



A D A T E L V Ű T U D O M Á N Y

Tartalom

Kivonat	3
Bevezetés	5
A változások	5
A magyarországi helyzet	7
Az ELKH stratégiája a kutatási adatok tárolása, archiválása tekintetében	10
1. A kutatási adatkezelés mint infrastruktúra (szuprastruktúra)	10
2. ELKH központi szabályozási rendszer (Referencia Modell)	10
3. A nemzeti adatinfrastruktúrák kapcsolódása a nemzetközi adatinfrastruktúrákhoz	12
4. A kutatási adatok kezelésének, tárolásának szintjei	13
5. Központi ELKH adatkezelési és -tárolási szolgáltatások	14
6. Az ELKH kutatási adatkezelési, adatrepozitóriummi tevékenységének főbb irányai a különféle szinteken	15
7. Metaadatsémák, adatformátumok, szabványok, ajánlások kezelése	17
8. Adatkezelési terv és adatkezelési biztos (felügyelő)	18
9. FAIR követelmények	18
10. A kutatási adatok minőségének és értékének becslése	19
11. Az adat- és repozitóriumminőség-ellenőrzés szervezeti struktúrája	20
12. Az adatkurátori tevékenység elismerési rendszere a tudományban	20
13. A kutatási adatok mint objektumok (research object)	21
14. A kutatási workflow hazai és nemzetközi megosztása	21
Utószó	22
Köszönetnyilvánítás	22

Adatelvű tudomány (tézisek)

Dr. Kovács László

laszlo.kovacs@sztaki.hu

SZTAKI DSD

Számítástechnikai és Automatizálási Kutatóintézet

Elosztott Rendszerek Osztály

Budapest 2020. július 30.

Vers. 1.3.1

Kivonat

Jelen dokumentum az Eötvös Loránd Kutatási Hálózat (ELKH) kutatási adatkezelési stratégia alapvetését, egy középtávú vízióját vázolja fel. A kutatási adatok hazai kezelésének, tárolásának, felhasználásának mikéntjére országos, többszintű hierarchikus adatkezelési infrastruktúra (szuprastruktúra) létesítését és fenntartását javasolunk. Bemutatjuk a kutatási adatkezelés infrastrukturális rendszerének komponenseit és azok funkcionális összefüggéseit. Foglalkozunk a rövid és közép távon megvalósítandó központi-országos és helyi-intézményi adatszolgáltatások rendszerével, valamint az ELKH intézményekkel és projektekkal szemben támasztott, elvárt adatkezelési szabályokkal. Elemezzük a hazai adatinfrastruktúrák kapcsolatrendszerét, beágyazódását a nemzetközi adatinfrastruktúrák világába.

Az ELKH kutatási adatrepozitóriumi tevékenységének főbb irányai a különféle szinteken:

V.	Nemzetközi (ágazati) szint	Az ELKH teljes intézményi súlyával és konkrét egyedi forrásokkal támogatja a hazai kutatók, kutatócsoportok, intézmények, ágazati adatkezelési infrastruktúrák releváns nemzetközi, elsősorban európai (ágazati) adatkezelési infrastruktúrákhoz való intézményesült (szervezeti, szabályozási, policy, adatkapcsolati, projekt szintű stb.) kapcsolódását.
IV.	ELKH (országos) szint	ELKH (országos) létrehozza, fenntartja és folyamatosan fejleszti, az országos, a teljes intézményhálózat egészére vonatkozó adatkezelési és adatrepozitóriumi központi szolgáltatásokat. ELKH Referencia Modellt (központi szabályozási rendszert) dolgoz ki, mely modell és a benne foglalt úttervek alapján fokozatosan vezeti be a szabályozásokat a szférában. A központi ELKH adatszolgáltatások és a Referencia Modell szabályozásai segítségével támogatja, alakítja a magyar kutatási adatrepozitóriumi hálózatok interoperábilis viszonyait országos szinten. Az elvárások betartásának ellenőrzésére intézményeket alakít ki.
III.	Ágazati (országos) szint	Az ELKH konkrét egyedi forrásokkal támogatja az ágazati (országos) szintű adatkezelési infrastruktúrák létrehozását és azok hosszú távú fenntartását. Elősegíti és preferálja, hogy ágazati szintű adattároló megoldások a nagy központosított országos adatfeldolgozó kapacitások (MTA felhő, KIFÚ felhő, szuperszámítógépek) környezetében jöjjenek létre. Központi adatkezelési szabályokat dolgoz ki az ágazatokkal karöltve és gondoskodik a szabályozások betartásának, az ágazati adatkezelési rendszerek interoperabilitásának létrejöttéről, az (adatkapcsolati, szemantikus, működtetési, működési, stb.) interoperabilitások ellenőrzéséről.
II.	Intézményi szint	Az ELKH központi előírásokat és elvárásokat fogalmaz meg az intézményi szintű adatvagyon védelemről. Betartatja és ellenőrzi az intézményi szinteken létrejövő adatkezelési rendszerek ágazati rendszerekkel történő interoperabilitásának biztosítását. Az ELKH konkrét egyedi forrásokkal támogatja az intézményi szintű kutatási adatrepozitóriumok hosszú távú fenntartását. Az olyan ágazatokban ahol az intézményi és az ágazati szint de facto egybeesik, az ágazati (országos) szintű ELKH tevékenységek érvényesülnek (lásd. Ágazati (országos) szintű ELKH tevékenységek fentebb)
I.	Projekt és/vagy mérőműszer szint	Projekt szinten az ELKH konkrét egyedi forrásokkal támogatja a (leginkább nagy mérőműszerek és berendezések, valamint az adatkezelésben leginkább elmaradott szférák, diszciplínák stb. esetében) a kutatási adatkezelési és repozitóriumi rendszerek létrejöttét.

