

AZ ELKH ADATREPOZITÓRIUM PLATFORM PROJEKT A KUTATÁSI ADATTÁROLÁS ÉS MEGOSZTÁS ORSZÁGOS FÖDERÁLT RENDSZERE

THE ELKH ARP RESEARCH DATA REPOSITORY PROJECT, NATIONAL FEDERATED REPOSITORY NETWORK FOR SHARING RESEARCH DATA

Kovács László

műszaki doktor, osztályvezető

ELKH Számítástechnikai és Automatizálási Kutatóintézet Elosztott Rendszerek Osztály, Budapest

laszlo.kovacs@sztaki.hu

ÖSSZEFOGLALÁS

Az Eötvös Loránd Kutatási Hálózat Adatrepozitórium Platform projekt országos kiszolgálást nyújtó központi rendszert hoz létre, mely többek között az ELKH intézeteit szolgálja ki kutatási adatrepozitóriumi és vele szorosan integrált adat- és metaadat-kezelési szolgáltatásokkal.

Az ELKH ARP-projekt alapfeladata, hogy megtervezze és megteremtse az országos föderált kutatási adatrepozitórium-hálózat központi rendszerkomponenseit, hálózati szolgáltatásait, amelyek segítségével egységes országos hálózati rendszerbe szervezhetők az ELKH meglévő és leendő kutatási adatrepozitóriumai.

A rendszer a központban biztosít kutatási adatrepozitórium (tárolási) szolgáltatást is azon intézmények és/vagy tudományágak, kutatócsoportok számára, amelyek nem tudnak, vagy nem akarnak saját kutatási adatrepozitóriumot felállítani vagy üzemeltetni.

A projekt kitűzött céljait, a létrehozandó rendszer átfogó logikai architektúráját, a rendszerkomponensek funkcióit, együttműködési mechanizmusait mutatom most be, valamint azt, hogy a rendszer milyen módon segíti elő a magyar kutatási adatkezelés FAIR-elveknek megfelelő szervezését.

Az ELKH ARP-rendszer támogatni kívánja a kutatási adatkezelés új, FAIR-érzékeny szintjét lehetővé tévő kutatási objektum (Research Object, RO) magyarországi bevezetését és széles körű alkalmazását. A kutatási objektumok szoftveres megvalósulásának általunk használt formája az RO-Crate. Az RO-Crate kutatási objektumkezelési infrastruktúra szervesen illeszkedik a föderált kutatási adatrepozitórium-hálózat tárolási és egyéb szolgáltatásaihoz.

Az ELKH ARP-repozitóriumhálózat föderációs mechanizmusai, integrált adatkezelési módok (protokollok és policyk) hosszú távon biztosítják az országos rendszer további fejlődése során a komponensek közötti harmonikus együttműködést adatkapcsolati, szemantikus, szervezeti szervezési és működtetési szinteken.

ABSTRACT

ELKH ARP Research Data Repository project aims to establish a central data repository system supporting the whole ELKH research network, providing research data repository and related, integrated data- and metadata-management services.

The mission of the ARP project is twofold: it designs and establishes the central system components and the network services of a national federated research data repository network and on the other hand creates the federated mechanisms and system components that enable to organize the existing and future research data repositories of ELKH into a single, uniform national network system.

The system will also centrally provide research data repository services to the institutions and/or disciplines, research teams that cannot or will not establish and manage their own research data repository, filling an important gap by doing so.

In the article we introduce the goals set by the project, the logical architecture of the system to be built, the functions and interoperability mechanisms of the system components, and the way the system facilitates the implementation of the FAIR principles in the process of managing national research data.

The ELKH ARP system will support the introduction and widespread use of RO – Research Object, that enables the new, FAIR-sensitive level of research data management in Hungary. To achieve this, RO-Crate, the actual software implementation of the concept of RO, is fully supported within the system. The RO infrastructure that is to meet the requirements to apply RO-Crate technology in Hungary forms an integral part of the federated research data repository network (storage and other) services.

The ELKH ARP repository network federated mechanisms, integrated data management methods (protocols and policies) will ensure, in the course of further development of the system, the interoperability between the components on data transfer, semantic, institutional, and operational levels in the long term.

