

Az ELKH Adatrepozitórium Projekt (ARP) bemutatása

Kacsuk Péter

SZTAKI

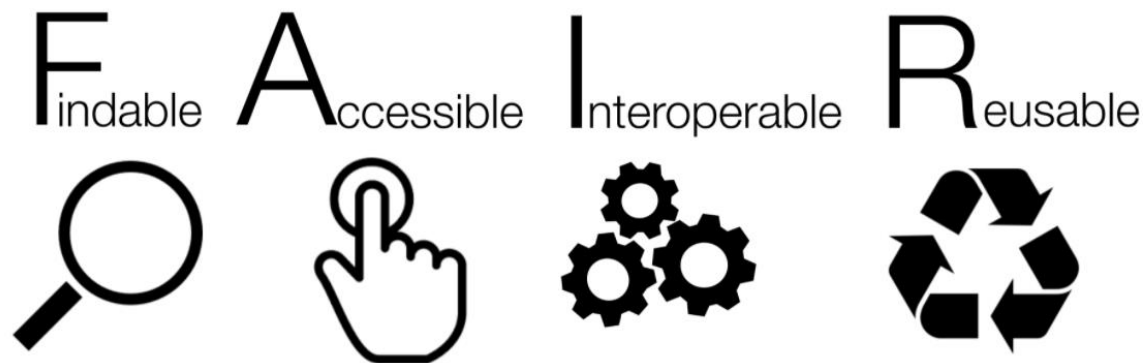
ELKH ARP projekt vezetője



Motivációk és előzmények

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

A projekt célkitűzései

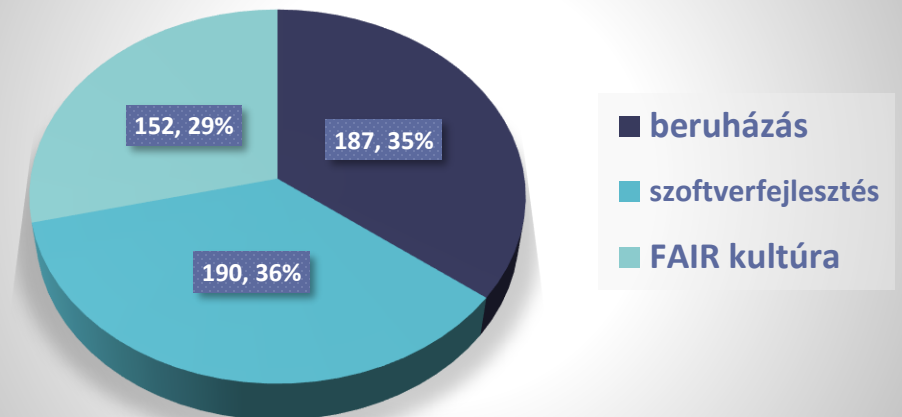


- Olyan új **adatrepozitórium infrastruktúra-szolgáltatás** alapjainak megteremtése, ami a teljes ELKH intézményhálózat folyamatos, hosszútávú kutatási adatkezelését képes lesz támogatni
 - ELKH Cloud projekt -> számítási infrastruktúra megteremtése
 - ELKH ARP projekt -> FAIR adatkezelési infrastruktúra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

