



A HUN-REN Adatrepozitórium Platform (ARP) projekt ismertetése

Dr. Kacsuk Péter DSc
HUN-REN SZTAKI

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
- HUN-REN Cloud, mint **számítási** felhő sikere
- Ugyanakkor a HUN-REN Cloud nem adattárolási felhőként működik
- HUN-REN intézmények adatkezelési munkafolyamata sok helyen nagyon elavult

A projekt célkitűzései

F_{indable}



A_{ccessible}



I_{nteroperable}



R_{eusable}

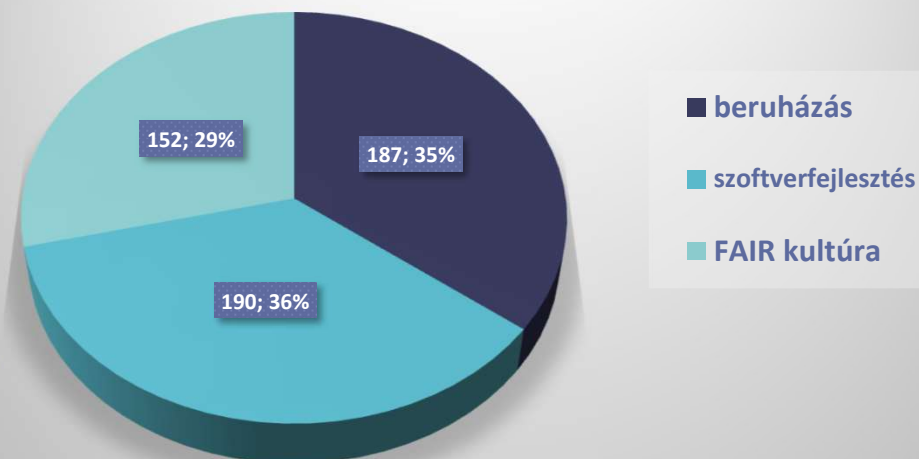


- Olyan új **adatrepozitórium infrastruktúra-szolgáltatás** alapjainak megteremtése, ami a teljes HUN-REN intézményhálózat folyamatos, hosszútávú kutatási adatkezelését képes lesz támogatni
 - HUN-REN Cloud projekt -> számítási infrastruktúra megteremtése
 - HUN-REN ARP projekt -> FAIR adatkezelési infrastruktúra megteremtése
 - A kettő együtt ugyanazt adja a magyar kutatóknak, mint az EOSC európai szinten
- A FAIR adatrepozitórium **kultúra** megteremtése a HUN-REN intézményhálózatában
 - A meglévő **kutatási munkafolyamatok** kibővítése a FAIR adatrepozitóriumok használatával
- Bekapcsolódás hasonló célú **európai kezdeményezésekbe**

