

Útmutató kutatási adatok kezeléséhez

A kutatásfinanszírozók egyre nagyobb hangsúlyt helyeznek arra, hogy a támogatott projektek résztvevői ne csak az eredményeket publikálják, hanem az eredményeket megalapozó kutatási adatokat is elérhetővé tegyék. A tudományos folyóiratok szerkesztői is elvárják a cikkek szerzőitől, hogy a hivatkozott adatokra ne csak utaljanak, hanem azokat megtalálhatóvá és elérhetővé tegyék.

A tudományos közéletben egyre inkább elfogadottá válik az az irányelv, hogy a kutatási adatokat – a vonatkozó jogi és etikai szabályokat, valamint a kutatók érdekeit figyelembe véve – a lehető legnagyobb mértékben nyitottá és hozzáférhetővé kell tenni, azaz az adatkezelés legyen „nyílt, amennyire lehetséges, zárt, amennyire szükséges”.

A kutatási pályázatok többségénél – ha a projektben keletkeznek kutatási adatok – jelenleg is elvárás az adatkezelési terv rövid leírása, de van olyan finanszírozó, amelyik részletes adatkezelési tervet kér a pályázat benyújtásakor.

Az „Útmutató kutatási adatok kezeléséhez”

- segítséget nyújt a kutatóknak a kutatási adatok kezelésének megtervezéséhez és az azt rögzítő kutatásiadat-kezelési terv elkészítéséhez;
- ismerteti a kutatási adatokkal kapcsolatos legfrissebb trendeket és elvárásokat és
- ajánlásokat és javaslatokat tartalmaz kutatás- és kutatótámogató szándékkal.

Az ajánlások és javaslatok az egyes tudományterületeken különböző ütemben építhetők be a kutatások menetébe.

Kutatási adat

A kutatási adatok a tudományos közösség által létrehozott, rögzített, elfogadott és megőrzött tényadatok, amelyek a kutatási eredmények hitelességét támasztják alá.

Keletkezésük módja alapján a kutatási projektekben létrejöhetnek

- megfigyelésen alapuló;
- kísérletek során létrehozott;
- szimulációs eljárás során keletkezett és
- meglévő adatforrások felhasználásával (összegyűjtésével, válogatásával, értelmezésével, feldolgozásával) képzett

kutatási adatok.

Feldolgozottságuk szerint csoportosítva a kutatási adatok lehetnek

- nyers vagy elsődleges adatok (pl. közvetlenül mérésből vagy gyűjtésből származó adatok);
- feldolgozott vagy másodlagos adatok (az elsődleges adatokból származtatva, miután a kutató azokat feldolgozta, pl. újrakódolta, kombinálta, kategorizálta, számításokat végzett velük).

Formátumuk szerint a kutatási adatok lehetnek

- digitálisan keletkezett adatok;
- nem digitálisan keletkezett, de digitalizált adatok és
- nem digitálisan keletkezett és nem digitalizált adatok (pl. kézzel írt jegyzetek, terepnaplók).

Példák kutatási adatokra:

- táblázatok
- mérések, alkalmazások, szimulációk eredményei, az azokból létrehozott adatfájlok
- fotók, filmek, diapozitívok
- rajzok
- audio- és videófelvételek, azok szöveges leiratai
- fehérje- vagy génszekvenciák
- kérdőíves felmérések válaszaiból létrehozott adatfájlok
- interjúfelvételek és -leiratok
- kutatási folyamat során megszerzett és/vagy előállított – digitális és nem digitális – tárgyak
- szövegkorpuszok

Kutatási dokumentáció

A kutatási dokumentáció a kutatási adatok mellett keletkező, azokhoz szorosan kapcsolódó, azokkal együtt kezelendő, tárolandó fájlok összessége.