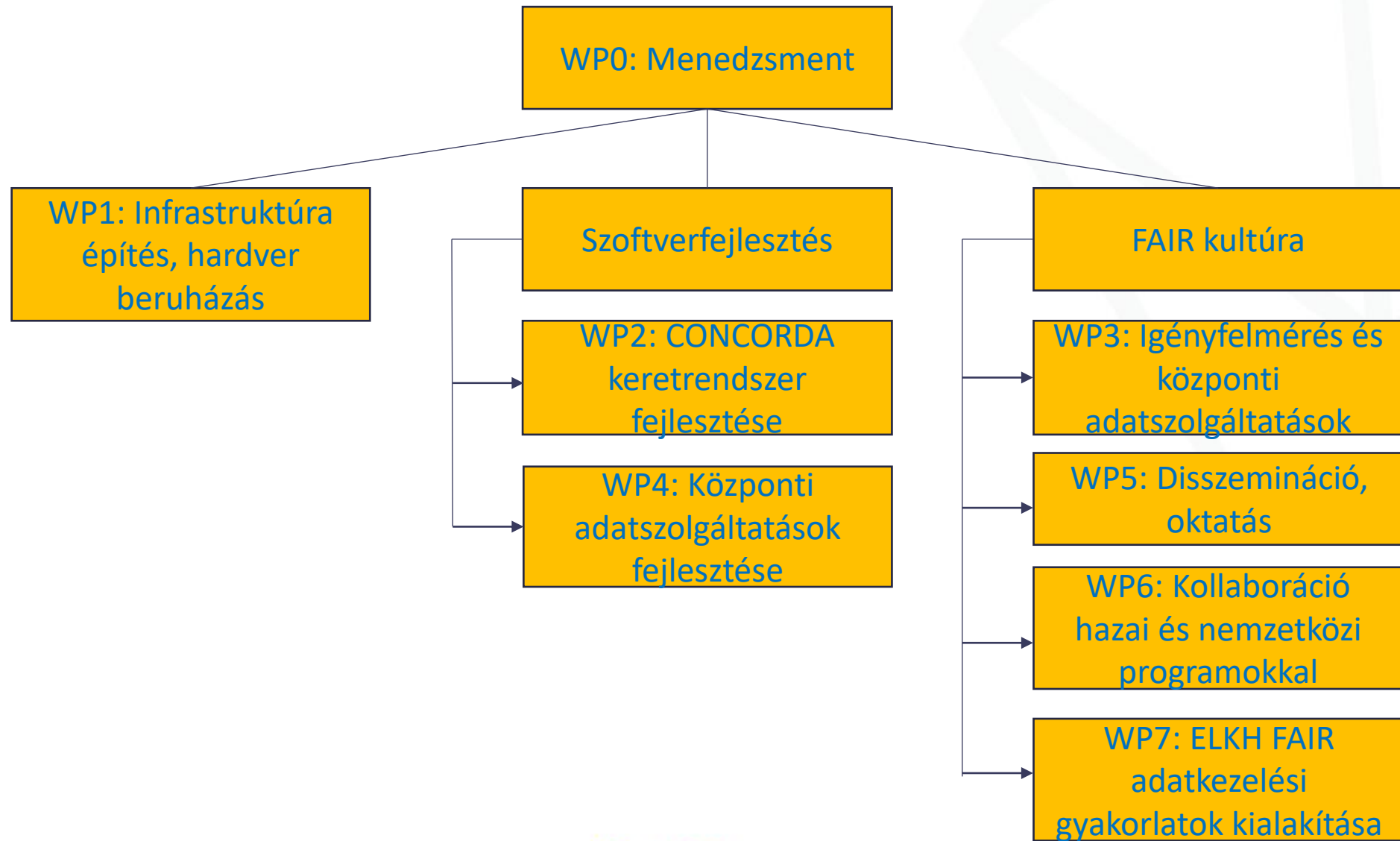
A projekt megvalósítása

- 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