Bevezetés

A 21. század jelentős változásokat hozott a kutatási adatkezelés mikéntje tekintetében a tudományok teljes szférájában. A tudományos megismerés megfigyeléseken illetve a kísérleteken nyugszik. Az empirizmus, a mért, megfigyelt adatokon alapuló tudomány a modern tudományok művelésének fundamentuma már a 17. század óta.

A változások

Korunkban a kutatási adatok területén lévő változásokat a következők jellemzik leginkább:

1. A kutatási adatok **mennyisége** drasztikusan megnövekedett, leginkább az alkalmazott intelligens, számítógép vezérelt mérőműszerek, szenzorok, az automatikus adatgyűjtési, adatkeletkeztetési lehetőségek miatt.
2. A kutatási adatok **megosztási és tárolási** igénye általános lett és felértékelődött az ezt lehetővé tevő gépi és humán technológiák jelentősége. A kutatási adatok már nem csupán egy adott projektet szolgálnak ki, hanem az adott tudományágban az adatokat megosztva, párhuzamosan illetve egymás után következő akár több kutatási projektet is.
3. A kutatási adatok **multidiszciplináris felhasználása** egyre gyakoribbá válik. A kutatási adatokat nem csupán egy adott tudományág használja, hanem azokat egymással akár korábban kevéssé kommunikáló, érintkező, eltérő tudományokban is felhasználják. E multidiszciplináris felhasználás miatt a kutatási adatok formátumainak, szemantikájának, stb. a keletkezés tudományágában közismert, hallgatólagos, implicit tulajdonságai egyértelműen rögzítettek, expliciten definiáltakká, leírtakká válnak avagy kell válniuk.
4. **A kutatási adatok szakszerű metaadatolása** az adatok újrafelhasználhatósága, az adatkezelés EU által is megkövetelt FAIR kritériumainak biztosíthatósága miatt, mind az adatkezelési metodológiákat, mind az adatokra vonatkozó minőségi paraméterek megállapítását az adattudomány és az adatkezelési gyakorlat egy-egy kulcsterületévé teszi.
5. Az adatheldolgozás **automatikus (számítógéppel segített) feldolgozása** általánossá vált a tudományok minden területén, mely összefüggésben van természetesen a nagy adatmennyiségek

kezelhetőségével, ugyanakkor mindez a véletlen (serendipity) felfedezések, a számítógéppel segített hipotézis generálás új terepét nyitja meg.

6. A **kutatási adatok keletkeztetése, tárolása, feldolgozása stb. esetén univerzális megoldások nem vagy kevéssé lehetségesek** (ellentétben pl. a tudományos publikációk kezelése területével). A kutatási adatok fajtái, formátumai, szemantikája, az adatfeldolgozás módszertanai, kialakult szokásai, a tárolásra használt repozitóriumok erősen diszciplína, sőt diszciplínákon belüli tudományos iskola avagy szakmai csoportosulás függők.
7. A fejlett országok tudományai (nemzeti és nemzetközi) **kutatási adatkezelési infrastruktúrák** évtizedes távlatokban rendelkezésre álló meglétére és felhasználására építhetnek. A kutatási adatkezelési infrastruktúrák a kutatási infrastruktúrák lényegi komponensei és mint ilyenek folyamatos fenntartás és fejlesztések tárgyai.

A tudományművelés mikéntje, az alkalmazott tudományos módszertanok tekintetében érzékelhető változásokat leginkább a következők jellemzik:

1. A **tudományos kísérlet** fogalma megváltozott, kibővült. A tudományos kísérletek illetve a természet, társadalom stb. megfigyelése során keletkezett, mért, összegyűjtött kutatási adatokat a tudományos hipotézisek alátámasztására, validálására, falszifikálására stb. használták/ják. A számítógépes adatfeldolgozási lehetőségek mellett megjelent a számítógépes modellek (pl. digitális iker stb.) létrehozásának lehetősége, mely esetekben a számítógépes modellekkel is kísérletek végezhetők, akár olyan kísérletek is, melyeket a valóságban nem, vagy csak nagyon nagy befektetések révén lehetne megvalósítani. A valódi kísérletek, a megfigyelt adatok valamint a számítógépes kísérletek során nyert adatok összevetése a számítógépes modellek identifikációját, majd később identifikált modell alapú predikciókat, új tudományos következtetéseket is lehetővé tesz.
2. A tudományművelés standard, hipotézis vezérelt módja mellett terjedőben van az **adatvezérelt tudomány művelés** paradigmatis változása, mely a tudományos módszertant drasztikusan változtatja meg és pl. a nagy adathalmazokban történő mintázatkeresés alapján a dedukció mint módszertan mellett az indukció szerepének újra- és felértékelődéséhez vezet.

A magyarországi helyzet

A tudományos adatkezelés és adatkezelés mai magyarországi helyzetét a következők jellemzik:

1. Az adatkezelés minden modern módja napi szinten itthon is elérhető akár saját kutatási környezetben, akár a nemzetközi kollaborációs projektek, kooperációk miatt, mely nem csupán az **újfajta adatok, adatformátumok** stb. használatát, de a nagyon **nagy adatmennyiségek kezelését** is napi problémává tette.
2. Az adekvát adattárolás, az adatrepozitóriumok tekintetében **évtizedes (10-15 éves) lemaradás** érzékelhető a fejlett európai országokhoz viszonyítva. Megmutatkozik mindez az adattárolók és repozitóriumok lényegében minden szintű minőségi és mennyiségi hiányában és az alkalmazott tárolási technológiák, módok nem megfelelő használatában, leginkább a releváns informatikai szakértelem, adattárolási szakértők elégtelen alkalmazása miatt. Az egyes tudományokban, tudományágakban, intézetekben általában belső, saját erőből próbálják megoldani az adattárolási, adatfeldolgozási stb. feladatokat, a külső intézmények, informatikai szakemberek igénybevétele nélkül. (Az okok feltárása nem tárgya e munkának.)
3. A kutatási adatok esetén az **adatvesztések aránya** meglehetősen nagy (lásd. debreceni egyetem felmérése: a megkérdezettek több mint 30%-a válaszolt úgy, hogy saját környezetében tapasztalt már kutatási adatvesztést).
4. **Nagyfokú adatkezelési szakértelem hiány** mutatkozik a magyar tudományosság szinte minden szférájában és diszciplinájában. A diszciplináris kutatók informatikusokkal való együttműködése szegényes. A modern digitális adatkezelési informatikai technológiák, módszerek, gyakorlatok ismeretének hiánya elmaradott adatkezelési, adattárolási rendszerek meglétét, aktuális használatát jelenti Magyarországon. A magyar bölcsészet- és társadalomtudományok művelői jelentős hátrányban vannak ilyen tekintetben a természet- és műszaki tudományok művelőihez képest. A magyar memóriaintézmények jelenlegi helyzete pedig tragikus, ezen intézményekben mindenhol (minőségi) informatikushiány mutatkozik. Mindez az alkalmazott adatkezelési rendszerek minőségében, technológiai fejletlenségében csapódik le: a state-of-the-art ismeretének hiányában még akkor sem alkalmaznak fejlett technológiákat ha az elvileg financiálisan elérhető lenne.

