

ELKH ARP

A CONCORDA - 2 kutatási adatrepozitórium rendszer

Kovács László

SZTAKI DSD

laszlo.kovacs@sztaki.hu

(Elhangzott a Networkshop 2022 konferencián (2022 Debrecen))

SZTAKI DSD Elosztott Rendszerek Osztály dsd.sztaki.hu

DSD

CONTENTS

Projects

Services

Products

References

Publications

People

Contact

DSD
DEPARTMENT OF
DISTRIBUTED
SYSTEMS

MTA SZTAKI

en hu

Copyright © 1994-2013 MTA SZTAKI - DSD.

A tudomány







- A tudomány művelése egyre inkább digitális alapú
- A szférában különféle típusú digitális objektumok keletkeznek
- Megjelenik a digitális objektum menedzsment - elvárás
- Nagyfokú (folyamat és objektum) transzparencia - elvárás
- A digitális objektumok menedzsmentje a fizikai objektumok menedzsmentjéhez képest sokkal bonyolultabb, munkaigényesebb és speciális szakmai hozzáértést igényel

- Kik, milyen intézményekben fogják a digitális objektumok kezelését elvégezni
 - kutató intézet, egyetem?
 - könyvtár?, levéltár?, múzeum?
 - új memória intézmény?
- Hálózati kutatási infrastruktúrák már léteznek különféle fenntartásban
 - domain specifikus közösségi intézmények
 - profit és/vagy nem profit orientált cégek
 - kutatóintézetek, egyetemek

- adatvirtualizáció
 - adattárolás a felhőben/felhő federációban
 - adattárolás a felhőhöz kapcsolt repozitóriumokban
 - tudományok adatainak virtuális egyesítése
- multisémás interoperábilis adattárolás és adathozzáférés
- kapcsolt adatok kezelése, adatfelhő integráció
- szemantikus adattárolás, szemantikus repozitórium

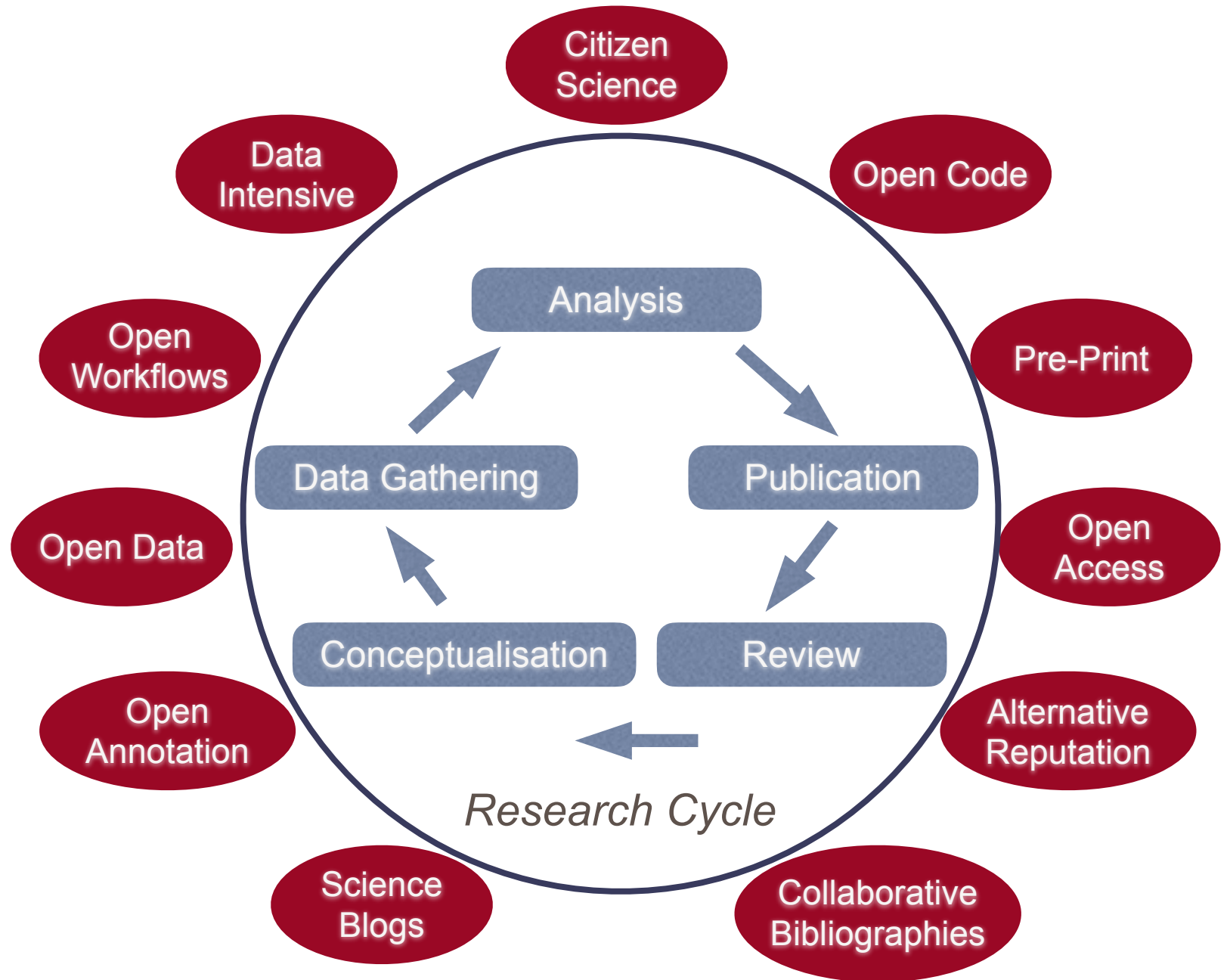
- Adataggregációs tendenciák
 - nemzeti szintű adataggregáció - egységes országos szintű aggregációs igény (a magyar adatok nemzetközi láthatósága növelése érdekében)
 - intézményközi/diszciplináris adataggregáció (multidiszciplináris kutatások érdekében)
 - aggregáció szemantikus technológiák alkalmazásával (magasabb szintű tudáslétrehozás érdekében)
 - aggregáció az egységes kereshetőség érdekében

- standardizált metaadasémák és sémaregiszterek
- a kutatásokhoz szükséges standardizált adatkészletek
 - névterek, ontológiák, lexikonok, tezauruszok, ...
- adatkészletek karbantartásának (javítás, tisztítás, ...) támogatása
 - egyéni, csoportos, közösségi (crowdsourcing)
- hosszú távon is biztonságos tárolás és digitális megőrzés
 - OAIS/ISO szabványos megoldások

- tudománypolitikai ajánlások és mozgalmak (pl. Open * mozgalmak, Science 2.0)
- adatkezelési szabványok és ajánlások terjedése (világ, EU, nemzeti, közösségi szinteken)
- kutatómódszertani javaslatok, szabványos adatkezelési/ tudományos workflow követelmények terjedése
- adatkezelési/tudományos workflow-k megosztása
- FAIR adatkezelés, etikai normák előtérbe kerülése

SCIENCE 2.0

- Open Access
- Open Data
- Open Code
- Open Annotation
- Open Lab Note/Workflows
- Alternative Reputation System
- Collaborative Bibliographies
- Science Blog
- Data Intensive
- Citizens Science
- Pre-Print



ELKH ARP (Adatrepozitórium) projekt

Az ELKH intézményhálózat kutatási adatkezelését támogató adatrepozitóriumok rendszerének, mint egy folyamatosan működő kutatási federált adatrepozitóriummi infrastruktúra alapjainak megteremtése, központi szolgáltatásainak kifejlesztése és működésbe állítása.