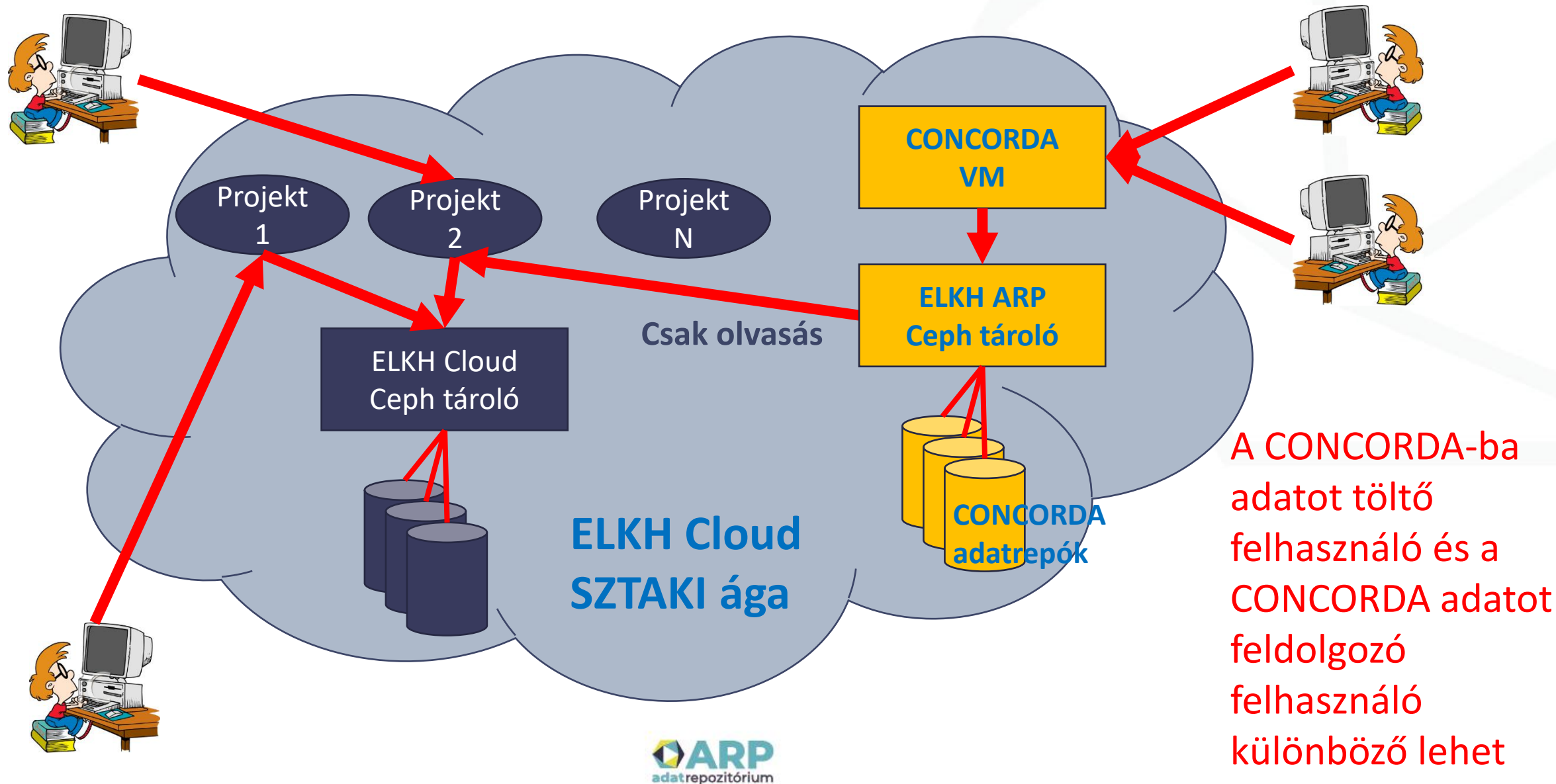
Munkacsomagok



WP1: Infrastruktúra építés, hardver beruházás (Pintér Ádám, WDC)