Kulcsszavak: kutatási adat, repozitórium, FAIR, kutatási objektum, RO-Crate

Keywords: research data, repository, FAIR, research object, RO-Crate

Az Eötvös Loránd Kutatási Hálózat Adarepozitórium Platform (ELKH ARP) projekt az intézményhálózat kutatási adatkezelését támogató, új hálózati szolgáltatásrendszer alapjainak megteremtését célozza. A projekt a kutatási adatok kezelésének alapvető feltételeit, az adattárolás és adatmegosztás országos szervezeti, humán, és műszaki infrastruktúráját hozza létre.

Lehetővé válik a kutatási adatok mint értékkel bíró adatvagyon biztonságos digitális tárolása, azok használata, újrafelhasználása, másodlagos és többedleges hasznosítása a kutatási és egyéb, leginkább a gazdasági szféra számára. A kutatási adatok értékének tudatosítása, az értéknek megfelelő adekvát adatkezelési és tárolási erőforrás használat fontos és felvállalt küldetése a projektnek.

A kutatási adatok tárolásának, kezelésének fejlesztése, a szükséges *know-how* létrehozása és a tervezett szakmai képzés megteremti a ELKH-ban keletkező ku-

tatási adatok jobb és hatékonyabb felhasználhatóságát. A hazai kutatási adat-repozitóriumok létrehozása és az ehhez szükséges fejlesztési és felhasználói kompetenciák kialakítása javíthatja a kutatók (európai) K+F-pályázatokban várható sikerességét.

Ezen belül is a legfontosabb a kiépülőben lévő EOSC (European Open Science Cloud) páneurópai kutatási infrastruktúrába való bekapcsolódásunk, amelyre a tervezett ELKH ARP-projekt megvalósítása nyújt lehetőséget.

1. A VÍZIÓ: FÖDERÁLT KUTATÁSI ADATREPOZITÓRIUM-HÁLÓZAT

Az ARP-projekt keretében az ELKH kutatási adat-repozitóriumait egységes hálózati rendszerbe kívánjuk fűzni. Ez az egységes rendszer az ARP-szoftverplatform által kiszolgált föderált kutatási adat-repozitórium-hálózat. Cikkemben a víziónról és annak egyes megvalósítási lépéseiről írok.

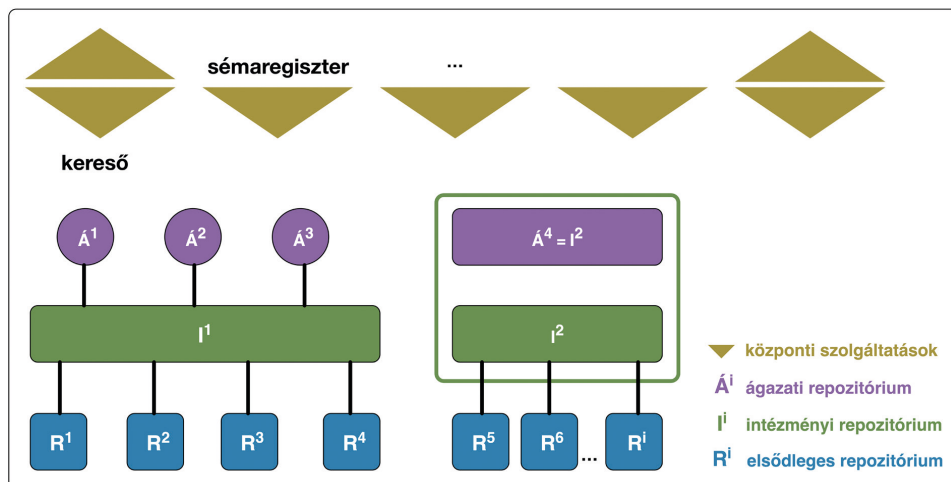
A kutatási adat-repozitóriumok hálózatának kialakítása mellett, azok hazai alkalmazástechnológiáinak továbbfejlesztését elsősorban az alábbi részterületeken célozzuk meg:

- a repozitóriumok, a repozitóriumi funkcionálisok kutatási munkafolyamatokba (workflow) történő integrálása;
- a kutatási munkafolyamat során menet közben kinyerhető, a folyamatot és a felhasznált entitásokat, erőforrásokat jellemző metaadatok rögzítése, kezelése, felhasználása, megosztása;
- az összetett kutatási objektumok (Research Objects, RO), laboratóriumi riportok, memók stb. létrehozása, megszerzése, kezelése, tárolása és felhasználása, a kutatások reprodukálhatóságának elősegítése érdekében, mind a kutatási adatok, mind pedig az alkalmazott tudományos módszerek és eljárások tekintetében.