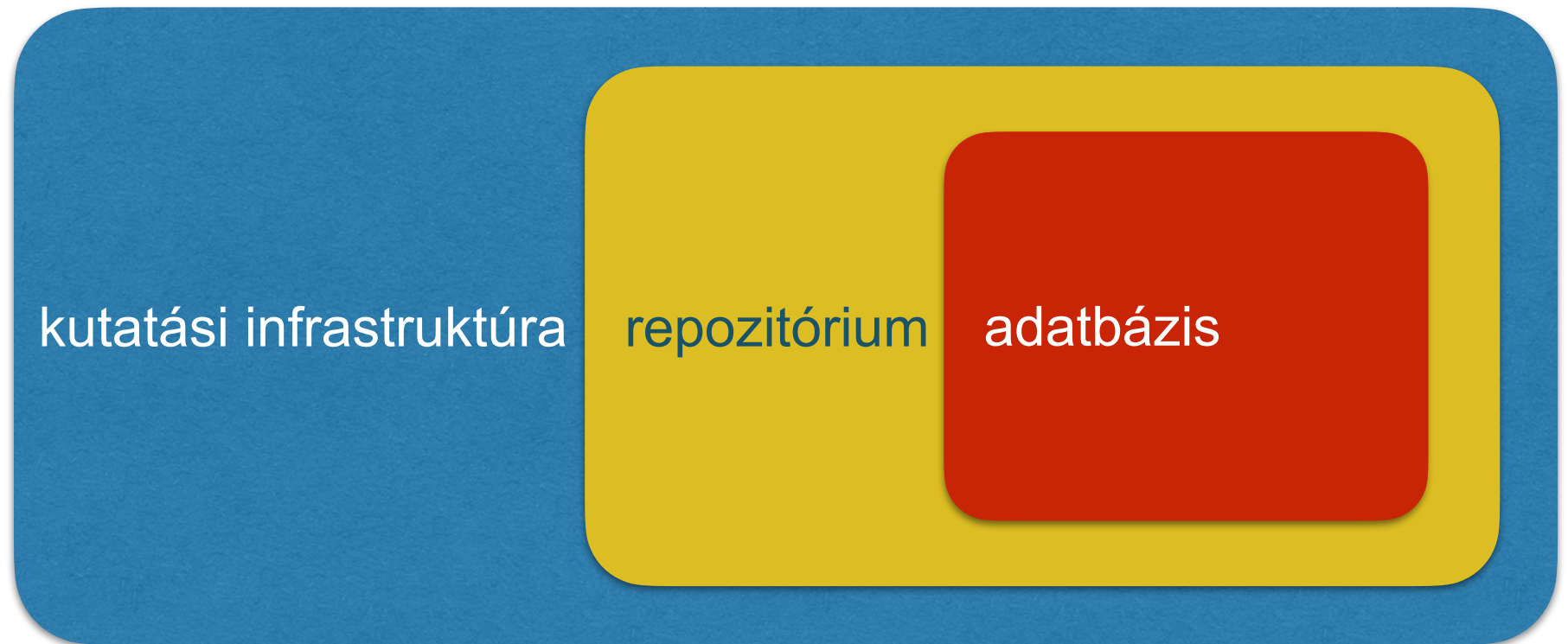
Az ELKH intézményhálózat kutatási adatkezelését támogató adatrepozitóriumok rendszerének, mint egy **1. folyamatosan működő** **2. kutatási** **3. federált** adatrepozitórium **4. infrastruktúra** alapjainak megteremtése, **5. központi szolgáltatásainak** kifejlesztése és működésbe állítása.

A repozitórium fogalma

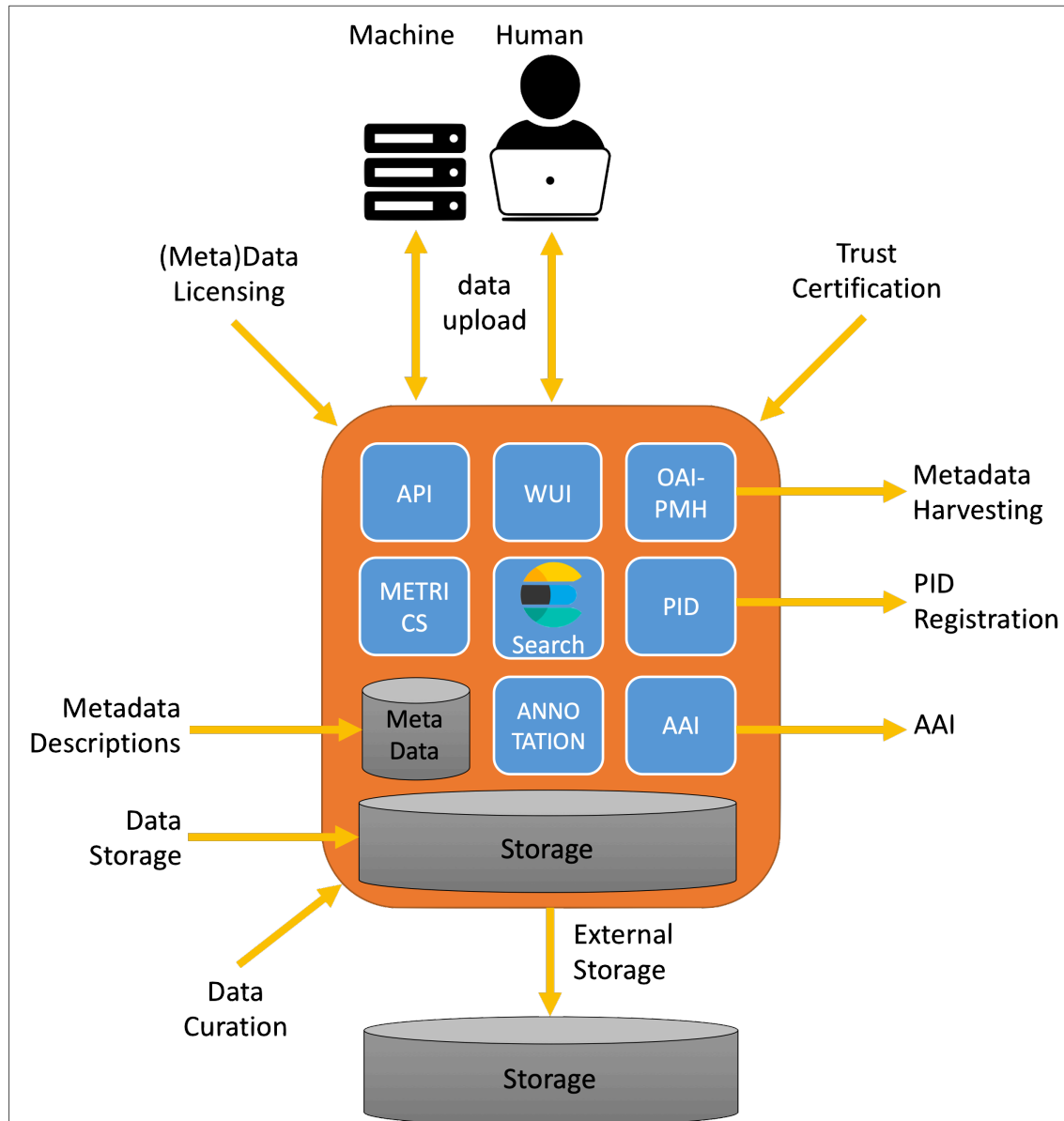
“A digital repository is an infrastructure component that is able to store, manage and curate Digital Objects and return their bitstreams when a request is being issued.”