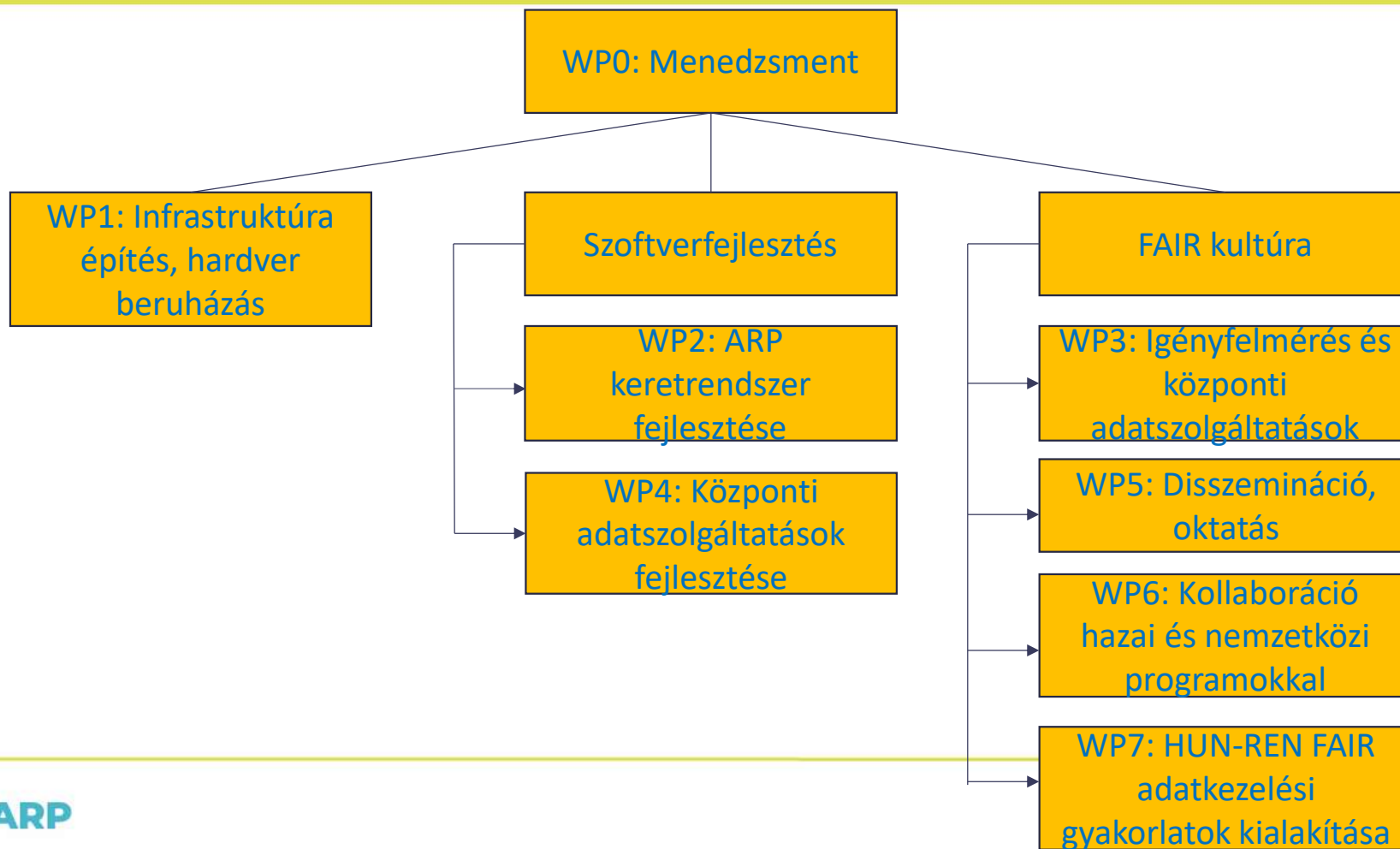
A projekt megvalósítása

- A projekt megvalósításában résztvevő HUN-REN intézmények:
 - SZTAKI, TK, WIGNER FK
- A projekt megvalósításának tervezett időtartama:
 - 18 hónap: 2021. nov. 1. – 2023. ápr. 30. -> **2023. dec. 31.**
- A projekt költségvetése: 529 millió Ft
 - Beruházás: 187 mFt
 - Szoftver fejlesztési költség: 190 mFt
 - FAIR adatrepó kultúra kialakítása: 152 mFt
- **A projekt támogatója: HUN-REN**