A kutatásba fektetett anyagi források, eszközök veszteségének hosszabb távú minimalizálását, egyben nagyobb fokú hasznosítását, illetve hasznosulását célozzuk meg.

1.1. Föderált repozitórium-hálózat – sematikus architektúra

A hálózatba kötött, összekapcsolt kutatási adat-repozitóriumok hierarchiát alkotnak, melynek legalsó részén az elsődleges adattárolók (R^1 - R^i , primary data repository) helyezkednek el. Az elsődleges tárolók magasabb szintű (intézményi [I^1], ágazati [A^1]) tárolókkal állnak adat- és egyéb pl. szervezeti, szervezési kapcsolatban. Bizonyos (nagy) intézmények egyben ágazati kiszolgálást célzó repozitóriumokat is üzemeltetnek. Ilyenkor az intézményi és ágazati repozitóriumok egybeeshetnek (pl. $I^2 = A^4$).



1. ábra. Az országos ELKH ARP föderált repozitórium-hálózat sematikus architektúrája
(a szerző saját szerkesztése)

A kutatási adatrepozitóriumok egyike a kitüntetett, és központilag létrehozott ARP-adatrepozitórium, amely ágazatfüggetlen, országos központi kiszolgálást nyújt (hiánypótlásként).

Az összekapcsolt tárolók közötti adatkapcsolat szervezett, vezérelt és kézben tartott adatátviteli lehetőség, amely egyaránt vonatkozik a tárolt adatokhoz társított metaadatokra és a tárolt adatokra magukra. A rendszerbe kapcsolt repozitóriumok föderációja az egységesen megszervezett adatkapcsolati hálózaton alapul, de a föderációt magasabb (szemantikus és működési interoperabilitási) szinteken működő egyéb föderációs mechanizmusok is támogatják.

A föderált rendszerarchitektúra legfelső rétegében helyezkednek el az ARP-projekt központi szolgáltatásai (közös kereső, sémaregiszter, ontológiaregiszter stb. – az architektúrávázlaton háromszögekkel jelölve). A háromszögek irányultsága mutatja, hogy a szolgáltatás elsősorban milyen irányokba (hálózaton belül, hálózaton kívül) szolgálhat.

2. ELKH ARP-PLATFORM

Az ELKH ARP-projekt keretében fejlesztjük ki az ARP-(szoftver)platformot. Az ARP-platform logikai architektúráját több forrásból (kutatói személyes interjúk, online kérdőív, szakértői háttértudás és a projekt keretében elvégzett szakmai elemzések, automatikus szoftverevaluációk stb.) származó elvárások, követelmények alapján terveztük meg, amelynek során figyelembe vettük a digitális

könyvtári, repozitóriumi szakterület és a tudományos kommunikáció (scholarly communication) aktuális fejlődési trendjeit is.

A logikai architektúrában az ARP-platform részrendszereinek azonosítását és a részrendszer-funkcionalitások meghatározását rögzítettük. A részrendszerek közötti fontosabb adat- és vezérlésáramlások meghatározásával a rendszer integráltságának milyenségét határoztuk meg.

Az ARP-szoftverplatform lazán (adat- és vezérléskapcsolatokkal) csatolt rendszerek integrációjával jön létre. A komponensek között preferáltan szabványos adatformátumokat használó adatátvitteleket terveztünk. Az ARP saját, specifikus adatformátumainak, illetve protokollkapcsolatainak mennyisége minimalizált.

Az ARP-platform bővíthetősége, új részrendszerek könnyű illeszthetősége (később, a projekt további fázisaiban) fontos tervezési szempont volt, pontosan ezért választottuk a lazán csatolt komponensekből történő rendszerépítkezést.