(RDA Data Foundation and Terminology Working Group)

- kutatási infrastruktúra <> repozitórium
- adatbázis <> repozitórium
- adattárház <> repozitórium



- a tartalom fajtája alapján:
 - full text publikáció, preprint, tézis, disszertáció, technical report, whitepaper, learning object, research object, **adatrekord**, **nyers kutatási adat**, **származtatott kutatási adat**, ...metaadat rekord
- a tartalom származási helye alapján:
 - személyes adat, folyóirat, intézet, részleg, projekt, ágazat, nemzeti, nemzetközi,...
- a tartalom felhasználási célja alapján:
 - hozzáférés biztosítása, keresés megkönnyítése, hosszú távú digitális tárolás, magán/csoportos/intézeti/céges tulajdonlás, megosztás, újrafelhasználás, ...



Korszerű repozitórium (egy példa)

Digital Repository

From: Presentation:
Mark van der Sanden & Heinrich Widmann
2020 EOSC-hub Week 20.05.2020

A vízió

- Évtizedes lemaradásban a fejlett országokhoz képest
- Intézményi repozitóriumok országos hálózata létezik a publikációk számára (MTA: REAL, SZTAKI: ePrints,...)
- Infrastrukturális jelentőségű rendszerek és rendszerkapcsolatok:
 - **MTMT** (adathivatkozás kezelhetősége megoldott)
 - **ELKH Cloud** (2. fázis) üzemel
 - **ORCID** (Open Researcher and Contributor ID) - kutató azonosítás használatban van
 - **DOI** (Digital Object Identifier) azonosítás terjed
- MTA rendelet a nyílt hozzáférésről
- A tudományos adatkezelés **FAIR** követelményei behatolnak
 - Findable, Accessible, Interoperable and Re-usable

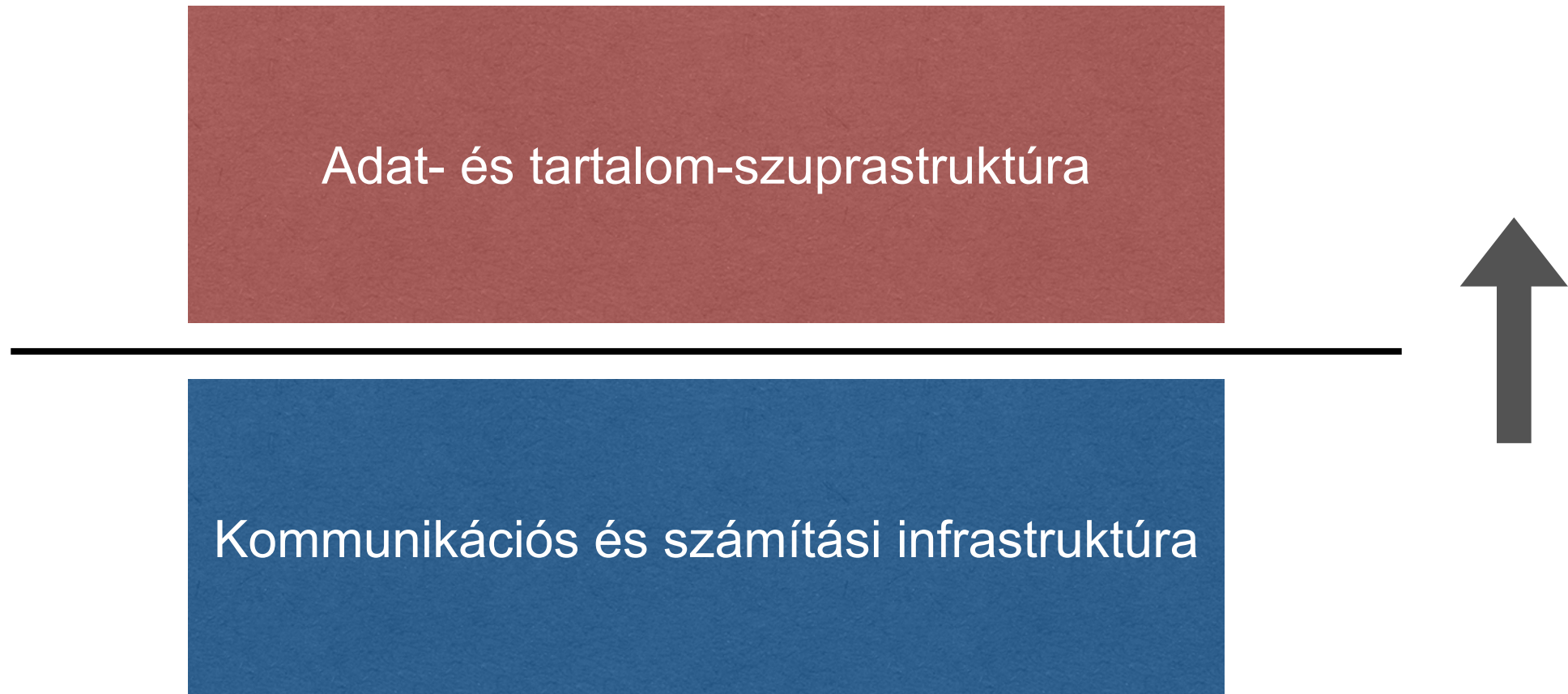
full text repozitórium
(REAL + intézeti repozitóriumok)

adat repozitórium
(néhány példa: CONCORDA, KDK...)