ARP költségek felosztása



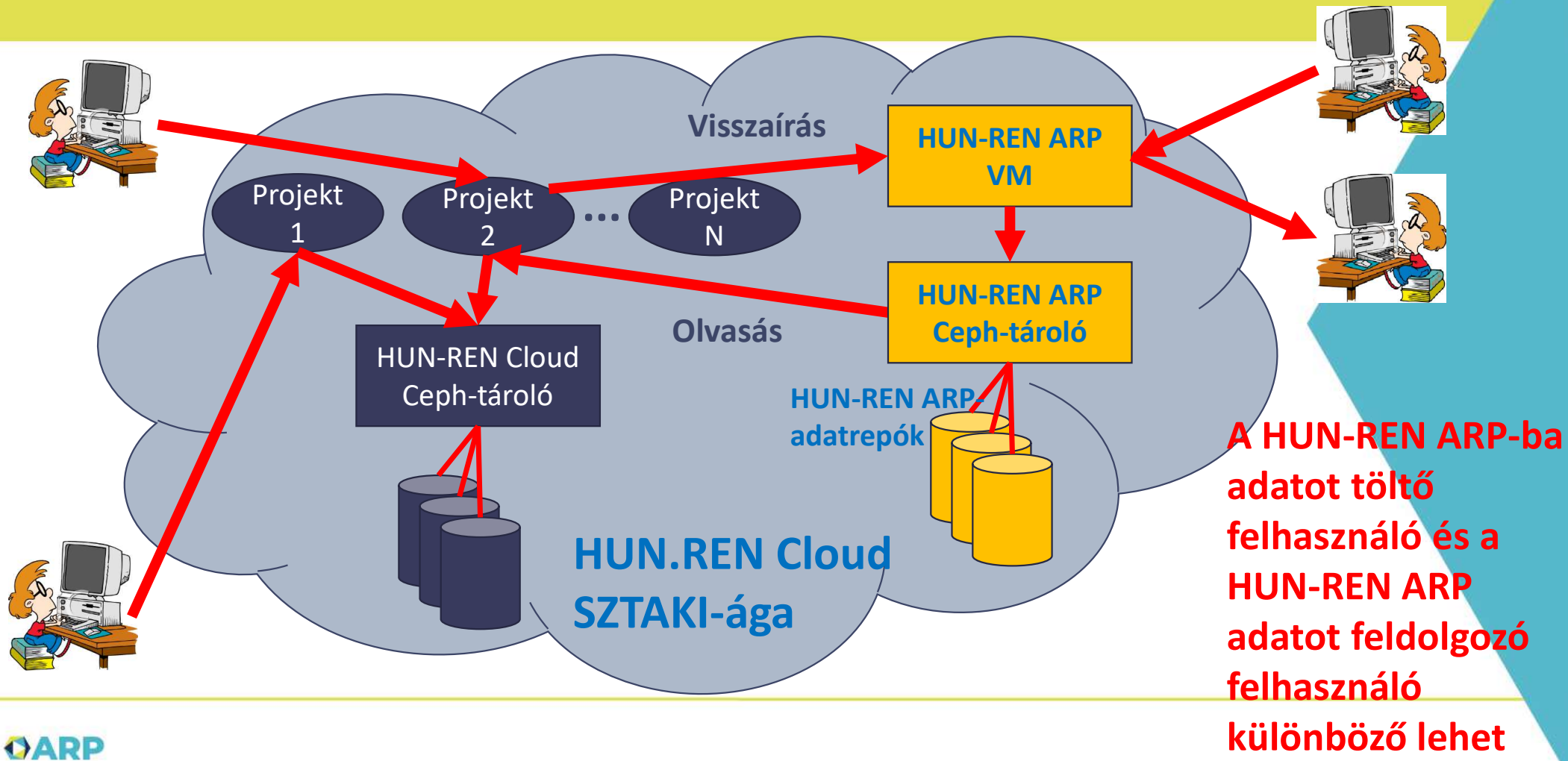
Munkacsomagok



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

- Irányelvek
 - Kétszintű adattárolás létrehozása
 - Központi ARP tároló
 - Lokális, intézeti tárolók
 - A központi ARP tároló a HUN-REN 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 (ami komoly probléma az EOSC esetén)