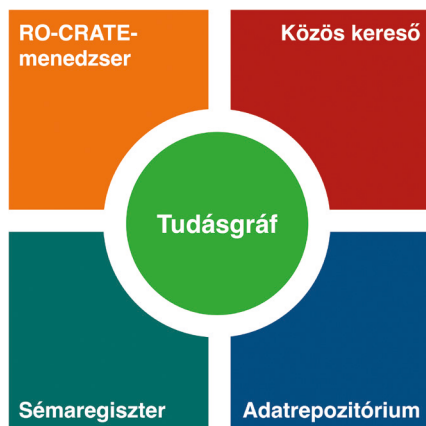
2.1. ELKH ARP-platform – főbb rendszerkomponensek

Az ARP-platform meghatározó komponensei a következők:

1. Az ARP-adatrepozitórium a Harvard Dataverse-kódbázison alapuló, az eredeti Dataverse Repozitórium rendszer funkcionális bővítése révén létrejövő kutatási adatrepozitórium, amely központi repozitálási (tárolási) szolgáltatásokat nyújt.
2. Metaadatséma-regiszter, az országos kutatási adatrepozitórium-hálózatban használt metaadatsémák egyértelmű definícióját kezelő, létrehozó, tároló és azokat szolgáltató rendszerkomponens.
3. A föderált repozitóriumhálózat központi keresője. A hálózati rendszerben tárolt kutatási adatok megtalálását teszi lehetővé, bárhol is legyenek az adatok eltárolva a hálózatban.
4. Az RO-Crate¹-menedzser az RO-Crate-csomagolás formátumú kutatási objektumkezelő, a kutatási adattermékek FAIR²-elvárásoknak megfelelő objektumorientált kezelését, tárolását és felhasználását segíti elő adatcsomag- és adatfájl-szintű szolgáltatásaival. (Csomagolási, metaadatolási, formátum-ellenőrzési, objektumkivonatoló adatkezelési stb. szolgáltatások.)
5. Tudásgráf. Fenti rendszerkomponenseket a központi tudásgráf mint a komponensek között megosztott adatszerkezet kapcsolja össze, és teszi lehetővé a komponensek közötti belső kommunikációt és a rendszeren belüli szeman- tikusan helyes adatmegosztást.

¹ RO-Crate – Research Object Crate, adatsomagolási formátum

² FAIR – Findable, Accessible, Interoperable, Reusable



2. ábra. Az ELKH ARP-platform főbb központi rendszerkomponensei
(a szerző saját szerkesztése)

Az ARP-platformban a központi tudásgráf mint a föderált rendszerben tárolt adat-entitásokhoz társított metaadatok központi tárolója teszi lehetővé a közös keresés implementálását, ami részletes és fázettázható keresést nyújt a teljes hálózatban, különféle granulációs szinteken: a DC- (Dublin Core) metaadatséma-alapú közös globális kereséstől, a diszciplínaszpecifikus sémavezérelt keresésen keresztül egészen a fájl szintű keresésig.

A rendszerben Harvard Dataverse-szoftverrendszeren alapuló tárolási, adat-repozitóriumi központi szolgáltatást terveztünk. A gazdag metaadatolást metaadatséma-regiszter és a hozzá kapcsolódó ontológiaregiszter támogatja. E két regiszterre dekomponált részrendszer lehetővé teszi a nemzetközileg elterjedt diszciplínaszpecifikus metaadatsémák hazai alkalmazását a föderált repozitóriumphálózat minden kapcsolódó repozitóriumban és egyben a hazai-intézményi-ágazati specialítások mint sémaalkalmazási profilok (application profile) létrehozását és gondozását is.

Az ARP-platform minden komponense támogatja a hazai bevezetésre és használatának elterjesztésére kiválasztott RO-Crate kutatási objektum adatsomagolási implementációs technológiát, mely könnyen használható, a járatos szemantikus webtechnológiák, illetve ajánlások, szabványok használatával hoz létre kutatási objektum adatsomagokat, egyben megfelel a jelenleg kialakítás alatt álló, az EOSC keretében is értelmezett FAIR Digital Object (FDO) magas szintű koncepcionális modell elvárásainak.

Az RO-Crate csomagok szerializált mini tudásgráfokban ábrázolják az adatfájlokhoz társított metaadat-szerkezeteket. Ezeket az ARP minden általa kezelt RO-Crate-adatsomaggól kivonja, és azokat az ARP központi keresőjének tudásgrádjába integrálja. Ezáltal megteremtődik az alacsony granulációs szintű

adatkereshetőség és a diszciplináris szintű specifikus keresési szolgáltatás központi támogatása. Az interdiszciplináris adatfelhasználás segítése mellett tehát az ARP-rendszer a lehető legerőteljesebben kívánja támogatni a mindennapi, diszciplínaszpecifikus adatfelhasználást, megkönnyítve a kutatók saját adatállományainak saját használatát is.