országosan egységes
(generikus)
repozitórium-hálózat



ágazati
(tudományterületenkénti)
repozitórium-hálózat



<u>V.</u>	<u>Nemzetközi szint</u>
<u>IV.</u>	<u>ELKH (országos) szint</u>
<u>III.</u>	<u>Ágazati (országos) szint</u>
<u>II.</u>	<u>Intézményi szint</u>
<u>I.</u>	<u>Projekt és/vagy mérőműszer szint</u>

- eltérő adatkezelési célok
- eltérő adatkezelési funkciók
- eltérő szuprastrukturális komponensek

A kutatási adatrepozitóriumok rendszerét mind a négy magyar kezelési szinten **egységes és interoperábilis** adatkezelési szuprastrukturális szemlélettel szervezzük

<u>V.</u>	<u>Nemzetközi szint</u>
<u>IV.</u>	<u>ELKH (országos) szint</u>
<u>III.</u>	<u>Ágazati (országos) szint</u>
<u>II.</u>	<u>Intézményi szint</u>
<u>I.</u>	<u>Projekt és/vagy mérőműszer szint</u>

- a kutatási adatkezelési szuprastruktúra fejlesztési terv
 - országos szinten:
 - infrastruktúra szervezési elvek
 - logikai architektura
 - adatkezelési, metaadatkezelési szabályzatok
 - a repozitóriumokkal szembeni minőségi és/vagy funkcionális elvárások rendszere
 - ágazatonként:
 - ágazati adatkezelési, metaadatkezelési szabályzatok
 - metaadatséma szabványosítási terv
 - technológia felhasználási ajánlások

- adatszolgáltatások illeszkedő (interoperábilis) rendszere
- infrastrukturális alap-szolgáltatások
 - CONCORDA-2 központi kutatási adatrepozitórium
 - országos közös kereső
 - országos metaadat sémaregiszter
 - ARP portál
- alap-adatok (névterek, tezauruszok,...)
- interoperabilitás
 - szabványos interfészek és kommunikációs protokollok
 - szemantikus interoperabilitás
 - működési interoperabilitás
- szabványosság konformancia központ, konformancia ellenőrzési szolgáltatások
- fenntarthatóság (működési, üzleti, finanszírozási modellek, intézményi működés)

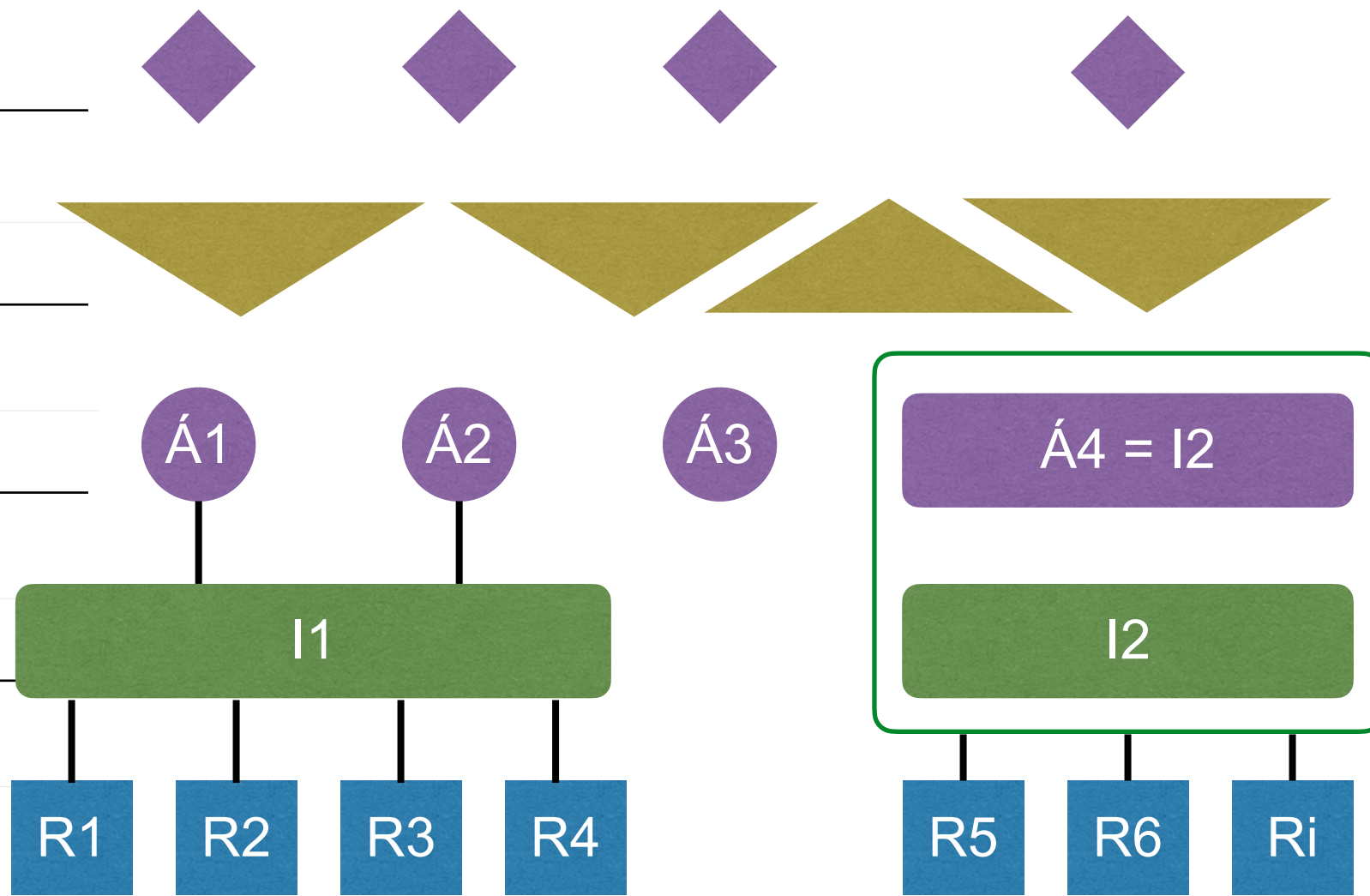
V. Nemzetközi szint

IV. ELKH (országos) szint

III. Ágazati (országos) szint

II. Intézményi szint

I. Projekt és/vagy mérőműszer szint



- Projekt/mérőműszer szint:
 - a kutatási adatok keletkeztetése, adatgyűjtés
 - elsődleges tárolás
 - ARP: 2 db kihelyezett elsődleges (NAS) tároló (pilot)
 - az elsődleges adatfeldolgozás kiszolgálása
 - repozitóriumokhoz és adatfeldolgozási rendszerekhez való hálózati adattranszfer