5. A kutatási adatok feldolgozása tekintetében nem előre megtervezett és kialakított, adatmérnöki módszerekkel kimunkált **adekvát adatfeldolgozási munkafolyamatok (workflow)** használatosak, hanem a kutatói “szabadság” hamis tudatában, az adatfeldolgozási informatikai technológiákban legtöbbször alulképzett és/vagy alulinformált egyéni kutató számára legkényelmesebb, legkönnyebben elérhető ad-hoc, házilag elkészített módszereket, szoftvereket alkalmazzák (pl. adatfeldolgozás saját laptopon), függetlenül attól, hogy ténylegesen rendelkezésre áll-e az országban vagy nemzetközileg esetleg fejlettebb adatfeldolgozási workflow/kapacitás/egyéb lehetőség avagy sem. Ez esetben az adatfeldolgozás megfelelő dokumentálása nem, vagy csak nehezen lehetséges, így ez a feldolgozási reprodukálhatóság egyik gátló tényezőjévé válik.
6. **A kutatási adatok a projekt befejezése után.** A kutatási adatok felhasználásának horizontja csak ritkán nyúlik az adott projekten időben túl. A projektek befejezése után az összegyűjtött, keletkezett tudományos adatok további sorsa erősen kérdőjeles. A kutatási adatok sorsa sokszor függ a kutató egyéni hozzáállásától. Az intézményesített, szigorú adatkezelési szabályokkal körbeárasztott előre megtervezett racionális (és gazdaságos) adatkezelés ritka. A digitális adatok kezelése a memóriaintézményekben is kezdetleges állapotban van napjainkban.
7. **A kutatási adatok tulajdonlása** tekintetében, leginkább a bölcsészettudományokban, de néha még memóriaintézményekben is tisztázatlan viszonyok tapasztalhatók. Az egyéni kutatók (a kutatói és magánélet és -érdekek egybemosódása miatt) sokszor tekintik szinte magántulajdonnak az általuk összegyűjtött, kezelt kutatási adatokat és az, hogy mi történik az adatokkal pl. a monográfia megjelenése után, sokszor csak a véletlenül múlik. Saját, egyéni adatkezelési tudást, gyakorlatot, tapasztalatot, informatikai rendszereket alkalmaznak (értékes kutatási adatok tárolása asztalfiókban, digitális adatok tárolása laptopon, személyi számítógépen, saját diszken stb.), mely nem csupán az inadekvát adatkezelési módszerek (pl. adattárolás Word fájlban) miatt, de az adatok hosszú távú megőrizhetősége, az adatok túlélése miatt is életveszélyes. A kutatási adatok újkeletű, digitális formája miatt e fenti esetben a kutató, a professzor halála utáni hagyaték feldolgozása és kezelésének korábban kialakult évszázados rendszere illetve gyakorlati lehetősége is megkérdőjeleződik.
8. A digitális kutatási adatok keletkeztetése (mérőrendszerek), tárolása (adatrepozitórium), feldolgozása (szoftverrendszer) és szállíthatósága (kommunikációs hálózat) egymással többek között adatkapcsolatban lévő

nagyrendszerek interoperabilitását követeli meg. Magyarországon ezen nagyrendszerek együttes tervezése, létrehozása, alkalmazása, a harmonikus adatkapcsolati, fenntartási együttműködések biztosítása nem jellemző. E rendszerek történetileg általában függetlenül alakulnak ki és független életciklussal rendelkeznek.

9. **Kutatási adatkezelés infrastrukturális (szuprastrukturális) szemlélete** is hiányzik. Országos szintű, egységesen tervezett, egész tudományágakra szabott interoperábilis adatkezelési, -tárolási rendszerek hiányoznak. Az adatkezelés megoldása legtöbbször egy-egy projektre korlátozódik a projekt futamideje alatt. Az adatkezelés hosszú távú (műszaki, működtetési) fenntarthatósága mint szempont, nem megkövetelt és nem is finanszírozott sem intézményi sem pedig finanszírozó szervezet/hatóság szinteken.
10. A magyar kutatók a hazai kutatási adatrepozitóriumok hiánya és-vagy a nemzetközi kooperációs projektekben való részvétel miatt a kutatási adatokat jelenleg csak **külhoni repozitóriumokban** tudják elhelyezni. Ez szükséges és kívánatos is! Csak így biztosítható a nemzetközi tudományos életbe való egyértelmű kutatói betagozódás, nem beszélve többek között az **Open Science** mozgalmak és EU kezdeményezéseiknek való szükséges megfelelésről.
11. Magyarországon hiányoznak vagy hiányosak a kutatási adatok kezelését lehetővé tévő **funkcionális, architekturális, technológiai, működési (és egyes esetekben üzleti) modellek és rendszerek, szabványok, szabályzatok, jogok, adat- és információszolgáltatások, regiszterek, valamint ezek interoperábilis rendszere**, tehát mindaz, amely a tudományos adatkezelés országos és intézményi szintű rendszerkontextusát, tágabb értelmű nemzeti infrastruktúráját adná.
12. A **mesterséges intelligencia** témakör központi kormányzati felkarolása az adatkezelést és benne a tudományos adatok kezelését is kulcskérdéssé teszi, így egy előreláthatólag jótékony kormányzati hatást prognosztizálhatunk e területen.