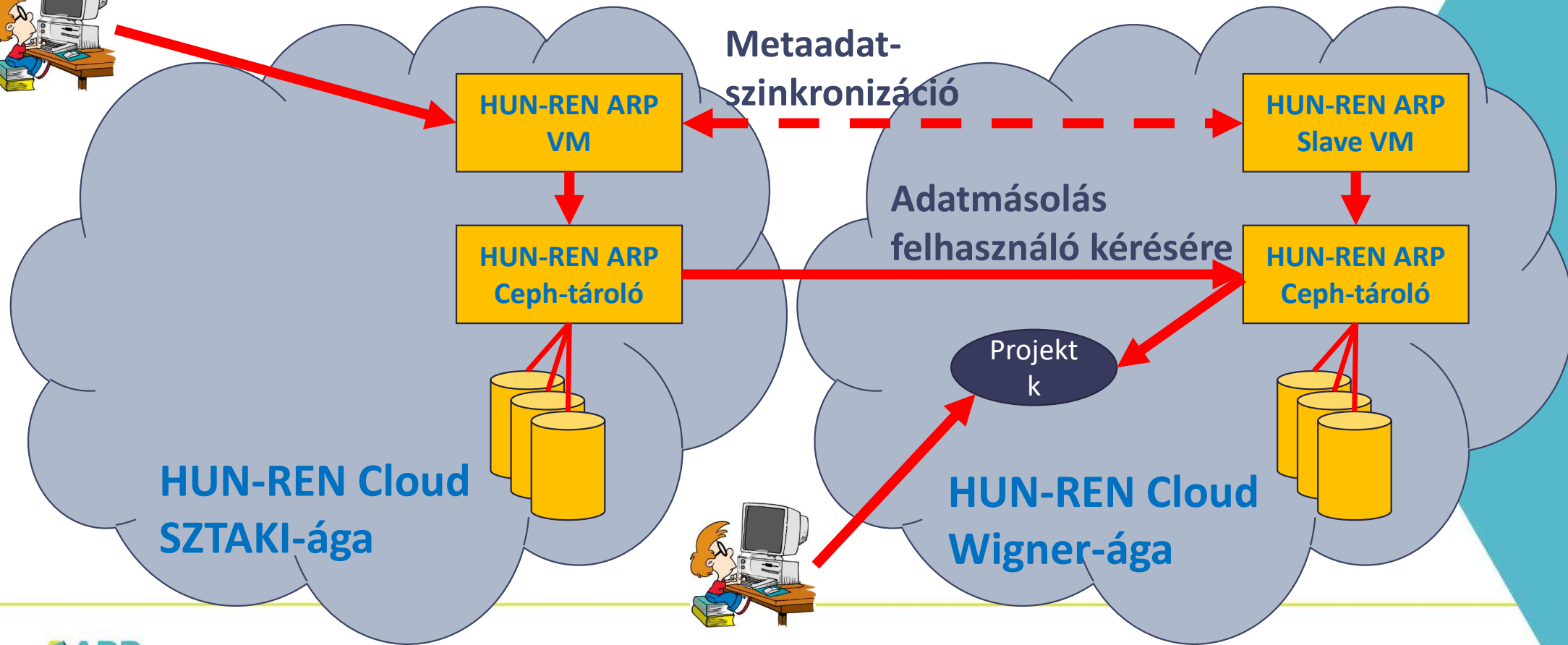
Az ARP mint a HUN-REN Cloud kiterjesztése



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 a SZTAKI ágon, ahol 3-szoros redundancia van
 - Felhasználói kérésre az adatok 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

A SZTAKI Cloud és a Wigner Cloud együttműködése (aszimmetrikus szerep)



Szalagos archiváló rendszer kiépítése

- Wigner FK Adatközpont ágon Oracle Tape Library (SL3000) robotizált szalagos tárolórendszer
- Hosszú távú adatmegőrzés (30 évre tervezve)
- Elérési protokoll: Network File System (NFS)
- Négy darab T10000D drive
- HUN-REN ARP számára 50 darab 8 TB-os T10000D szalag
- Szalagokra írásnál lehetőség van két kópiás beállításra (egy fájl két különböző szalagra íródik ki)
- Két kópiás beállítással összesen 200 TB szalagos területet eredményez HUN-REN ARP számára
- Hitachi VSP G200 diszkalrendszer (100 TB cache: elsődleges tárterület)