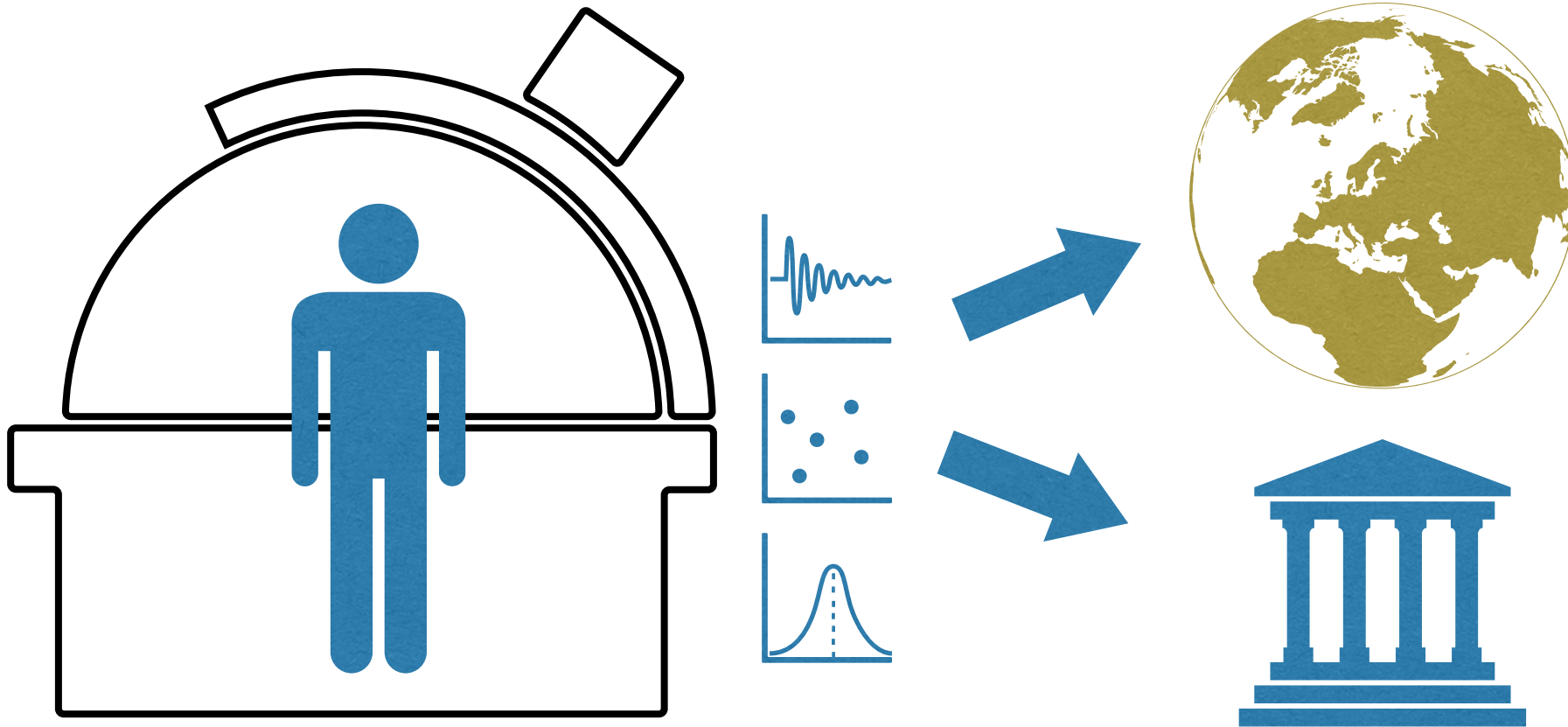
- Intézményi szint:
 - intézmény kutatási adatvagyonának védelme
 - a projektek közötti időintervallumokban történő adattárolás:
 - a jövőbeli projektekben történő adatfelhasználás biztosítása
 - a párhuzamos projektekben történő adatfelhasználás biztosítása
- Intézményi és ágazati szintek potenciális egyesítése
 - Az intézmény az ágazati szintű repozitóriumi feladatokat is elláthatja:
 - vezető intézmény ágazati szinten
 - az intézményi repozitórium fenntartási szolgáltatást (személyzet, adatkuráció, infrastruktúra, ...) biztosít “árva” repozitóriumok számára
 - gazdaságossági/fenntarthatósági megfontolások

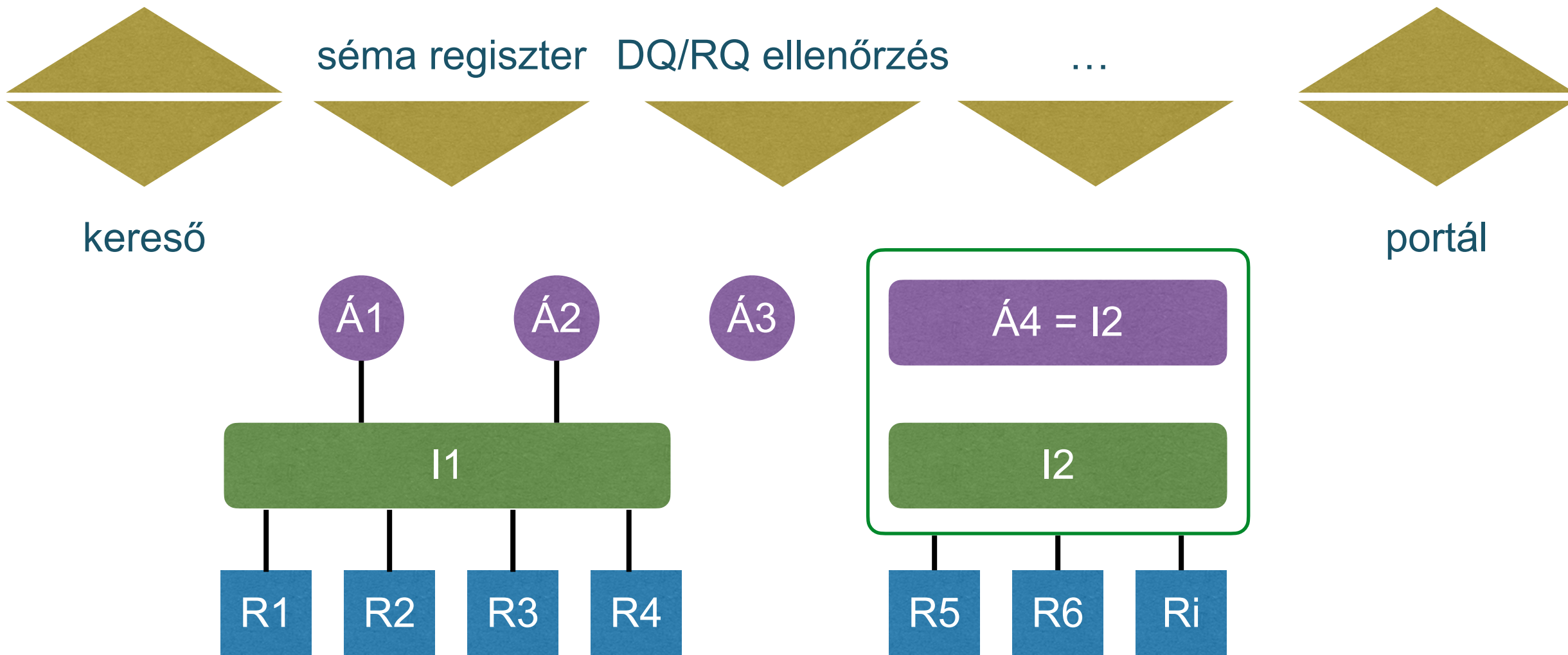
- **Ágazati szint:**
 - a kutatási adatok diszciplináris megosztásának biztosítása
 - országon belül
 - nemzetközileg
 - a nagy, központosított, országos adatfeldolgozó kapacitások (ELKH felhő, KIFÜ felhő, szuperszámítógépek) közvetlen és célszerű elérésének biztosítása
 - az intézményi és ágazati szintű adattárolók (funkcionális, fenntartási, ... szempontokból) egyesíthetők