Az ELKH stratégiája a kutatási adatok tárolása, archiválása tekintetében

1. A kutatási adatkezelés mint infrastruktúra (szuprastrukturá¹)

Az ELKH a kutatási adatkezelések, adattárolások, repozitóriumok rendszerét mind a négy magyar kezelési szinten (lásd. 4. pont később) **egységes és interoperábilis adatkezelési infrastrukturális (szuprastrukturális) szemlélettel szervezi országosan**, függetlenül attól, hogy a konkrét adatkezelő rendszert vagy rendszerkomponenst mikor, milyen projekt keretében, milyen felhasználói közösség számára tervezték és implementálták.

Az adatkezelés infrastrukturális megközelítése azon a felismerésen alapul, hogy az adatszolgáltatások egyedileg, önmagukban nem életképesek, hanem más adatszolgáltatások egymáshoz harmonikusan, informatikailag és a kezelt adatok tekintetében szemantikusan is illeszkedő rendszerében, egy mérnökileg megtervezett elosztott logikai hálózatban tudnak csak rendesen létezni és működni.

2. ELKH központi szabályozási rendszer (Referencia Modell)

Az ELKH középtávra (5-10 évre) vonatkozó központi szabályozási koncepcionális rendszert (Referencia Modellt) dolgoz ki. Ebben a modellben meghatározzuk

- a kutatási adatkezelési infrastruktúra fejlesztési útitervét,
- az infrastruktúra szervezési elveit rögzítjük,
- generikus logikai architekturális tervet készítünk,
- technológia felhasználási tervezetet hozunk létre,
- adatkezelési, metaadatkezelési szabályzatokat hozunk létre,
- metaadat szabványosítási útitervet készítünk
- a repozitóriumokkal szembeni minőségi és funkcionális elvárások rendszerét dolgozzuk ki,

¹ A “szuprastrukturá” olyan felettes, az adatok, szoftverek, logikai entitások stb. kezelésére szolgáló adat- és tartalom-infrastrukturát, interoperábilisan szervezett adat- és tartalomkezelő szolgáltatások rendszerét jelöli, mely segíti ezen entitások hatékony menedzselését. A “szuprastrukturá” fogalom megkülönböztetését az infrastruktúra fogalmától az teszi szükségessé, hogy a köznapi nyelv infrastruktúrán leginkább az energiaellátást, a közlekedést, a kommunikációs stb. igényeket kiszolgáló, leginkább hardver jellegű hálózati rendszereket ért. Az infrastuktúra, mint “alant” elhelyezkedő rendszertől eltérően a “szuprastrukturá” a “felett”, a köznapi értelmű infrastruktúra felett elhelyezkedő, a saját szintjén infrastrukturális jelentőségű hálózatosan szervezett rendszert jelöli.

ágazatonként eltérő, de ágazatok között a szükséges mértékben harmonizált tartalommal.

A fejlesztési útitervben a fokozatosság és fontossági sorrend alapján határozzuk meg az országos rendszer komponensei létrehozásának tervezett menetrendjét. A rendszerszervezési elvek meghatározása és betartatása biztosítja az egységes és jól működő rendszerarchitektúra létrejöttét függetlenül attól, hogy egy-egy komponens mikor, milyen körülmények között és hogyan jön létre. A generikus architektúra nem csupán a rendszer felépítését, a rendszerkomponensek meghatározását, hanem a külföldi és hazai más, a rendszer működéséhez feltétlenül szükséges adatszolgáltatásokhoz való rendszerkapcsolatokat is definiálja.

A technológia alkalmazások előtt értékelni szükséges a technológiák képességeit, de leginkább azok hosszú távú fejlődési és alkalmazhatósági lehetőségeinek feltételrendszerét. A gyors technológia fejlődés rendszeres felülvizsgálatot igényel e területen.

Az adat- és metaadatkezelési szabályzatok kidolgozásánál figyelembe kell venni a nemzetközileg kialakult normák, elvárások mellett a magyar sajátosságokat, a kialakult hazai szokásokat, valamint természetesen az adatkezelésre allokálható pénzügyi lehetőségeket is.

A metaadatszabványosítási útiterv meghatározása azért fontos, mert a szabványosítási folyamatok egyrészt rendkívül idő- és forrásigényesek, másrészt az eltérő ágazatok más-más fejlettségi, szervezeti szinten állnak. Egyes tudományos közösségekben még fel sem merült a kutatási adatokhoz társítható metaadatok egységesítésének gondolata, másoknak azonnali létszükséglet az adatcserét lehetővé tevő ágazatilag vagy akár nemzetközileg szabványos metaadatséma megléte és használata.

A Referencia Modell a generikus (minden ágazatra érvényes) és a specifikus (egyes ágazatokra speciálisan vonatkozó) elvárások rendszerét egyértelműen megkülönbözteti. Ennek egyik fontos eleme a repozitóriumokkal szembeni minőségi és funkcionális elvárások meghatározása.

Az ELKH a Referencia Modellt folyamatosan karbantartja és igazítja a nemzetközi és hazai tudomány aktuális igényeihez, a gazdasági lehetőségek változásához, követve természetesen a technológiai fejlődést is.

Az ELKH e Referencia Modellben összefoglalt tervek elvárásainak megfelelően megépülő, konkrét adatkezelő, repozitórium rendszereket támogatja. A megféleléseket akár a konkrét intézményi, intézményhálózati

központi ELKH elvárások rendszerének kialakítása és betartatása, akár a támogatási, finanszírozás rendszereibe beépülő elvárások mentén kényszeríti ki.