Példák kutatási dokumentációra:

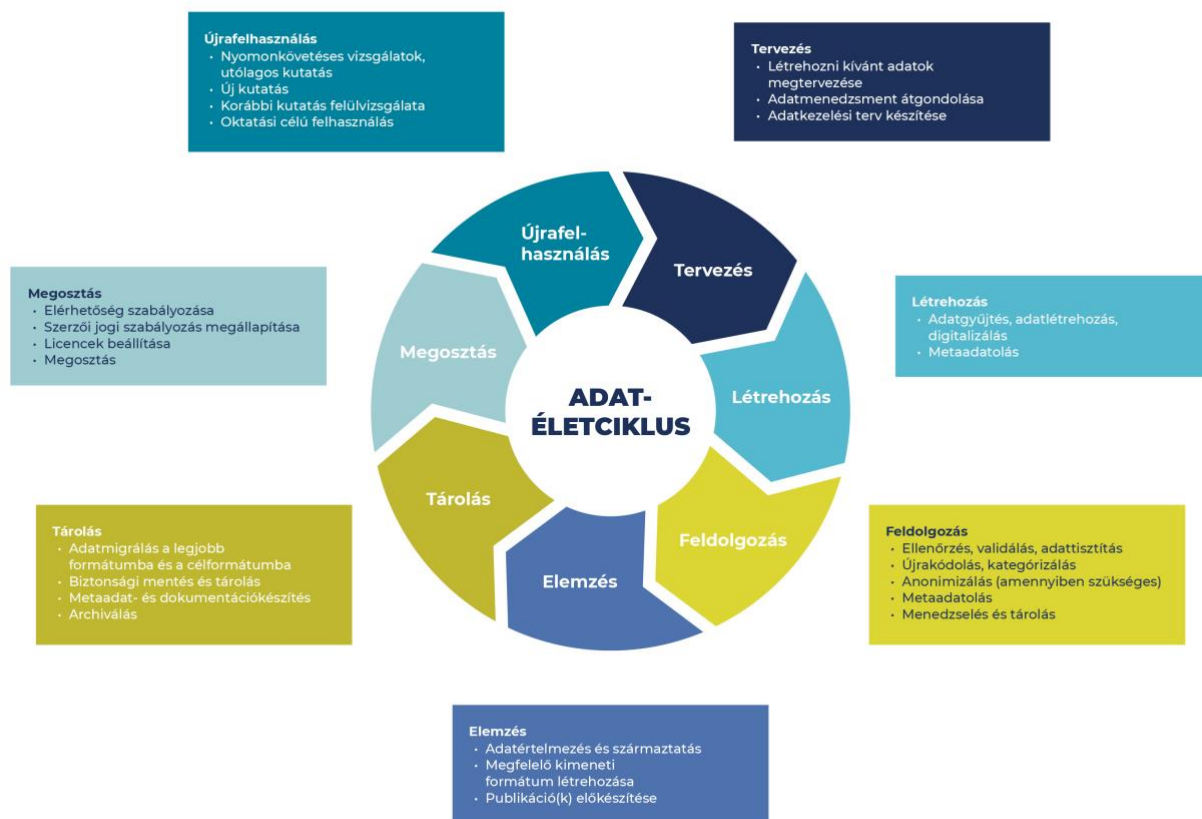
- kutatási tervek
- jegyzetek, vázlatok
- módszerek és munkafolyamatok leírásai
- kérdőívek, interjú vezérfonalak, kódkönyvek
- modellek, algoritmusok, kódok, szkriptek, kutatásra kifejlesztett szoftverek
- laboratóriumi jegyzetek, naplók, feljegyzések, jegyzőkönyvek

Kutatási adatok élelciklusa

A kutatási adatok élelciklusa hét fő szakaszra bontható:

1. Adatkezelés megtervezése
2. Adatgyűjtés, adatlétrehozás
3. Adatfeldolgozás
4. Adatelemzés
5. Adattárolás
6. Adatmegosztás
7. Adatok újrafelhasználása

A kutatási adatok életciklusa



Forrás: https://rdmkit.elixir-europe.org/data_life_cycle, magyarra fordítás, grafikai megjelenítés a HUN-REN ARP-projekt keretében

Kutatásiadat-kezelés

A kutatásiadat-kezelés a kutatási adatokkal kapcsolatos döntések és tevékenységek összessége a kutatás tervezési szakaszától egészen a hosszú távú tárolásig. A kutatási adatok kezelésébe beletartozik a kutatási adatok gyűjtésének, feldolgozásának, tárolásának, másokkal történő megosztásának, archiválásának (hosszú távú tárolásának), elérhetővé és újrafelhasználhatóvá tételének módja, helye és egyéb körülményei is. A jó kutatásiadat-kezelés támogatja és optimalizálja a kutatást. Ha a kutatási adatok nyitottak és ellenőrizhetők, az nagyobb láthatóságot biztosít a kutatásnak és növelheti a kutatómunka hatását.