- 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

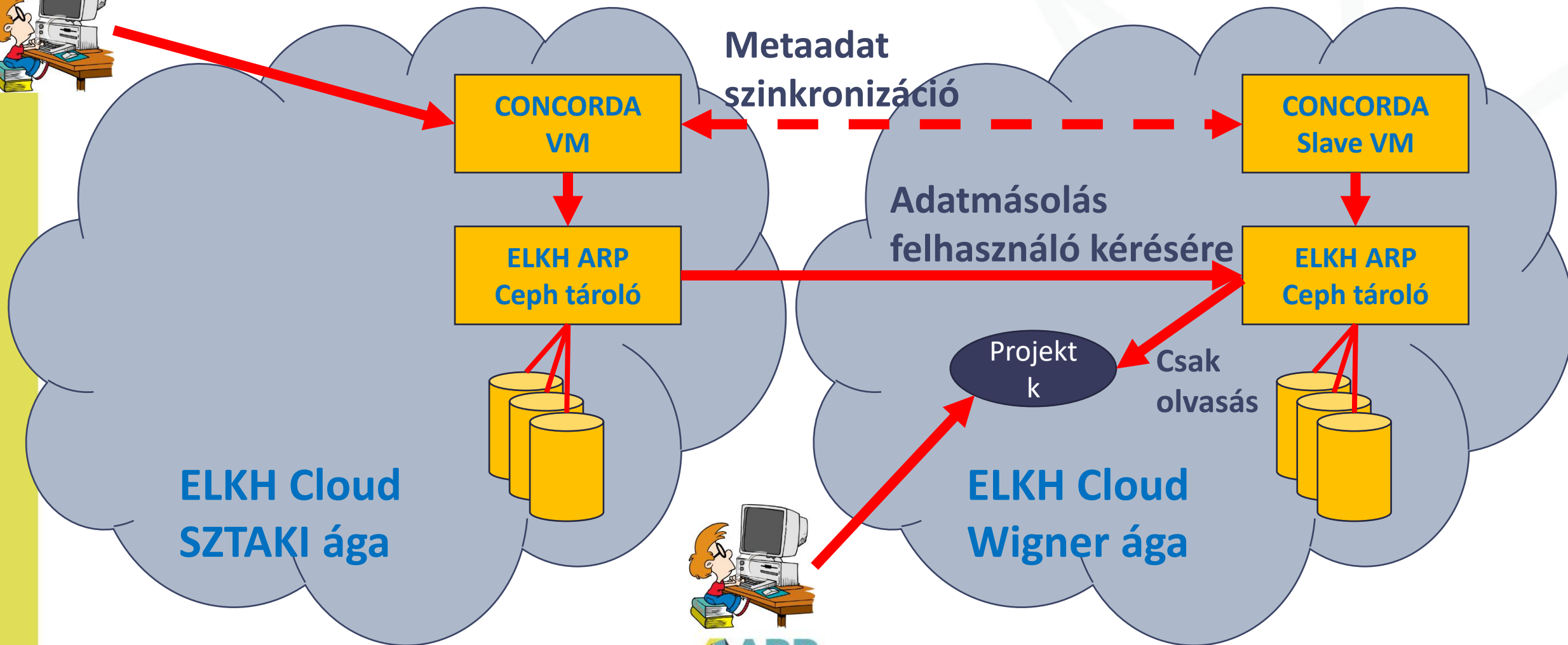
Az ARP, mint az ELKH Cloud kiterjesztése



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 (CONCORDA) 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

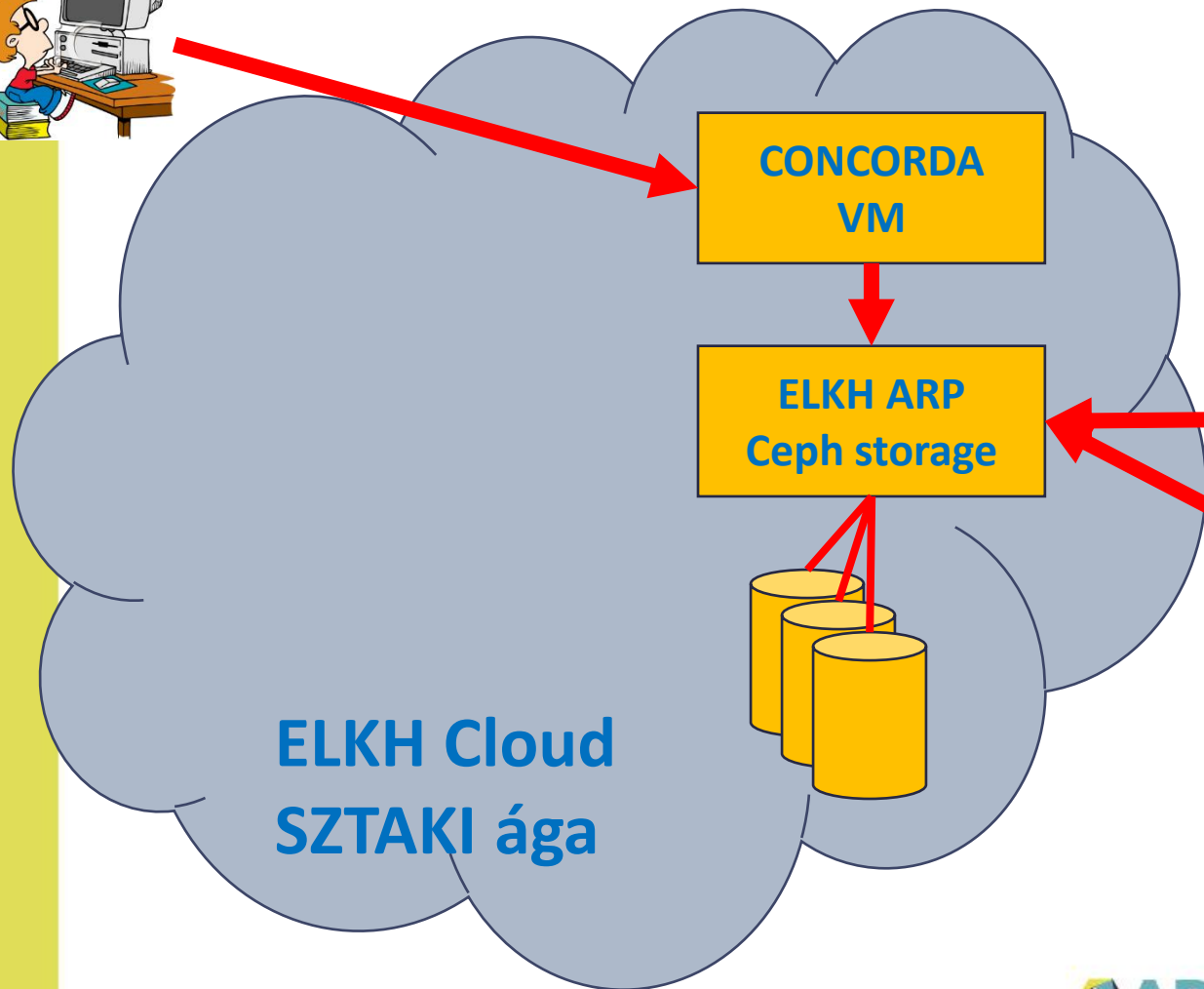
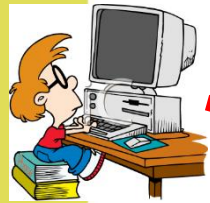
SZTAKI Cloud és Wigner Cloud együttműködése (aszimmetrikus szerep)