A hosszabb távú, átgondolt koncepcionális tervek tudják csak biztosítani azt, hogy a felépülő ágazati rendszerek és komponensek, adatszolgáltatások stb. egymással illetve az ágazaton kívüli országos szintű szolgáltatásokkal interoperábilisan (adatkapcsolati, szemantikus, működési, stb. interoperabilitás) épüljenek fel, akár most keletkeznek akár később.

3. A nemzeti adatinfrastruktúrák kapcsolódása a nemzetközi adatinfrastruktúrákhoz

Az ELKH a tudományterületeken, ágazatonkénti magyar kutatási adatkezelési infrastruktúrák, többszintű (lásd. A kutatás adatok kezelésének, tárolásának szintjei) tárolók létrehozását támogatja.

Felmerül a kérdés, hogy ha a kutatók a kutatási adatokat sokszor szükségszerűen és célszerűen (lásd. magyarországi helyzet 10. pont) tárolják a tudományterületükben járatos illetve a konkrét projektben értelmezett nemzetközi tárolókban, akkor miért kívánatos párhuzamosan ekvivalens magyar tárolókat és adatkezelési infrastruktúrákat létrehozni.

Az ELKH a kutatási adatok hazai felhasználását prioritásnak tekinti és ennek megfelelően biztosítani kívánja a magyar kutatási adatok hazai hasznosításának, a (gazdasági, stb.) "adatönrendelkezés" teljes szabadságát.

Ez azt jelenti, hogy a kutatási adatok csak előzetes, tudatos döntés, válogatás, szűrés stb. után kerülhetnek nemzetközi tárolókba. Ezen elsődleges adatválogatási tevékenységek alapja a magyar adatfelhasználás prioritásának biztosítása, legyen az, az adatok tudományon belüli avagy más szférákban (pl. gazdaságban) történő potenciális felhasználása. ***Az ELKH ajánlást dolgoz ki a nemzetközi kutatási adatexport ellenőrzési rendszerének létrehozását elősegítendő.***

Egyes tudományterületeken a kutatási adatokat nem lehetséges nemzetközi tárolókban elhelyezni, mert nincs (vagy még nincs) releváns (ágazati) nemzetközi tematikus tároló. Ilyen esetekben az egyetlen lehetőség a létrehozott hazai tárolókban történő adatelhelyezés.

A magyar kutatási adatkezelési infrastruktúrákat különféle (fizikai, logikai, szemantikus, szervezeti stb.) szinteken is össze kell kapcsolni az ekvivalens nemzetközi adatkezelési infrastruktúrákkal, tehát a magyar adatkezelési infrastruktúrák nem csupán adatátadás (adat import-export) lehetőségeket nyújtanak, de **megteremtik a potenciális nemzetközi szervezeti kapcsolódásokat**, kapcsolatokat a magyar kutatási szféra teljes egésze számára. Így lehetséges csak a tudomány nemzetközi munkamegosztásában megfelelő nyomatékkal részt venni.

Fenti szempontok mellett a tükrözött adatokat tartalmazó magyar és nemzetközi adattárolók **e redundanciával szolgálhatják a kutatási adattárolás hosszú távú biztonságát** is.

4. A kutatási adatok kezelésének, tárolásának szintjei

A tudományos intézményekben a kutatási adatkezelés alapvetően diszciplináris alapokon, ágazatilag eltérően szervezett. Az ELKH figyelembe veszi ezt a jelenlegi, történetileg kialakult struktúrát és a következő adattárolási szinteken kívánja támogatni a magyar kutatási adatok kezelését:

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

Az egyes adattárolási szinteken legfontosabb adatkezelési funkciók a következők:

V.	Nemzetközi (ágazati) szint	A nagy nemzetközi adatkezelő infrastruktúrák időben tartósan fennmaradva egy-egy tudományág évtizedes adattárolási igényeit biztosítják vagy a nemzetközi kooperatív projektek kiszolgálását végzik a projekt időtartama alatt.
----	----------------------------	---

IV.	ELKH (országos) szint	ELKH (országos) szinten a teljes intézményhálózat egészére vonatkozó adatkezelési és adatrepozitórium központi szolgáltatások helyezkednek el (pl. ELKH kereső: a kutatási adatok központi, országosan egyablakos, kereső szolgáltatása; metadatséma regiszter; szabványosítási, adatminőségbiztosító és ellenőrzési szolgálatok; hosszú távú digitális megőrzési szolgálat; stb.).
III.	Ágazati (országos) szint	Az ágazati szintű adattároló megoldások elsődleges feladata a kutatási adatok megosztásának biztosítása mind országon belül, mind nemzetközileg, valamint a nagy központosított országos adatfeldolgozó kapacitások (MTA felhő, KIFÜ felhő, szuperszámítógépek) közvetlen és célszerű elérése. Az intézményi és ágazati szintű adattárolók (funkcionálisi szempontból) egyesíthetők, ha pl. az intézmény országosan a részdiszciplína egyértelműen vezető intézete, így egy gazdaságos tárolási megoldás keletkezhet biztos intézményi felügyelet és menedzselés mellett.
II.	Intézményi szint	Az intézményi adattárolási szint alapfeladata az intézmény kutatási adatvagyonának védelme, a projektek közötti időintervallumokban történő adattárolás és a jövőbeli valamint párhuzamos projektekben történő adatfelhasználás biztosítása.
I.	Projekt és/vagy mérőműszer szint	Projekt szinten a kutatási adatok keletkeztetését, gyűjtését és elsődleges tárolását kell megoldani, valamint biztosítani kell az adatfeldolgozási rendszerekhez való hálózati adattranszfer megfelelő sávselességű lehetőségét (illetve célszerűen egyes más esetekben az ellentétes irányú folyamatot, mely során az adatfeldolgozó szoftvereket szállítják a (nagy)tömegű adattárolókhoz közel).