FAIR kutatási adat-kezelés

A FAIR betűszót a **F**indable (megtalálható), az **A**ccessible (hozzáférhető), az **I**nteroperable (szabványos) és a **R**eusable (újrafelhasználható) szavak kezdőbetűi alkotják. A rövidítést és a hozzá kapcsolódó alapelveket egy kutatókból és kutatási intézményekből álló konzorcium vezette be 2016-ban a *Scientific Data* folyóirat *FAIR Guiding Principles for scientific data management and stewardship*¹ című tanulmányban.

A FAIR alapelvek alapvető célja a tudományos adatok újrafelhasználásának támogatása. Mivel a tudományos adatok mennyiségének, összetettségének, valamint keletkezési és létrehozási sebességének növekedése miatt a kutatók az adatok kezelésében egyre inkább számítanak a gépi támogatásra, ezért az adatok kezelésének azon kell alapulnia, hogy a számítógépes rendszerek emberi beavatkozás nélkül vagy minimális emberi beavatkozással megtalálják és elérjék a különböző kutatási adatokat, együtt tudjanak velük működni és újra fel tudják használni azokat. A FAIR adatkezelés tehát - az eredeti koncepció szerint - nem elsősorban az emberek (kutatók) általi, hanem az algoritmusok általi adatelérést hivatott elősegíteni.

A FAIR adatkezelés alapelvei

Megtalálható (Findable)		Könnyen megtalálható az adat emberek és gépek számára egyaránt? <ul style="list-style-type: none">• Rendelkezik egyedi azonosítóval (pl.: DOI-val)?• Gazdagon ellátott metaadatokkal (leíró adatokkal)?• Jegyezve, indexálva van valamilyen nyilvános, kereshető szolgáltatásban (pl. repozitóriumban)?
Hozzáférhető (Accessible)		Könnyen hozzáférhető, letölthető az adat emberek és gépek számára egyaránt? <ul style="list-style-type: none">• Az adat az adott tudományterületen elfogadott, elterjedt formátumban van?• Segíti szabványos protokoll az adathoz történő hozzáférést?• Van lehetőség a felhasználó (letöltő) azonosítására, amennyiben az adat nem mindenki számára szabadon elérhető?• A metaadatok hozzáférhetőek lesznek akkor is, ha az adatok esetleg már nem?
Szabványos (Interoperable)		Érthetőek és munkafolyamatokba integrálhatóak az adatok? <ul style="list-style-type: none">• A (meta)adatok FAIR alapelveket követő szókészleteket használnak?• A (meta)adatok tartalmazzák más (meta)adatokra vonatkozó minősített kereszt-hivatkozásokat?
Újrafelhasználható (Reusable)		Van lehetőség az adatok jövőbeni újrafelhasználására? <ul style="list-style-type: none">• Tartozik az adatokhoz, az azok értelmezését segítő, releváns információkat tartalmazó leírás?• Rendelkezésre áll egyértelmű felhasználási engedély (licenc)?• Jól dokumentált a (meta)adatok eredete?• Megfelelnek a (meta)adatok a szakmai közösség szokásainak, elvárásainak?

A FAIR adatkezelés nem azonos az Open Data (nyílt kutatási adatok) fogalmával: a kutatási adatoknak nem feltétlenül kell mindenki számára szabadon hozzáférhetőnek lenniük ahhoz, hogy FAIR-nek tekinthessük őket. A megtalálhatóság és hozzáférhetőség elve elsősorban a kutatási adatokat leíró metaadatokra vonatkozik, hiszen azok segítségével tudnak bekerülni az adatok a tudományos vérkeringésbe.

¹ Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

Metaadatok

A metaadatok a kutatási adatok minél teljesebb beazonosítására szolgálnak. A hosszú távú adattárolásnál (repozitálásnál) a kutatási adatokkal együtt tárolt metaadatok segítik a kutatási adatok kereshetőségét, megtalálhatóságát. A metaadatok előállhatnak manuálisan, algoritmus vagy mérőműszer által automatikusan, vagy ezek kombinációjával.