A kutatási objektumok és azok RO-Crate-alapú csomagolása nem csupán a kutatási adatok mint adatentitások kezelését és gazdag metaadatolását segíti, de hosszabb távon alkalmas a teljes kutatási tevékenység archiválására is.

A kutatási tevékenységek, a humán résztvevők azonosításával, a kutatási munkafolyamatok leírásával (workflow), eredetinformációkkal stb. gazdagított metaadatolása, a FAIR adatkezelési követelmények jövőbeli várható fejlődési fázisaiban, a kutatási tevékenység reprodukálhatósága irányában lesz szükséges és elvárt alapkövetelmény akár már intézményi szinten is. Várhatóan újabb, hatékonyabb szoftvereszközök jelennek meg az RO-Crate-alapú kutatási objektum kezelésre, ami által a kutatók számára egyszerűbb lesz a kutatási adat- és metaadat-kezelés, valamint a teljes kutatási folyamat hosszú távú archiválása, reprodukálása és/vagy megosztása. Mindez a kutatásba fektetett erőforrások hatékonyabb és takarékosabb felhasználását alapozza meg.

Az ARP-platform biztos jövőállóságát teremthetjük meg a kutatási objektum RO-Crate-alapú technológiájának hazai bevezetésével és teljes körű országos támogatásával.

2.2. ARP-szolgáltatások

Az ARP-platformon központi hálózati szolgáltatásokat hozunk létre, melyek közül számos a hálózati föderáció megteremtését szolgálja. Ezek közül elsődleges a közös kereső szolgáltatás.

2.2.1. Közös keresés központi szolgáltatás

Az ARP-projekt központi kutatási adatkeresőt hoz létre, amely nem csupán a központi Adatrepozitóriumban elhelyezett adatcsomagokat teszi kereshetővé, de a föderált hálózat minden bekapcsolt repozitóriumban, minden hálózati csomópontban is lehetővé teszi a keresést. A központi kereső különféle adatrészletezettségi (granulációs) szinteken működik.

A föderált hálózatba kötött repozitóriumokban elhelyezett adatcsomagok egyedi metaadatsémáit közös DC- (Dublin Core) sémára képezzük le, ezáltal megteremtjük a teljes hálózatban a diszciplínafüggetlen (egyszerű) keresési lehetőségét.

A rendszer egy vagy több diszciplínaszpecifikus metaadatséma kiválasztása után az ezekkel leírt adatobjektumok közötti teljes sémaalapú keresést biztosít a hálózati csomópontok mindegyikében.

A harmadik, legalacsonyabb granulációs szinten a kereső képes az adatsomagok belsejében elhelyezett (fájlszintű) metaadatokban is keresni a korábban részletezett RO-Crate-formátumú adatsomagok ARP-rendszerben történő belső kezelése révén.

A közös kereshetőség implementációs alapja az, hogy a központi kereső begyűjti a hálózatba kötött repozitóriumokból az ott tárolt adatentitásokhoz társított metaadatrekordokat, és egy központi tárolóban helyezi el azokat. Ez a begyűjtési folyamat a metaadat-aratás. Aratáskor a metaadatrekordokat változatlan formában gyűjtjük be. A központban a szemantikus szintű feldolgozás után a metaadatrekordok közösített adattárolóba (központi tudásgráf) kerülnek. Ezt a központi metaadat-tárolót használja a központi kereső és a többi központi rendszerkomponens is. Az aratási folyamatok ütemezése az aratandó repozitóriumok, adatállományok, metaadatfajták stb. kiválasztásával előre tervezhető, konfigurálható, a folyamat automatizált.

Az ARP-projekt tehát egy országos, egyablakos keresőszolgáltatást hoz létre az eltérő granulációs szinteken az összes, a hálózatba bekötött repozitórium esetére.

2.2.2. Föderáció a metaadatolás tekintetében

A kutatási adatok (újra)felhasználásához, archiválásához, az adatkezelő és feldolgozó rendszerek közötti interoperabilitás megteremtéséhez, az adatfeldolgozások gazdaságos megszervezéséhez stb. szükséges a metaadatok sémáját, legalább ágazati szinteken, szabványosítani.