Az egyes szinteken alkalmazott tárolási, repozitórium technológia adekvát kiválasztásához, implementálásához, a repozitóriumok elvárt minimális minőségének biztosítása érdekében **az ELKH műszaki (funkcionális és minőségi) ajánlásokat és elvárásokat fogalmaz meg** (lásd. Referencia Modell).

5. Központi ELKH adatkezelési és -tárolási szolgáltatások

Az ELKH országos szintű központi adatszolgálatokat és támogató hálózati szolgáltatásokat hoz létre és tart fenn. Az ELKH által létrehozott és

központilag fenntartott szolgáltatások elsődlegesen a következők (nem teljes lista):

ELKH központi adatkereső: A kutatási adatok központi, országos, egyablakos kereső szolgáltatása;
Országos metadatséma regiszter: Metadatséma, adatfomátum és alkalmazási profil, szabványosított névterek és szótárak regisztere;
Adatminőség és repozitórium ellenőrzési és értékelési szolgálat;
Hosszú távú digitális kutatási adatmegőrzési szolgálat;
Központi ELKH adatportál: Az adatkezelési infrastruktúra fejlesztési útitervek, adatkezelési szabályzatok, elvárások, legjobb gyakorlatok, nemzetközi infrastrukturális kapcsolatok, reprodukálható adatfeldolgozási workflow-k, szoftver eszközök stb. központi portálja;

6. Az ELKH kutatási adatkezelési, adatrepozitóriumi tevékenységének főbb irányai a különféle szinteken

Az országosan elkülönülő szinteken egymásra épülő, egymással adat- és szervezeti-kapcsolatokban lévő rendszerkomponensek helyezkednek el. Az egyes szinteken az ELKH egymástól eltérő támogatási tevékenységeket végez:

V.	Nemzetközi (ágazati) szint	Az ELKH teljes intézményi súlyával és konkrét egyedi forrásokkal támogatja a hazai kutatók, kutatócsoportok, intézmények, ágazati adatkezelési infrastruktúrák releváns nemzetközi, elsősorban európai (ágazati) adatkezelési infrastruktúrákhoz való intézményesült (szervezeti, szabályozási, policy, adatkapcsolati, projekt szintű stb.) kapcsolódását.
----	----------------------------	---

IV.	ELKH (országos) szint	ELKH (országos) létrehozza, fenntartja és folyamatosan fejleszti, az országos, a teljes intézményhálózat egészére vonatkozó adatkezelési és adatrepozitóriumi központi szolgáltatásokat. ELKH Referencia Modellt (központi szabályozási rendszert) dolgoz ki, mely modell és a benne foglalt útitervek alapján fokozatosan vezeti be a szabályozásokat a szférában. A központi ELKH adatszolgáltatások és a Referencia Modell szabályozásai segítségével támogatja, alakítja a magyar kutatási adatrepozitóriumi hálózatok interoperábilis viszonylatait országos szinten. Az elvárások betartásának ellenőrzésére intézményeket alakít ki.
III.	Ágazati (országos) szint	Az ELKH konkrét egyedi forrásokkal támogatja az ágazati (országos) szintű adatkezelési infrastruktúrák létrehozását és azok hosszú távú fenntartását. Elősegíti és preferálja, hogy ágazati szintű adattároló megoldások a nagy központosított országos adatfeldolgozó kapacitások (MTA felhő, KIFÜ felhő, szuperszámítógépek) környezetében jöjjenek létre. Központi adatkezelési szabályokat dolgoz ki az ágazatokkal karöltve és gondoskodik a szabályozások betartásának, az ágazati adatkezelési rendszerek interoperabilitásának létrejöttéről, az (adatkapcsolati, szemantikus, működtetési, működési, stb.) interoperabilitások ellenőrzéséről.
II.	Intézményi szint	Az ELKH központi előírásokat és elvárásokat fogalmaz meg az intézményi szintű adatvagyon védelemről. Betartatja és ellenőrzi az intézményi szinteken létrejövő adatkezelési rendszerek ágazati rendszerekkel történő interoperabilitásának biztosítását. Az ELKH konkrét egyedi forrásokkal támogatja az intézményi szintű kutatási adatrepozitóriumok hosszú távú fenntartását. Az olyan ágazatokban ahol az intézményi és az ágazati szint de facto egybeesik, az ágazati (országos) szintű ELKH tevékenységek érvényesülnek (lásd. Ágazati (országos) szintű ELKH tevékenységek fentebb)
I.	Projekt és/vagy mérőműszer szint	Projekt szinten az ELKH konkrét egyedi forrásokkal támogatja a (leginkább nagy mérőműszerek és berendezések, valamint az adatkezelésben leginkább elmaradott szférák, diszciplínák stb. esetében) a kutatási adatkezelési és repozitóriumi rendszerek létrejöttét.

Az országos szintű repozitórium megoldások esetében (ágazati és ELKH szinteken) a **központi (pl. MTA) felhő infrastruktúra alapú implementációk preferáltak**, ha az nem ütközik ellentétes műszaki követelményekbe (pl. big data tárolás speciális követelményei).

Az **ELKH pilot (demonstrációs) projektek** keretében egyedileg támogatja az egyes jellemző és legjobb gyakorlatok bemutatására leginkább alkalmas

konkrét adatrepozitóriumok létrehozását, általános repozitórium támogatási politikája keretében.