- **ELKH (országos) szint:**
 - a teljes kutatási intézményhálózat egészére vonatkozó adatkezelési és adatrepozitóriumi központi szolgáltatások biztosítása
 - **1. központi tároló (CONCORDA-2) - hiánypótlás**
 - **2. központi kereső:** a kutatási adatok központi, országosan egyablakos, kereső szolgáltatása
 - **3. metadatséma és alkalmazási profil regiszter**
 - metaadat, interfész, protokoll,... szabványosítási és konformancia ellenőrzési szolgálatok
 - adat- és repozitórium-minőség ellenőrzési szolgálatok
 - hosszú távú digitális megőrzési szolgálat
 - ...
 - a kutatási szféra/repozitóriumi szakmai témakör képviselete
 - országon belül
 - nemzetközi irányokban

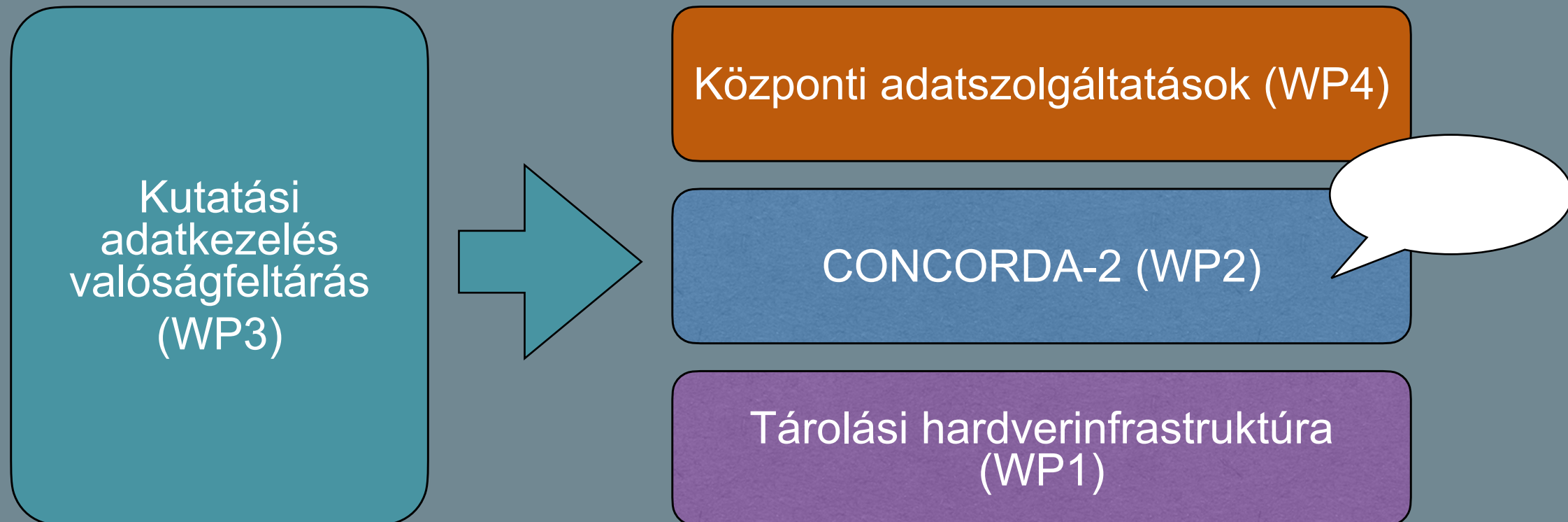
Kettős tárolási logika

(párhuzamos hazai és nemzetközi tárolás a totális hazai adatszuverenitás érdekében)





CONCORDA - 2 központi adatrepozitórium





- Alap: Harvard Dataverse Repository
- 2020. évben:
 - CONCORDA szerver létesítése (v. 4.20)
 - Dataverse kétnyelvűsítés, GUI design csere, hibajavítások
- Pilot projektek (HRDA pályázat adatkezelés és tárolás célú pilot projektek számára, 7 db támogatott projekt)
- Teszt célú CONCORDA használat (4+ HRDA pilot projekt esetében, ...), aktív operátori segítség, adatmigrációs scriptek fejlesztése, hibajavítás
- Tanulságok levonása (HRDA workshop keretében, ...)

Search User Guide Support English Sign Up Log In

 **CONCORDA**
Concentrated Cooperation on Research Data

Metrics 477 Downloads Contact Share

The root dataverse for CONCORDA.

Search this dataverse... Advanced Search

Dataverses (30)
 Datasets (1,181)
 Files (2,228)

Dataverse Category
Organization or Institution (12)
Research Project (7)
Research Group (5)
Department (3)
Laboratory (1)

Publication Year
2022 (658)
2021 (519)
2020 (34)

Author Name
HF (1,149)
Data, GLD (3)
Fiáth, Richárd (3)
Nagy, Dénes Lajos (3)
Hegedűs, Gergő (2)
More...

Subject

1 to 10 of 1,211 Results Sort

A Three-Dimensional Analysis of Magnetic Nanopattern Formation in FeRh Thin Films on MgO Substrates: Implications for Spintronic Devices
Apr 6, 2022 - Wigner Research Centre for Physics

 Merkel, Dániel Géza; Hegedűs, Gergő; Gracheva, Maria; Deák, András; Illés, Levente; Németh Attila; Maccari, Fernando; Radulov, Iliya; Major, Márton; Chumakov, Aleksandr I.; Bessas, Dimitros; Nagy, Dénes Lajos; Zolnai, Zsolt; Graning, Sára; Sájerman, Klára; Szilágyi, Edit; Lengyel, Attila, 2022, "A Three-Dimensional Analysis of Magnetic Nanopattern Formation in FeRh Thin Films on MgO Substrates: Implications for Spintronic Devices", <https://doi.org/10.5072/FK2/HVYQSA>, CONCORDA, V2, UNF:6:bgzs99u5U4GqdM9ySXKVTg== [fileUNF]

Magnetic nanopatterns were successfully created in FeRh thin film deposited on MgO (100) substrates. Silica and polystyrene spherical masks, nominally 500 and 1000 nm in diameter, respectively were applied on the surface of the sample in order to locally shadow the film against t...