A szabványosítás elektronikus támogatása érdekében az ARP-projekt létrehozta a kutatási metaadatsémák és adatformátumok digitális regiszterét (metaadatséma-regiszter), amely nem csupán a szabványosított metaadatsémák nyíltan hozzáférhető (open access) tárolója, hanem támogatja a szabványosítás kialakításának kooperatív szakértői tevékenységét is.

A metaadatsémák és adatformátumok szabványainak kialakítása esetén elsődleges követelmény a járatos, már (kvázi) szabványosított nemzetközi szabványok és ajánlások hazai adaptálása a hazai követelményekhez.

A metaadatséma, adatformátum szabványosításokhoz központilag menedzselte kutatási részterületi ontológiák, névterek, szótárak, szöszedetek szabványosítását és közös, interoperábilis használatát is elősegítjük a metaadatséma és az adatformátum megosztások mellett. A központi metaadatséma-regiszter az ezekről az objektumokról szóló csoportos tudásfelhalmozást és tudáskezelést is támogatja. A metaadatséma-regiszter fontos társszolgáltatása a metaadatsémák egyértelmű dokumentálása a különféle szintű felhasználói csoportok (teljes vagy részdíszcipína, intézet, kutatócsoport, egyéni felhasználók) számára. Továbbá, újrafelhasználható, szabványos metaadatséma-leíró űrlap-definíciókat szolgáltat a potenciális gépi felhasználások számára.

Ez lehetővé teszi azt, hogy az országos rendszerbe kötött különféle repozitóriumokban azonos célra azonos vagy közel azonos (egyediesített) metaadatsémákat,

alkalmazási profilokat használnak az adatentitások metaadatolására. Ez a közös, (kvázi) szabványosított, permanens azonosítóval ellátott metaadatséma-használat biztosítja azt, hogy a teljes repozitóriumhálózatban legalább diszciplínánként kialakulhasson az adatentitás felhasználások átjárhatósága, az adatentitások megtalálhatósága és azok kompatibilis/azonos szemantikus értelmezése.

A metaadatséma-regiszter a tárolás mellett a séma kialakításához szükséges alapvető szerkesztési funkciókat is szolgáltatja, amelyekkel új sémák definiálhatók, avagy a konkrét felhasználási módokat (application profile, egyediesített séma) is meghatározhatóvá teszi.

A metaadatolás tekintetében tehát ez a központi metaadatséma-regiszter teszi lehetővé a szemantikus szintű hálózati föderáció létrejöttét. Ez azonban csak egy lehetőség, mert a központi metaadatséma-regiszter nem kényszeríti ki a regiszterben elhelyezett metaadatsémák kötelező használatát, ugyanakkor természetesen a föderált közös sémahasználatot a rendszer preferáltan támogatja.

2.2.3. Föderáció adatkapcsolati szinten

A föderált hálózatba kötött kutatási adatrepozitóriumok között a rendszer biztosítja az adatentitások, metaadat-entitások alkalomszerű (ad hoc) és szervezett (előre tervezett és ütemezett) mozgatását egyedi operátori tevékenység révén, avagy előre programozott módokon.

A rendszer az adatmozgatásokat OAI-PMH³- és OAI-ORE⁴-protokollok segítségével végzi. Az adatmozgatás egyedi rekord szinten és/vagy nagyobb adatmennyiségeket egyben mozgatva is lehetséges, beleértve teljes, repozitóriumszintű adatmigrációs feladatokat is. Az adatmozgatás időbeli ismétlődése és periodicitása révén adatfolyamok jönnek/jöhetnek létre a föderált rendszerben.