7. Metaadatsémák, adatformátumok, szabványok, ajánlások kezelése

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

A magyar szabványosítás elektronikus támogatása érdekében az **ELKH létrehozza a magyar kutatási metaadatsémák és adatformátumok digitális regiszterét (országos metaadatséma regiszter)** mely nem csupán a szabványosított metaadatsémák nyíltan hozzáférhető (open access) tárolója, hanem támogatja a szabványosítás kialakításának kooperatív, hálózaton értelmezett szakértői (humán) tevékenységét, sőt az ELKH intézmények egyedi használati profiljainak (a metadatsémák használata lokális szokásainak) pontos definiálását is. **Az ELKH fenntartja ezt az országos központi hálózati szolgáltatást és finanszírozza a szabványosítási útitervnek megfelelő csoportos szakértői szabványosítási munkákat.**

A metaadatsémák és adatformátumok magyar szabványai kialakítása esetén elsődleges követelmény az ágazatokban a járatos, avagy már (kvázi) szabványosított **nemzetközi szabványok és ajánlások hazai átvétele, adaptálása** a hazai követelményekre a hazai intézményi lehetőségeknek illetve elvárásoknak megfelelően.

A metaadatséma, adatformátum szabványosításokhoz központilag menedzselte kutatási részterületi ontológiák, névterek, szótárak, szószedetek szabványosítását, közös interoperabilis használatát is elő kell segíteni, a metaadatséma és az adatformátum megosztások mellett. A központi séma regiszter az ezekről az objektumokról való **csoportos tudásfelhalmozást és tudáskezelést is támogatja.**

A hazai szabványosítás munkák célszerű és gazdaságos véghezvitele megköveteli a megfelelő nemzetközi szabványosítási testületekben végzett aktív munkát, a megfigyelői státuszt meghaladó tevékenységeket. Csak így lehetséges elsajátítani a szabványosítás már korábban befektetett nemzetközi munkáit. **Az ELKH támogatja a szabványosítási útitervhez kapcsolódó,**

nemzetközi testületekben végzett ilyen irányú szakértői tevékenységeket.

8. Adatkezelési terv és adatkezelési biztos (felügyelő)

A kutatási projektek esetében, a EU által támogatott nemzetközi projektjei hasonló megközelítésével szinkronban az ***ELKH megköveteli a projektek adatkezelési tervének létrehozását a projekttervek részeként***, mely nem csupán a projekt élettartamára, de azt követően (a kutatási adatok továbbélése, újrafelhasználása fenntartása, esetleges hosszú távú archiválása stb.) is rendelkezik (lásd. FAIR követelmények)

Az ELKH létrehozza intézményi szinten az ***(intézményi) adatkezelési biztos (felügyelő)*** munkakört és megköveteli az ilyen munkakörben való foglalkoztatást minden ELKH intézményben.

Az adatkezelési biztos munkakörben alkalmazott kutatási adatkezelési szakértő támogatja, felügyeli az intézményekben végzett adatkezelési gépi- és humán technológiák kooperációjával megvalósított adatkezelést és biztosítja a kutatási adatok elvárt helyes, pl. FAIR stb. kezelését. Az adatkezelési biztos véleményezi és támogatja az adat- és tartalomkezelési infrastruktúrák (szuprastruktúrák) megvalósítására vonatkozó műszaki és stratégiai tervek létrehozását, valamint felügyeli ezen suprastruktúrák napi üzemeltetését.

9. FAIR követelmények

A FAIR-követelmények (Findable, Accessible, Interoperable and Reusable) a kutatási adatok projektek futása közbeni és utáni megtalálhatóságát, a széles körű hozzáférés biztosítását, az adatok csereszabotosságát, illetve az újrahasznosításhoz szükséges feltételek és metaadatok meglétét követelik meg. ***Az ELKH különféle eszközökkel illetve szabályozásokkal támogatja a FAIR követelmények betartását.***

Mindez azt jelenti, hogy az ***adatkezelési tervben egyértelműen definiálni kell*** azt, hogy a projekt során milyen adatokat gyűjtenek be, azokat hogyan, milyen módszerrel dolgozzák fel, valamint milyen új vagy származtatott adatok keletkeznek.

Az adatformátum szabványoknak való megfelelés, a használt szabványok egyértelmű meghatározásán túlmenően az adatkezelési terv tartalmazza az adatok közzétételének, megosztásának tervezett módozatait, a hozzáférés biztosításának módszereit, jogi, műszaki, szervezeti feltételeit.

Információ szükséges arról is, hogy a kutatási adatok feldolgozási folyamata során milyen adatkurátori (válogatás, szűrés, aggregáció stb.) munkát terveznek végezni, a projekt befejezte után hogyan fogják az adatokat megőrizni és/vagy újrahasznosítani, különféle időintervallumokat feltételezve, valamint azt, hogy az újrahasznosítást milyen származási/nyomkövetés jellegű metaadatok (provenance metadata) támogatják.

Az adatkezelési terv, elvárás szerint, foglalkozik az adatkezelés tervezett költségeivel, a projektek lezárulása után felmerülő, hosszú távon jelentkező fenntartási költségekkel is.

10. A kutatási adatok minőségének és értékének becslése

A kutatási adatok minőségi paramétereinek becslésére, legelőször is az **adatminőséget leíró metaadatok rendszerét, minőségi metrikákat kell meghatározni és szabványosítani**, valamint kialakítani az adatminőségi osztályok definícióját és létrehozni az adatminőség meghatározás módszertanát.

Az adatminőség ellenőrzése nem csupán az alkalmazott minőségi metrikák és módszertanok kialakítását igényelik, de érdekmentes konkrét **minőségmeghatározás intézményesült rendszerét** is ki kell alakítani legalább az ELKH intézményhálózaton belül.

A **kutatási adatok értékének (közgazdasági) becslésére** az ELKH módszertant fejleszt ki, mely nem csupán az adatok újra létrehozásának, pótlásának lehetőségeit veszi számba de az adatok felhasználásának potenciális (hasznosítási) lehetőségeit is. A kezelt **adatok értékessége fokozatosan tervezési szemponttá válik** az adatkezelési és tárolási szuprastruktúrák tervezése és üzemeltetése valamint az adatkezelések során.

