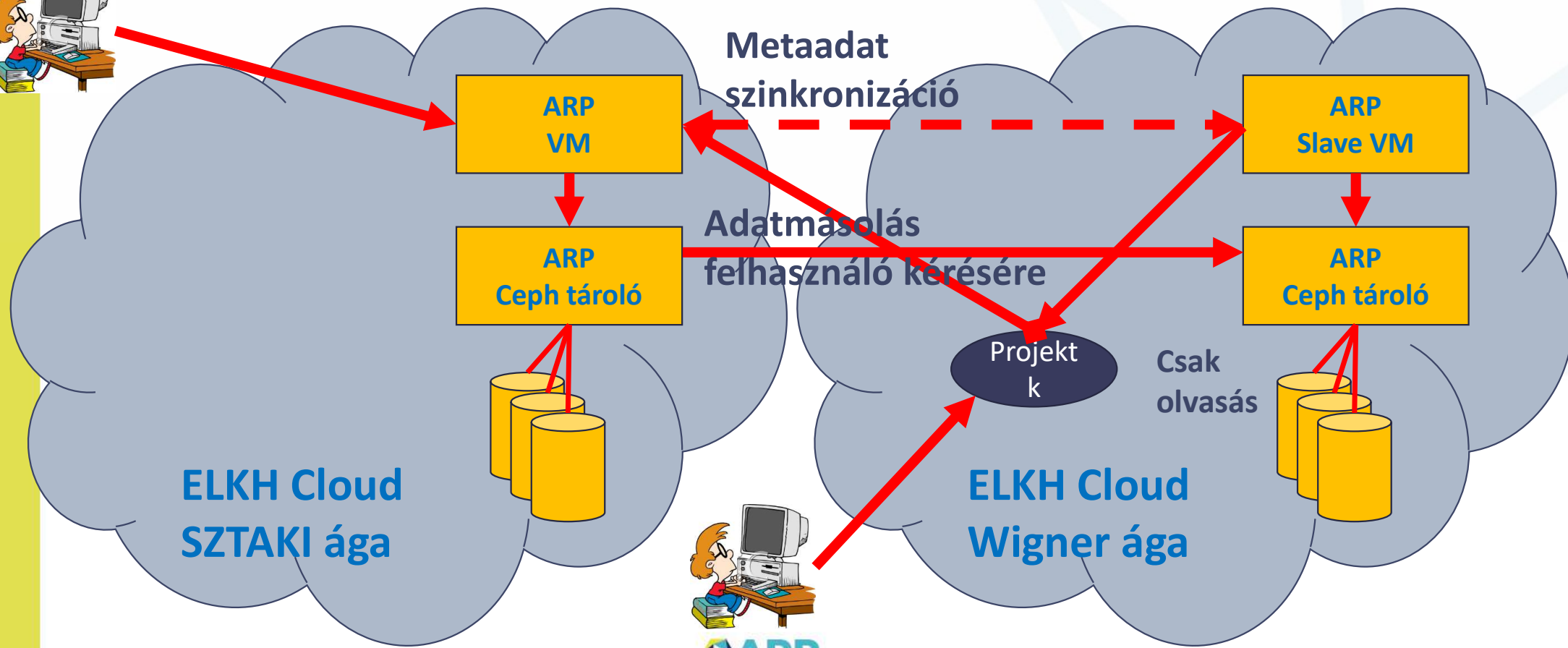
- Meta-adatok szempontjából
  - A két felhő szimmetrikus szerepet játszik
  - Minden meta-adat automatikusan átmásolódik a SZTAKI felhőből a Wigner felhőbe
- Tárolt adatok szempontjából
  - A két felhő aszimmetrikus szerepet játszik
  - Elsődleges adattárolás a SZTAKI ágon, ahol 3-szoros redundancia van
  - Felhasználói kérésre az adatrepok egy része átmásolható a Wigner ágra (ahol szintén 3-szoros redundancia van)
    - **geo-redundancia**
      - Wigner ágon futó projektek fel tudják dolgozni
  - Kérhető szalagos archiválás, ami a Wigner ágon van

# Az ARP, mint az ELKH Cloud kiterjesztése



Az ELKH ARP-ba adatot töltő felhasználó és a ELKH ARP adatot feldolgozó felhasználó különböző lehet