Reversible control of magnetism in FeRh thin films
Apr 5, 2022 - Wigner Research Centre for Physics

 Merkel, Dániel G.; Lengyel, Attila; Nagy, Dénes Lajos; Németh, Attila; Horváth, Zsolt E.; Bogdán, Csilla; Gracheva, Maria A.; Hegedűs, Gergő; Sajti, Szilárd; Radnóczy Z. György; Szilágyi Edit, 2021, "Reversible control of magnetism in FeRh thin films", <https://doi.org/10.5072/FK2/ZC2AK2>, CONCORDA, V2, UNF:6:Cvr6lrS6zc+AXoY5zRPfYA== [fileUNF]

The multilayer of approximate structure MgO(100)/[nFe51Rh49(63 Å)/57Fe51Rh49(46 Å)]10 deposited at 200 °C is primarily of paramagnetic A1 phase and is fully converted to the magnetic B2 phase by annealing at 300 °C for 60 min. Subsequent irradiation by 120 keV Ne+ ions turns the...

- a tárolt objektumok egyértelmű és perzisztens azonosítása a weben (DOI, handle, ARK,...)
- felhasználó azonosítás (egyéniesített szolgáltatások céljára)
- az objektumok lehetséges felhasználásainak egyértelmű szabályozása (licenszek)
- objektumokhoz extra tartalom (overlay content) társítása (annotáció, review, comment,...)
- fenntartói - felhasználói kommunikáció
 - szolgáltatási minőség fejlesztése, új igények kiszolgálása

- az objektumok exportja, importja (interfészek és protokollok)
- objektumok nagy tömegű exportja, importja
- a repozitórium aktivitások mentése, loggolása, monitorozása
- az objektum transzformációk és a belső repozitóriumi aktivitások külső transzparenciájának biztosítása

- a repozitórium működésének leírására szabványosított leíró metaadatok (metrikák) bevezetése
 - hálózati és más külső szolgáltatások összehangolása
 - külső szervezetek számára meta-, statisztikai adat szolgáltatások
 - működés optimalizáció
- a tárolt tartalom fajták ismerete (self-awareness) és specifikus kezelése (TEI XML, ...)
- összetett (kompozit) objektumok kezelése
 - research object (RO), tárolási objektum, hálózatos objektum modellel rendelkező objektumok (LD, ...)

- repozitórium beágyazhatósági, interoperabilitási funkciók:
 - kommunikációs interfészek, API, szabványos protokollok (adattranzferszintű interoperabilitás)
 - szemantikus szintű adateleírások (szemantikus interoperabilitás)
 - repozitórium működési transzparencia (működési interoperabilitás)
- repozitórium stabil fenntartása hosszú távon
 - működési, üzleti, finanszírozási modellek
- hosszú távú digitális tárolás? (csak a középtáv a megcélzott)

CONCORDA - 2 implementáció

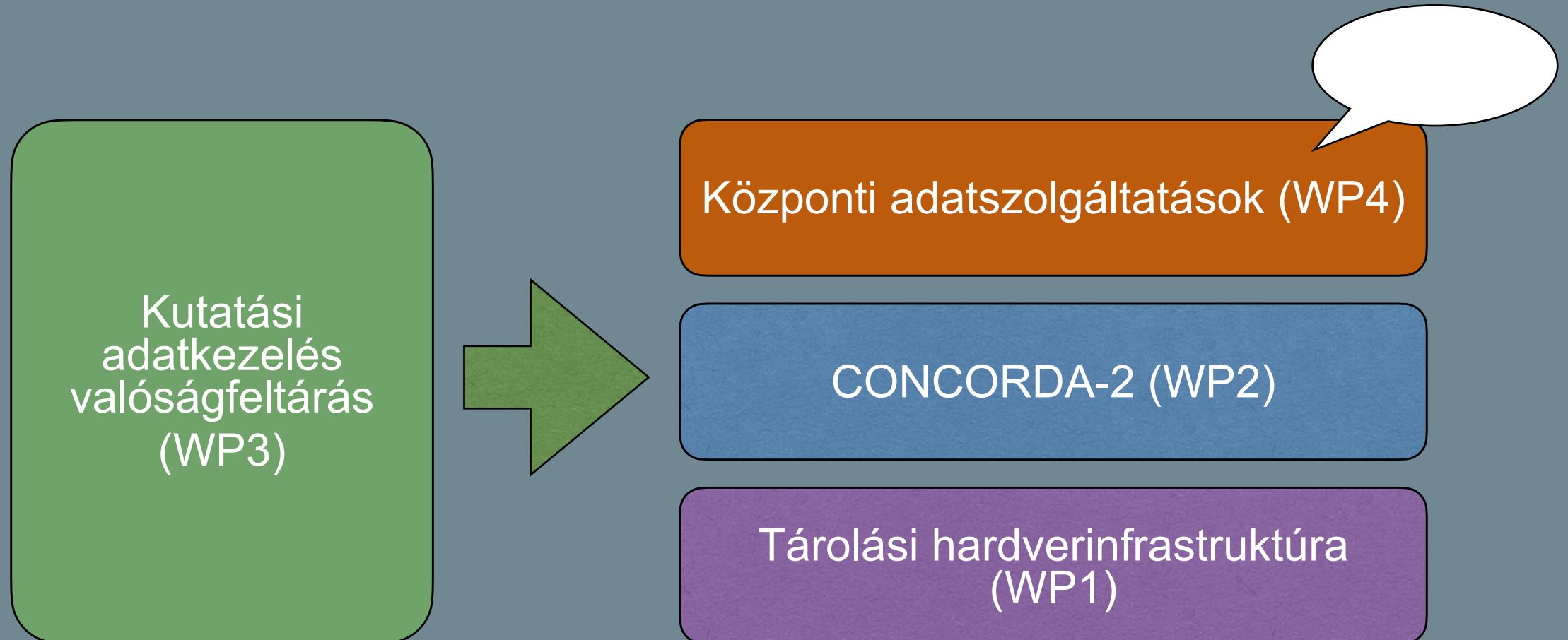
- A CONCORDA-2 szolgáltatás és a szolgáltatási szoftver környezet kialakítása
 - Szolgáltatási szerver környezet kialakítása
 - ◆ szerver cluster, egyéb szerverek,
 - Adattárolási háttér illesztése
 - ◆ georedundancia, biztonságos tárolás, backup, storage alrendszer illesztése ...
 - Szolgáltatás monitorozásának kialakítása
 - Applikációs szintű monitoring dashboard fejlesztése
 - ◆ belső működés, erőforrás használat, adatáramlás, hiba identifikáció,...
 - ◆ monitoring funkciók, monitoring GUI, szoftver szenzorok
 - Tesztelési szoftver környezetek kialakítása a szoftverek továbbfejlesztése, a verziók váltása, módosítása elősegítésére