11. Az adat- és repozitóriumminőség-ellenőrzés szervezeti struktúrája

Az ELKH az **adatminőség és a repozitóriumok minőségének meghatározása érdekében konkrét vizsgálati és mérőhelyeket, akkreditált minőségellenőrzési intézményt alapít** (Adatminőség és repozitórium ellenőrzési és értékelési szolgálat).

A repozitóriumok esetében a (műszaki, szervezeti, gazdasági, finanszírozási) fenntarthatóság valamint a repozitóriumok adatszolgáltatásainak interoperabilitása a legfontosabb minőséget befolyásoló tényezők.

Az ELKH egységes minőségi elvárásokat támaszt az intézményhálózatban mind a kutatási adatkezelés tekintetében, mind pedig az adattároló repozitórium rendszerekkel szemben. Az adatminőségi paraméterek felhasználása az adatkurátori munkák minőségének és mennyiségének meghatározása miatt is elengedhetetlen. A repozitóriumok rendszeres (periodikus) minőségellenőrzési és minőségbesorolási értékelési folyamaton esnek át.

12. Az adatkurátori tevékenység elismerési rendszere a tudományban

Amennyiben a kutató napi tevékenységének szignifikáns, netalántán túlnyomó részét az adatkurátori és persze az ehhez társuló klasszikus adatfeldolgozási (gyakorlati) munka alkotja (ez ma már nem csupán a memóriaintézményekben, de a kutatóhelyeken is gyakorta megjelenő jelenség), akkor a kutatói lét megalapozását és általános teljesítménymérését jelenleg kvázi egyetlen paraméterben mérő publikációs tevékenység csorbát szenvedhet. Leginkább azért, mert a (magyar) tudományosságban nem alakult ki a tudományos adatközlés és adatpublikálás, tudományon belüli, jutalmazási és elismerési rendszere.

A minőségi tudományos adatok publikálása mint olyan, jelenleg nem vagy alig érvényesíthető tudományos teljesítményként, sem egyénileg, sem pedig intézményi, diszciplináris szinteken. Az adatvezérelt tudományok jelenlegi felívelő korszakában ez sürgősen felülvizsgálandó kérdés, melyet a végtelig feszít az újabb keletű publikációs formák (pl. nanopublikáció, kapcsolt adatok publikálása) elterjedése.

Az ELKH kezdeményezi a magyar tudományosság szférájában a kutatási adat adatkurátori, adatpublikálási tudományos értékelési és jutalmazási rendszerének kialakítását.

13. A kutatási adatok mint objektumok (research object)

A kutatási adatok repozitóriumi tárolása, majd pedig az adatfeldolgozás során az adatentitásokhoz a keletkeztetés és adatfeldolgozás bármely fázisában társított metaadatok elválaszthatatlan kezelése szigorú elvárás.

Az ELKH támogatja az új generációs adatrepozitóriumok és a kutatási objektumok kutatás-fejlesztési munkálatait és gyakorlati használatba vételét.

Az új generációs repozitóriumok számtalan újszerű funkciója mellett többek között már kutatási objektumokat (research object) is kezelnek. A kutatási objektum a kutatási adatok (pl. fájlok) és a társított metaadatok szabványos csomagolásának egységes kompozit entitásai. A kutatási objektumok biztosítják a metaadat társítások szakszerű kezelésének módozatait. A kutatási objektumok hordozzák a kutatási folyamat és benne az adatkezelés teljes munkamentének (workflow) leírását, dokumentálását, sőt az adatfeldolgozók, (humán) közreműködők saját munkájáról is tartalmaznak információkat.

14. A kutatási workflow hazai és nemzetközi megosztása

Az ELKH támogatja az Open Science mozgalom keretében a kutatási workflow, ezen belül a kutatási adatkezelési workflow, mint leírható és megosztható objektumok hazai és nemzetközi megosztását a járatos workflow megosztó hálózati szolgáltatások, regiszterek segítségével (MyExperiment stb.), de csak abban az esetben, ha nincs közvetlen lehetőség a kutatási (adatkezelési) workflow és/vagy annak bizonyos komponenseinek innovációs folyamatokban történő hazai hasznosítására, szabadalmaztatására, és esetleges direkt (kommerciális) hasznosítására, **melyek előnyt élveznek a szabad megosztással szemben.**

Utószó

Az adatelvű, adatvezérelt tudományok jelenkori korszakában a kutatási adatkezelés, adatkeletkeztetés, tárolás, feldolgozás, továbbítás stb. célokra egységes adatkezelési infrastruktúrákat (szuprastruktúrákat) kell létrehozni és mint ilyeneket hosszú távon fenntartani a kutatási adatkincs megfelelően hosszú távú védelme és (újra)felhasználása, hasznosítása érdekében. Az ELKH elkötelezett a kutatási adatok korszerű tárolási, repozitórium technológiák, megoldások, legjobb gyakorlatok stb. hazai elterjesztése tekintetében. ***Mindezt az ELKH egységes és interoperábilis magyar kutatási adatkezelési-tárolási infrastruktúrák létrehozása és fenntartása keretében támogatja.***

Az ELKH támogatja a kutatási adatok kezelését nemzetközileg szervező, releváns szabványok és ajánlások, legjobb gyakorlatok kidolgozását és terjesztését stb. végző nemzetközi szervezet, az RDA (Research Data Alliance) keretében létrejött magyar nemzeti csomópont, a ***HRDA (Hungarian Research Data Alliance Node) hazai és nemzetközi szintű tevékenységét*** a magyar kutatási adatkezelési kultúra fejlesztése érdekében.

Köszönetnyilvánítás

A szerző ezúton is megköszöni az ELKH Repozitórium Munkabizottság valamint a HRDA tagjainak értékes hozzászólásait, véleményét, melyek felvillanyozó erővel hatottak. Név szerint külön köszönettel Holl András (MTA KIK), Gárdos Judit (TK), Mohácsi János (KIFÜ), Kacsuk Péter, Micsik András, Pataki Balázs, Virág Éva (SZTAKI) irányában.