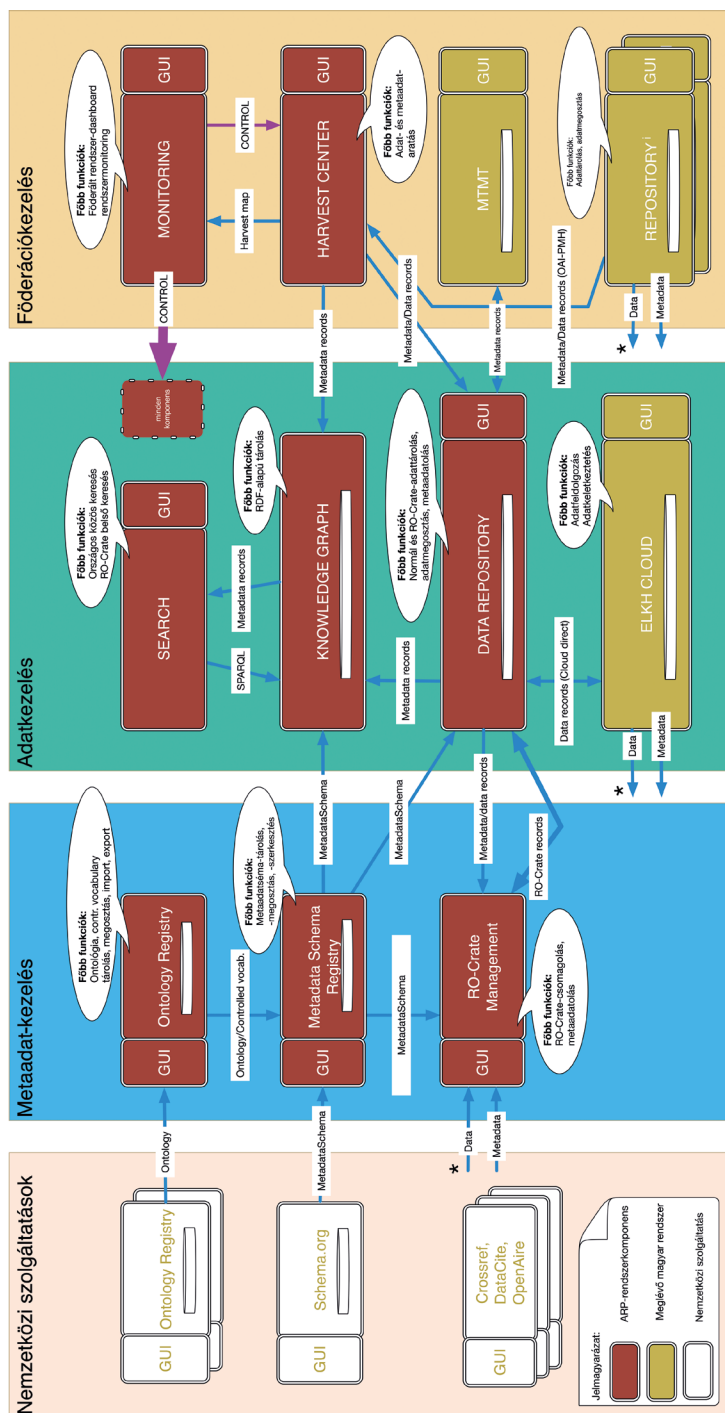
2.2.4. Föderáció a működtetés tekintetében

A hálózatba kötött repozitóriumok adatkapcsolati szintű föderációja az adatáramlások, aratási logikák, funkciók, aratási ütemezés stb. rendszerszintű konfigurálásával együttesen értelmezhető. Ezen a szinten a repozitóriumok gazdái, a repozitóriumok üzemeltetését végző operátorok páros (P2P) megállapodásainak rendszere biztosítja az adatkapcsolatok megfelelő konfigurálását, a repozitóriumok közötti adatáramlások, adatfolyamok racionális megszervezését és azok folyamatos fenntartását, felügyeletét.

A föderált rendszerben értelmezett adatfolyamrendszer operátori áttekintését hálózati szintű központi adatfolyamrendszer-megjelenítés (dashboard) segíti.

³ OAI-PMH – Open Archives Initiative Protocol for Metadata Harvesting, Nyílt Archívumok Kezdeményezés metaadatgyűjtő protokollja

⁴ OAI-ORE – Open Archives Initiative Object Reuse and Exchange, Nyílt Archívumok Kezdeményezés objektumfelhasználás és -csere protokollja



3. ábra. Az ELKH ARP-platform logikai architektúrája
(a szerző saját szerkesztése)

2.3. ELKH ARP-platform logikai architektúra

Az ARP-rendszer részletes logikai architektúrájának ábráján (3. ábra) az adatkezelés, metaadat-kezelés és föderációkezelés funkcionális komponensei és azok fontosabb adatkapcsolatai láthatók.