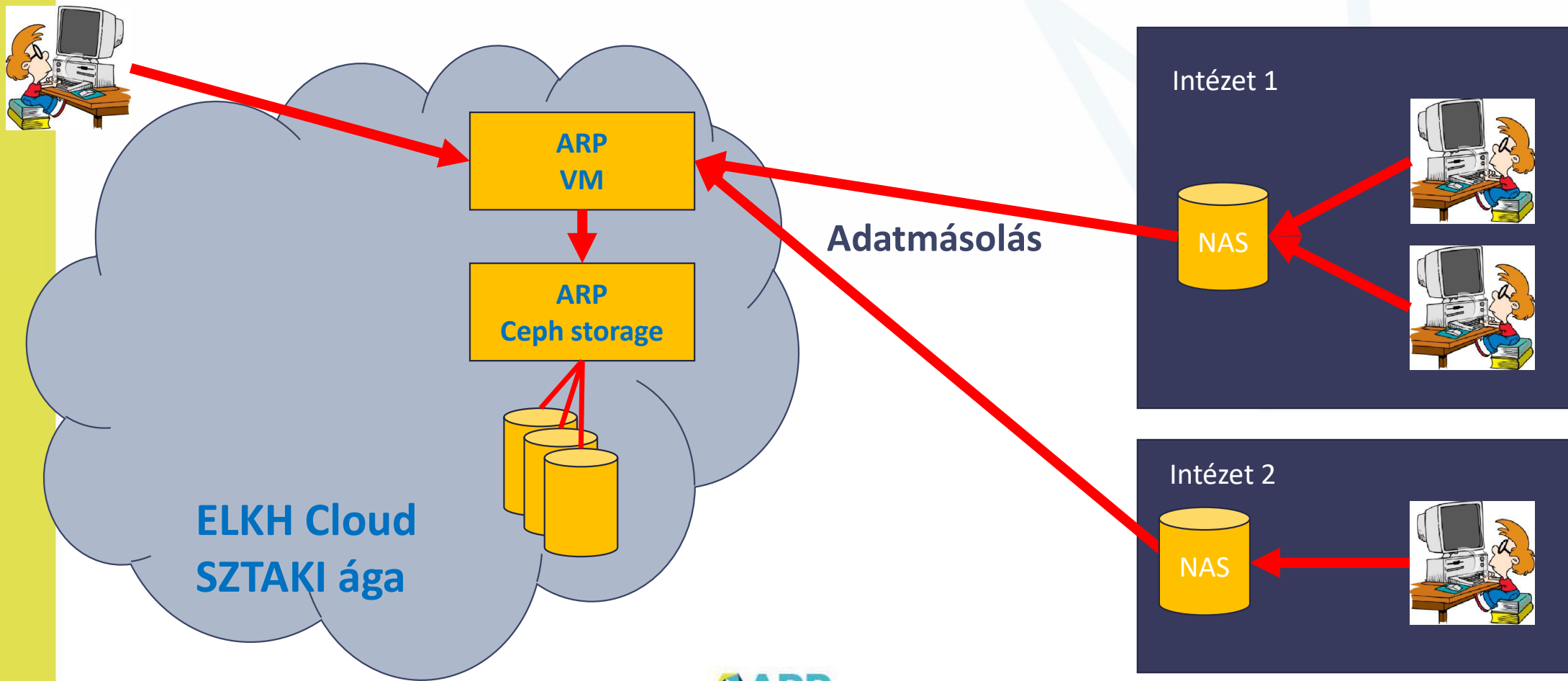
# SZTAKI Cloud és Wigner Cloud együttműködése (aszimmetrikus szerep)



# Lokális, intézeti adatkezelés

- Cél:
  - Az intézeti adatkezelési workflow-k korszerűsítése
  - Ehhez lokális, intézeti adattárolók biztosítása
  - Lokális adattárolókban primer adatgyűjtés
  - A projekt végén, vagy időszakonként nyers adatok metaadatolása és feltöltése a központi ELKH ARP-be
- Fizikai megvalósítás
  - 60 TB-os NAS szerverek alkalmazása
  - Hálózati összeköttetés a NAS és az ELKH ARP között
  - Két intézetben új NAS szerverek felállítása
  - Meglévő NAS szerverek használatának támogatása (pl. Régészeti Intézet)

# ELKH ARP és lokális NAS együttműködése



# Az ELKH ARP megvalósításának fő tulajdonságai

- Aszimmetrikus szerep a SZTAKI és a Wigner felhő között
  - A SZTAKI felhő az elsődleges ARP tároló, az adatok betöltése az ARP-be csak itt lehetséges
  - Adatarchiválás csak a Wigner felhőből lehetséges
- Adat biztonság
  - 3-szoros adattárolási redundancia mind a SZTAKI, mind a Wigner ágon
  - FAIR adattároláshoz szükséges metaadatok geo-redundáns tárolása
  - Adatok archiválási lehetősége a Wigner felhőhöz kapcsolt szalagos archiválás segítségével
- Adatfeldolgozási lehetőségek
  - Mindkét felhőben feldolgozhatók az ARP-ben tárolt adatok, de a Wigner felhő használata esetén:
    - Az adatokat először át kell másolni a SZTAKI felhőből a Wigner felhőbe (automatikus folyamat kérésre)
    - A feldolgozás eredményeképpen keletkező adatokat csak a SZTAKI felhő adatterületére lehet beírni

# Összefoglalás

- Az ARP megvalósulása esetén a magyar kutatók
  - Rendelkeznek majd a FAIR adatkezeléshez szükséges hardver és szoftver infrastruktúrával
  - Képesek lesznek a FAIR adatkezelést beépíteni a kutatási workflow-ikba és azt aktívan használni
  - Képesek lesznek FAIR adat repositóriumokat építeni és használni
  - A fenti képességek lehetővé teszik, hogy aktívan bekapcsolódjanak nemzetközi projektekbe, ill. az EOSC európai infrastruktúrába



**Köszönöm a figyelmet**

Kacsuk@sztaki.hu