- Adatkapcsolatok kialakítása (API és kommunikációs protokollok)
 - Metaadat sémaregiszter API
 - API és kereső protokoll fejlesztés a közös kereső integrációjához
 - Adatkapcsolati szolgáltató környezet létrehozása (különbéle célú külső szerverek és azok közötti adatkapcsolatok létesítése, adatkapcsolati workflow-k, adatbevitel, ...)
 - Az 1. és a 2. szintű tároló rendszerek közötti adatkapcsolati integráció
 - nagytömegű adatbevitel támogatása
 - OAI-PMH alapú interfészelés
 - MTMT interfészelés

- Funkcionális fejlesztések és hibajavítások
 - handle és ARK (Archival Resource Key) mint elsődleges permanens azonosítási rendszerek integrálása
 - DOI, mint másodlagos permanens azonosító és kezelési workflow kialakítása
 - Diszciplína specifikus automatikus sémakezelés (sémadefiníció import, értelmezés, a belső CONCORDA keresési és megjelenítési funkciók továbbfejlesztése)
 - Hibajavítások
 - Extra funkciók fejlesztése, meglévő funkciók módosítása (WP3 követelményanalízise alapján)

- Rendszerintegrációk
 - Metaadat sémaregiszter integráció
 - Közös kereső integráció
 - ARP Portál integráció
 - MTMT integráció
 - ELKH Cloud integráció
 - OAI-PMH alapú repozitórium integrációk
 - Dataverse szoftver verzió-interoperabilitási interfész és verzió váltási módszertan kifejlesztése
 - A szoftver rendszer egyéb komponensei verziókövetésének és karbantartásának menedzsmentje

Központi ARP szolgáltatások



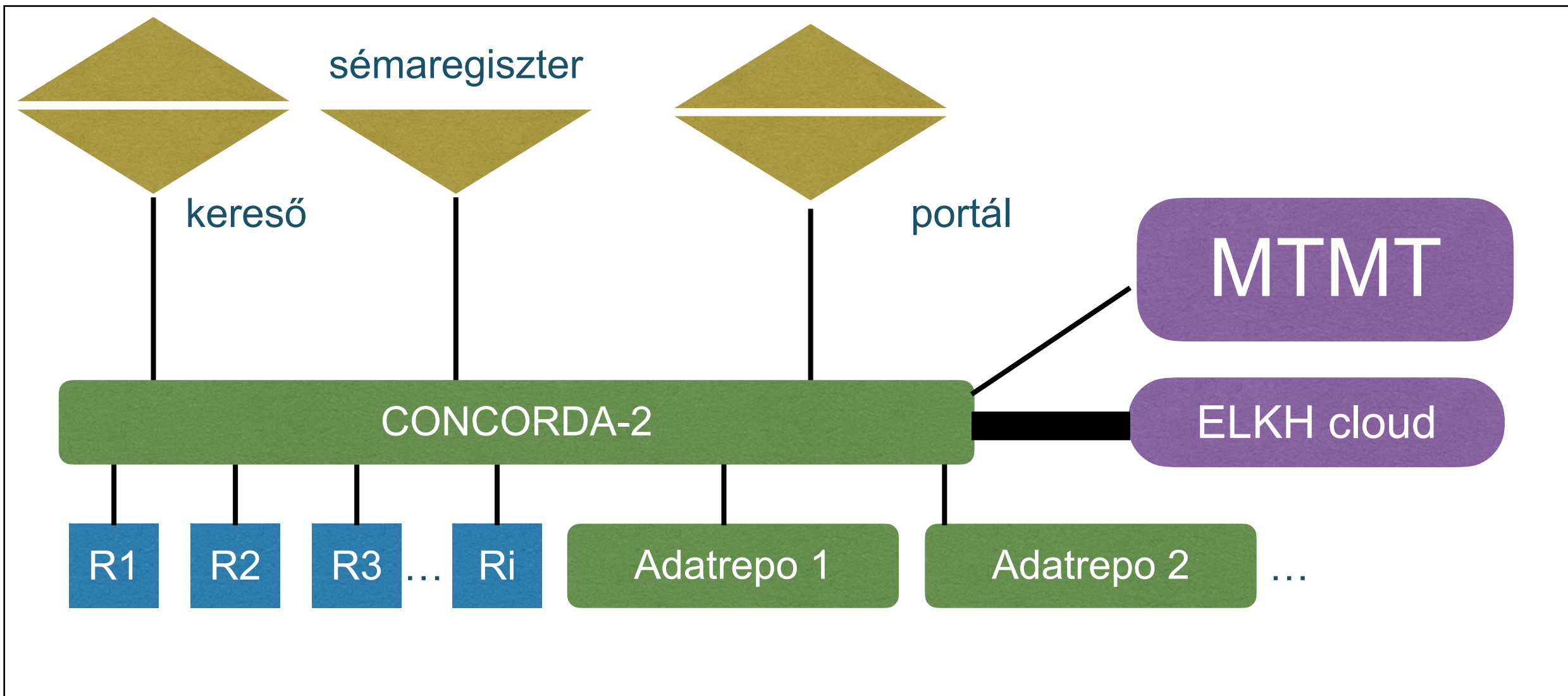
■ Közös Kereső

- Országos átfogású közös kutatási adatkereső
- Architektúra:
 - OAI-PMH harvesting / virtuális objektum mapping a CONCORDA-ba (teljes séma alapon)
 - A CONCORDA belső tárolási/keresési mechanizmusának továbbfejlesztése

■ Metaadat sémaregiszter

- séma definíciók (logikai séma + kódolási séma + applikációs profil) tárolása
- sémaverziók interaktív kezelése
- séma definíciók létrehozása csoportos kollaboráció támogatással (osztott sémaeditálás, kommentelés, csoportos döntéstámogatás)
- sémadefiníciók migrálása a CONCORDA-2-be
- sémadefiníció import-export külső regiszterekből (schema.org), interfészelés, formátum transzformációk, szemantikai gazdagítás

- Központi ARP portál
 - projektinformációk
 - felhasználói kommunikáció
 - segédprogramok, sciptek, szabványok, szabályzatok...
- Adatvizualizáció (ARP 2. fázis)
 - Egyedi, sémaspecifikus adatmegjelenítés, böngészés, adatanalízis célokra
 - interfészelés az adatanalízis és egyéb adatfeldolgozó rendszerek irányában
- Adattranszformációk (ARP 2. fázis)
 - integrált sémaspecifikus adatkonverziós rutinkönyvtár



ELKH ARP

A CONCORDA - 2 kutatási adatrepozitórium rendszer

Kovács László
SZTAKI DSD
laszlo.kovacs@sztaki.hu