Szalagos archiváló rendszer használata

- Cartridge Access Port: szalagok kivétele/újak behelyezése a rendszerbe
- Funkciók:
 - 1. Archiválás: elsődleges tárterületről => szalagra, archiválási ciklus négy óránként fut le automatikusan
 - 2. Felszabadítás: a fájl szalagon elérhető, de az elsődleges tárterületről kitöröljük (automatizált folyamat a telítettségtől függően)
 - 3. Visszatöltés: szalagról => elsődleges tárterületre, olvasási kérés esetén automatikusan
 - 4. Újrahasznosítás: A szalagokon törölt adatok helyére más adat írható (jelenleg ezt nem használjuk)



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

- WP2: Repozitóriumkezelő szoftverrendszer (keretprogram) fejlesztése
 - CONCORDA (Dataverse alapú) keretrendszer továbbfejlesztése -> **HUN-REN ARP**
- WP4: Központi adatszolgáltatások tervezése
 - metaadatséma-regiszter kifejlesztése
 - közös adatkereső rendszer többféle adatrepóhoz

WP3: Igényfelmérés és központi adatszolgáltatások (Meiszterics Enikő, TK)

Az információgyűjtést a SZTAKI és a TK munkatársai közösen koordinálták és végezték

- online kérdőíves megkérdezés és
- online interjú adatfelvételek megvalósításával,
- a teljes kutatási hálózatban (11 kutatóközpont, 7 kutatóintézet)
- és a támogatott kutatócsoportok kutatói körében.

A cél az volt, hogy mindegyik HUN-REN intézményt érjük el legalább az egyik módszerrel.



Kutatói igényfelmérés-információk feldolgozása

Online kérdőíves
felmérés adatai



A két adatfelvétel
releváns
információi
együttesen

szoftverspecifikáció készítése



Interjúk leiratai

tartalomelemzés



ARP Nagykövet-pályázat és program



ARP nagyköveteket keresünk!

 2023. február 23.

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

- 6 ARP nagykövet
- 6 intézmény
- Feladatok:
 - ARP kipróbálása, „megtanulása”, adatfeltöltés, tesztelés
 - képzésen való részvétel
 - intézményi munkafolyamat kialakítása
 - ismeretátadás intézményen belül és kívül

ARP tesztelés (alfa- és bétateszt)

Az ARP alfateszt során a résztvevők feladata volt

- felderíteni, hogy a különböző adaptált szoftver komponensek (Dataverse, Cedar) megőrizték-e az eredeti hivatalos verzió minden képességét
- ezek az installációk hogyan férnek össze az AROMA szoftvermodullal,
- feltárni és jelezni az AROMA felhasználói kézikönyv hiányosságait,
- megismerni az AROMA és az RO-Crate szoftver modulok lehetőségeit és korlátait

Az ARP bétatesztben résztvevők feladata

- a rendszer teljes működését lefedő munkafolyamatokkal
 - ellenőrizni a teljes infrastruktúra együttes működését,
 - felderíteni az esetleges integrációs hibákat, és
 - rávilágítani a hiányzó menedzsment folyamatokra.