Az ARP-platfom közvetlen adatkapcsolatban áll az ELKH Cloud rendszerével, amely az adatok feldolgozását végzi felhőtechnológiákon alapuló szoftverrendszerek segítségével, megteremtve így a kutatási adatokra az adatfeldolgozás, az adattárolás és adatarchiválás szerves egységét, integrációját az ELKH-intézmények (és igény szerint akár egyéb hazai kutatóintézetek, egyetemek) számára.

Az ARP-platform lehetővé teszi a nemzetközi kutatási infrastruktúrákhoz való közvetlen kapcsolódást adatkapcsolati, szemantikus és szervezeti szinteken egyaránt. Lehetőségeket biztosít a járatos nemzetközi hálózati szolgáltatások elérésére, kutatási adatok területén metaadatsémák, ontológiák, adatállományok, metaadat-állományok közvetlen importálására, exportálására.

A rendszer megkönnyíti a tudományos publikációk közzétételénél egyre inkább megkövetelt, a publikációkhoz tartozó repozitóriumi kutatási adatletéteket, adatpublikációkat és az adattársításokat, beleértve a folyóiratok *review* folyamatainak közvetlen, algoritmikusan integrált kiszolgálását (embargó, *blind review* stb.) is. A magyar tudományos szféra központi jelentőségű MTMT- (Magyar Tudományos Művek Tára) szolgáltatása felé kiépített kétirányú adatkapcsolati szint is ezeket segíti elő.

3. ÖSSZEFOGLALÁS

Az ELKH ARP-projekt keretében létrehozandó elosztott föderált hálózati rendszer a kutatási adattárolás és archiválás új minőségét (és új tárolókapacitását) teremti meg Magyarországon. Először készül egységes elvek, intézményhálózatra közös szabályok, szabványok, harmonizált technológiák alapján felépülő országos hálózati nagyrendszer egy olyan, mostanáig nagymértékben elhanyagolt területen, mint a kutatási adatok tárolása.

A kutatási adatok értékesége még az egyedi és/vagy *de facto* megismételhetetlen kutatások esetében sem tudatosult kellőképpen a múltban. A kutatási adatok tárolására, hosszú távú megőrzésére irányuló elvárások nem érvényesültek sem a kutatóintézet-hálózati, sem pedig az intézményi szinteken következetesen és rendszerszerűen.

A kutatási adatok egyéni és/vagy kiscsoportos birtoklása és hasznosítása nem teszi lehetővé a kutatási adatokban rejlő teljes pontenciál kiaknázását, a tudományos és persze más szférák (oktatás, gazdaság, ipar, kormányzat stb.) számára.

Különösen nehézkes a szemantikusan korrekt interdiszciplináris adatfelhasználás, az eltérő sémákkal rendelkező hibrid adatállományok összekapcsolása (linked data). Új adatfelhasználási lehetőségek, új tudományos eredményképzés végtelen tárháza nyílik meg az adatkezelés és -tárolás ezen új minőségi szintjének bevezetésével.

A kutatási adatok esetében az adatszuverenitás kérdése és koncepciója láthatóan világszerte mindenhol előtérbe került. Jól érzékelhető a nagy nemzetközi cégek versenyfutása a kutatási adatok mind nagyobb fizikai birtoklása érdekében.

A (kutatási) adatok direkt, gyors adatkapcsolati elérése lehetővé teszi az adatok gépi (akár mesterségesintelligencia-algoritmusokkal történő) (újra)feldolgozását.

A kutatási adatok másod-, harmad-, többedleges adathasznosításának gyakorlati lehetőségét csak úgy tudjuk biztosítani, ha a korábban kizárólag külhoni repozitóriumokban letárolt kutatási adatainkat – ami a nemzetközi tudományban való részvételünk érdekében elengedhetetlen továbbra is – a külhoni tárolással párhuzamosan hazai, az adatfeldolgozó rendszereink és szoftvereink számára közvetlen és elegendően nagy sávszélességű hálózati eléréssel ellátott saját adat-repozitóriumokban is letároljuk.

Ez az ún. kettős tárolási logika teszi lehetővé később a szabad adatfelhasználást, a teljes hazai adatszuverenitás megteremtését, akár a tudományon belüli, akár azon kívüli felhasználásra. E kettős tárolási logika követésének tényleges szakmai, műszaki és szervezeti lehetőségeit hozza létre, többek között, az ELKH ARP föderált adat-repozitórium-hálózat jelen projektben célzott megteremtése.