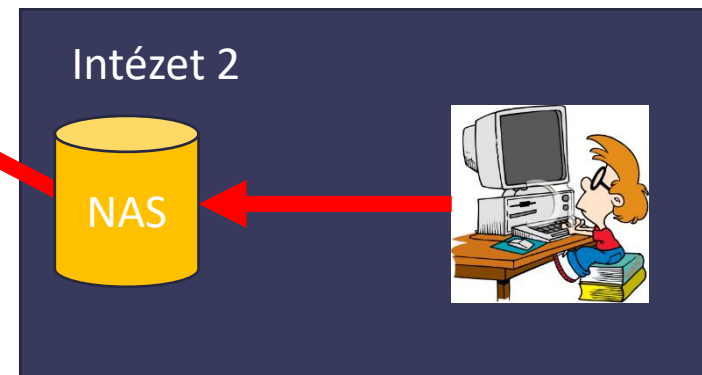
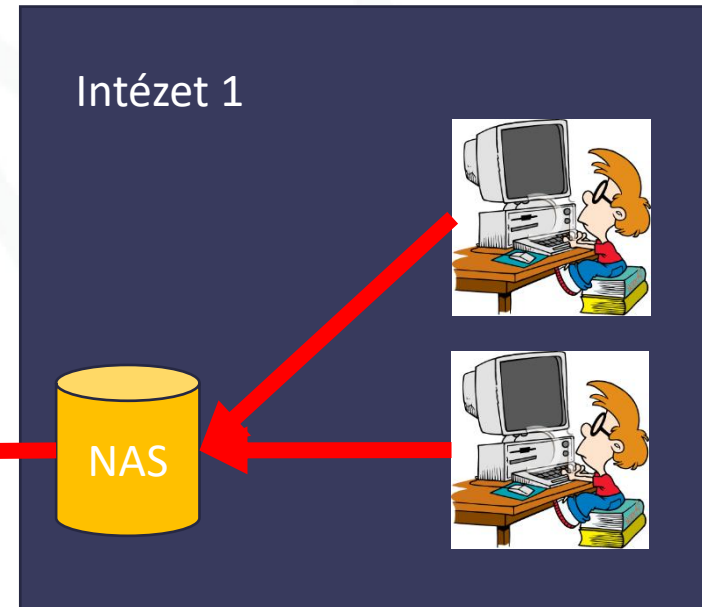
Lokális, intézeti adatkezelés