WP5: Disszemináció (Gárdos Judit, TK)

Rendezvények és külső megjelenések az ARP projektben

Megjelenések típusai:

1. Saját szervezésű (**7** esemény)
 - Pl. Tájékoztató az OTKA pályázatokhoz kötelezően feltöltendő kutatási adatkezelési tervekről
2. Külső szervezésű (pl. konferencia), ARP-os előadással való részvétel (**6** esemény)
3. Egyéb megjelenések (pl. publikáció)
 - Magyar Tudomány folyóiratban az ELKH ARP tematikus blokk Gárdos Judit szerkesztésében (**5** cikk)

Kincs ami van rendezvénysorozat

- Saját (HRDA-val közös) szervezésű **6 esemény**:
 - A kutatási objektum (**1** előadás)
 - Eltűnő értékek nyomában: kutatási adatok újrahasznosítása (**2** előadás)
 - Környezeti adatok - Föld, víz, levegő (**4** előadás)
 - Csillagászati adatok – határ a csillagos ég? (**2** előadás)
 - Mindenki az adatokért, az adatok mindenkiért! (**3** előadás)
 - Miből lesz az adattár? A tárgyi gyűjteményektől a digitális adattárak felépítéséig (**3** előadás)
- **Összesen 15 előadás**

Külső szervezésű események

1. 2022. évi Networkshop (**5** ARP előadás)
2. 4. Nyílt Tudományos Fórum - Nyílt tudományos gyakorlatok az élettudományokban és a társadalomtudományokban (**1** ARP előadás)
3. A Magyar Tudomány Ünnepe rendezvénysorozat (**4** ARP előadás)
4. OntoPortal Alliance meeting (**1** ARP előadás)
5. 2023. évi Networkshop (**8** ARP előadás)
6. kolozsvári Babes-Bolyai Egyetemen rendezett nemzetközi Aktivizmus c. konferencia (**1** ARP előadás)
7. Nemzetközi fejlesztői projekteken (Dataverse, Describo, CEDAR) (**4** ARP előadás)

Összesen 24 előadás

ARP információs portál I.

Élesítés: 2022 március

Tartalom

- Tevékenységek, infrastruktúra-fejlesztés bemutatása
- Rendszeresen frissülő hírek, események, ismeretterjesztő anyagok, előadások, publikációk, GYIK stb.
- ARP nagykövetek bemutatása
- Kutatási adatkezelési útmutató
- Hírlevél feliratkozási lehetőség
- Együttműködési felhívások

HUN-REN Adatrepozitórium Platform

EN|HU

ADATREPOZITÓRIUM

Rólunk Hírek Események Tudástár

Tudományos Adatrepozitórium a magyar kutatói közösség számára

A HUN-REN ARP egy olyan digitális infrastruktúra, amely lehetővé teszi a kutatási adatok hosszú távú megőrzését, szolgáltatását, valamint szabályozott és megbízható megosztását.

Bővebben >

Fókuszban

Nyílt tudomány, nyílt adatok Magyarországon

2023. nov. 16. 10:00 - 16:45

SZTAKI Nagytanács terem

2023 november 16-án rendezzük meg a HUN-REN Adatrepozitórium Platform (ARP) első két évét összefoglaló konferenciát.

Tovább >

Miből lesz az adattár? - A tárgyi gyűjteményektől a digitális adattárak felépítéséig

Kincs, ami van eseménysorozat

2023. nov. 7. 10:00 - 11:00

HTK Jogtudományi Intézet: Tárgyaló és Online

Az ARP és a HRDA közös Kincs, ami van

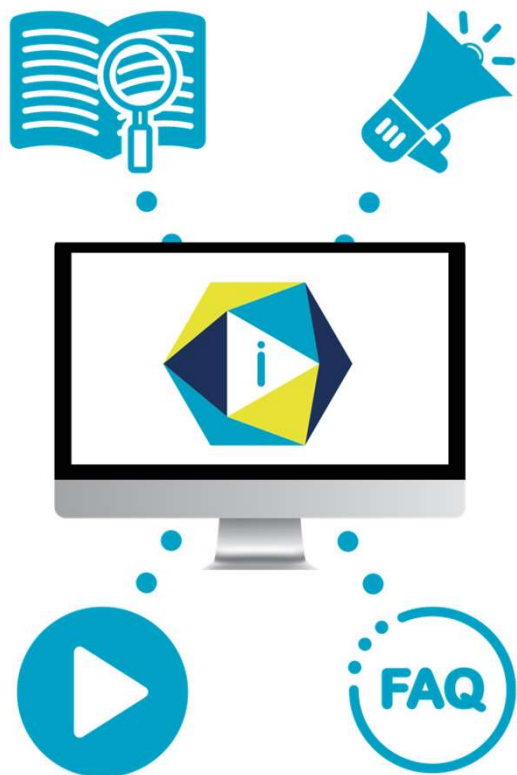
Kutatási adatok kezelése a bölcsészettudományok területén

2023. szeptember 22.

Scheibner Tamás, az ARP egyik nagykövete előadást tartott a kutatási adatok kezeléséről és az ARP-ról.

Tovább >

ARP információs portál II.



Tájékoztatás, ismeretterjesztés számokban

- **26** esemény meghívó
- **33** előadás
- **21** ismeretterjesztő tudástár elem
- **27** fogalomtár elem
- **23** hír
- **5** publikáció

WP6: Kollaboráció hazai és nemzetközi programokkal (Kacsuk Péter, SZTAKI)

5 fő tevékenységi terület:

- Együttműködés az RDA-val (Research Data Alliance)
 - Részvétel 4 RDA Interest Group munkájában
- Együttműködés a HRDA-val (Hungarian Research Data Alliance)
 - Közös események szervezése
 - Közös megjelenés az RDA-ban
- Együttműködés az EOSC-kal (European Open Science Cloud)
 - Részvétel az EOSC Symposium konferenciákon
- Együttműködés tudomány-specifikus hazai és nemzetközi adatrepó projektekkel
 - Együttműködés előkészítése magyar ERIC képviselőkkel
- Együttműködés az amerikai Dataverse és a nemzetközi RO-CRATE fejlesztőkkel

WP7: HUN-REN FAIR adatkezelési gyakorlatok kialakítása (Meiszterics Enikő, TK)

- Háttérinformációk összegyűjtése, elemzése, közzététele
 - Információs adatbázis építése a kutatói igényfelmérésben megnevezett repozitóriumokból, adatbázisokból, platformokból, szoftverekből (a kutatás lezárása után közzétesszük az ARP repozitóriumában)
 - Tudományos folyóiratok, kiadók open data kutatási adat elvárásainak feltárása (cikk a Magyar Tudományban)
 - Közösen szerkeszthető fogalomtár létrehozása, gondozása (elérhető a projekt honlapjáról: <https://science-research-data.hu/fogalomtar>)
 - PID tanulmány (a kutatás lezárása után közzétesszük a REAL-ban)



© Can Stock Photo

WP7 aktivitások

- **Szakmai egyeztetések repozitórium létrehozását tervező intézményekkel**
 - Témák
 - ARP-projektben elért eredmények bemutatása
 - A másik fél tapasztalatainak megismerése
 - Egyeztetés az együttműködési lehetőségekről
 - Partnerek
 - SZTE Klebersberg Kúnó könyvtára, PTE ETKK, ELTE DH
- **Policy-k, ajánlások kidolgozása**
 - Útmutató a HUN-REN kutatók számára kutatási adataik kezeléséhez (a projekt oldaláról online elérhető)
 - HUN-REN adatkezelési keretszabályzat
 - ARP Term of Use

Összefoglalás

Az ARP eredményeképpen a magyar kutatók

- Rendelkezni fognak a FAIR adatkezeléshez szükséges hardver és szoftver infrastruktúrával
- Képesek lesznek a FAIR adatkezelést beépíteni a kutatási munkafolyamataikba és azt aktívan használni
- Képesek lesznek FAIR adat repozitóriumokat építeni és használni
- A fenti képességek lehetővé teszik, hogy aktívan bekapcsolódjanak nemzetközi projektekbe