- Cél:
 - Az intézeti adatkezelési workflow-k korszerűsítése
 - Ehhez lokális, intézeti adattárolók biztosítása
 - Lokális adattárolókban primer adatgyűjtés
 - Projekt végén, vagy időszakonként nyers adatok metaadatolása és feltöltése a központi ELKH ARP-be, mint CONCORDA adatrepó
- 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



Adatmásolás



WP2 és WP4: szoftver fejlesztés (Kovács László, SZTAKI)

- WP2: Repozitóriumkezelő szoftverrendszer (keretprogram) fejlesztése
 - CONCORDA keretrendszer továbbfejlesztése -> CONCORDA-2
- WP4: Központi adatszolgáltatások tervezése
 - metaadatséma-regiszter kifejlesztése
 - közös adatkereső rendszer többféle adatrepóhoz
 - adattranzformációs funkciók könyvtára
 - Real-time adatelemzési és vizualizációs szolgáltatások
- További cél:
 - a tudományterületek egyedi igényeit is figyelembe venni

WP3, WP5-7: FAIR kultúra kialakítása

- WP3: Igényfelmérés és központi adatszolgáltatások tervezése (Meiszterics Enikő, TK)
 - Kérdőíves felmérés, melyre 17 ELKH intézményből 129 kitöltött válasz érkezett
 - 48 interjú 16 ELKH intézményben
 - Begyűjtött tapasztalatokat betápláljuk a CONCORDA-2 tervezési folyamatába
- WP5: Disszemináció, oktatás (Gárdos Judit, TK)
- WP6: Együttműködés hazai és nemzetközi programokkal és projektekkel (Kacsuk Péter, SZTAKI)
- WP7: ELKH FAIR adatkezelési gyakorlatok kialakítása (Meiszterics Enikő, TK)

WP5: ELKH ARP weblap science-research-data.hu



Rólunk [Aktualitások](#) ▾ [Kapcsolat](#) [Érdekességek](#) ▾

Hírek



📅 2022. április 1.

Handle típusú perzisztens azonosítók bevezetése

Az ELKH ARP alapjául szolgáló Concorda adatrepozitórium mostantól handle típusú perzisztens azonosítókat oszt az adatsomagok számára.

[Tovább →](#)



📅 2022. január 27.

Lezárult a HRDA, az MTA KIK és az ELKH Titkárság kutatási adatok archiválását támogató projektje

Kutatásiadat-archiválási pilot-projektek az Eötvös Loránd Kutatási Hálózathoz tartozó kutatóközpontokban, -intézetekben és -csoportokban címmel,



📅 2022. január 27.

Elindult a nemzetközi kutatási infrastruktúrákba való bekapcsolódást támogató ELKH Adatrepozitórium fejlesztése

Az Eötvös Loránd Kutatási Hálózat (ELKH) három kutatóhelye, a Számítástechnikai és Automatizálási

WP6: Kapcsolódás nemzetközi és hazai programokhoz és projektekhez

- 4 fő tevékenységi terület:
 - **Együttműködés a HRDA-val**
 - Együttműködés az RDA-val
 - Együttműködés az EOSC-kal
 - Együttműködés tudomány-specifikus hazai és nemzetközi adatrepó projektekkel

Az eddigi eredmények a számok tükrében

- Eltelt **5** hónap
- Készítettünk **48** interjút **16** ELKH intézményben
- Kérdőíves felmérésünkre **129** kitöltött válasz érkezett **17** ELKH intézményből
- A CONCORDA repozitóriumkezelő szoftverrendszernek már **51** felhasználója van, akik összesen **2.5 TB** tárhelyet használnak, összesen kb. **7300 fájlal**. Jelenleg **1199 adatcsomag** van használatban és ebből **1180 nyilvános**.
- A felhasználók **16** intézményből vannak

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 repozitóriumokat építeni és használni
 - A fenti képességek lehetővé teszik, hogy aktívan bekapcsolódjanak nemzetközi projektekbe