A metaadatok a kutatási adatokra vonatkozó tények és információk összességét jelentik - adatok az adatról. A legáltalánosabb metaadatok az adat elnevezése, készítője, forrása, a készítés dátuma és módszere, de metaadat az eredet, az időbeli hivatkozások, a földrajzi elhelyezkedés, a hozzáférési feltételek vagy a felhasználás feltételei is.

A következő konkrét példa segítségével könnyű megérteni a metaadat fogalmát. A digitális fényképezőgépek a fényképezés pillanatában automatikusan készítenek és onnantól kezdve tárolnak bizonyos információkat az elkészült digitális képpel, azaz digitális fájljal kapcsolatban. Ilyen információk lehetnek:

- a kép készítésének időpontja;
- a kép felbontása;
- a fájl típusa;
- a fájl mérete.

Ezek az adatok metaadatok, mégpedig automatikusan generált metaadatok, amelyek elsősorban a digitális dokumentumról szóló technikai információkat hordozzák.

Modern fényképezőgépek megfelelő beállítás esetén a kép készítésének pontos helyét is tárolják. Ezen túl felhasználóként mi is megadhatunk további információkat, így például egy fotó esetében megadhatjuk a képen szereplő tárgyak nevét vagy az eseményt, amelyhez kapcsolódóan a fotó készült. Ezzel további metaadatokat hozunk létre, amelyek már nem technikai, hanem leíró információkat adnak a digitális képről. Ezek a leíró adatok segítik a kép további felhasználhatóságát és visszakereshetőségét is.

Metaadatok automatikusan és manuálisan hozzárendelve tartozhatnak tehát minden digitális fájlhoz, így a digitálisan elérhető kutatási adatokhoz is. A digitálisan keletkező metaadatok (a fényképezőgépes példában például a kép készítésének percre, vagy másodpercre pontos időpontja) nem csupán a kutatási adatok iránt „kívülről” érdeklődőt támogatják az adatok keresésében és validálásában, hanem a kutató számára is rendkívül pontos és fontos információval szolgálhatnak a kutatás egyes aspektusait tekintve.

Metaadatséma

Mind a digitális dokumentumok (pl. kutatási adatok), mind az azokat tartalmazó digitális gyűjtemények rendkívül sokfélék lehetnek, és nagyon sokféle szinten és módon különbözhetnek egymástól. Így az azokat leíró metaadatok is sokfélék lehetnek, amely miatt utóbbiak használatában egységes szabványok kidolgozása vált szükségessé - az átjárhatóság és a könnyebb kereshetőség biztosítására. Ennek köszönhetően jöttek létre metaadatsémák, amelyek nagyban hozzájárulnak ahhoz, hogy a kutatási adatok és egyéb digitális információforrások könnyen kereshetőek és valóban megtalálhatóak legyenek.

A metaadatséma metaadatelemek meghatározott készlete (elemkészlet) és az ahhoz kapcsolódó szabályok összessége.

Az egyik legelterjedtebb a *Dublin Core* metaadatséma. A Dublin Core elemkészlete leíró metaadatokat tartalmaz, általános, minden tudományterületre alkalmazható, ennek köszönhetően rendkívül széles körben használt. Jellemző rá az egyszerűség és a rugalmasság, a séma jól érthető és átlátható.

A név első eleme a workshop helyszínére utal, ahol 1995-ben kidolgozták. Az Ohio állambeli Dublinban az OCLC/NCSA Metadata Workshop keretében szakértői csoport gyűlt össze azzal a céllal, hogy olyan metaadat-elemkészletet állítsanak össze, amely elégséges és megfelelő az elektronikus információforrások leírására. A második elem (*Core*, azaz *mag*) arra utal, hogy a kialakított elemkészlet olyan alap, amely tovább bővíthető.

Elemkészlete időközben ISO szabvánnyá vált, Magyarországon 2004-ben az „MSZ ISO 15836 Információ és dokumentáció. A Dublin Core metaadat elemkészlete” címen adták ki.

A Dublin Core 15 alapeleme kutatási adatokra adaptálva

1.	Cím	Az adatfájl/adatcsomag címe
2.	Létrehozó	Az adatfájl/adatcsomag szellemi tartalmának alkotója vagy alkotói
3.	Tárgy- és kulcsszavak, jelzetek	Az adatfájl/adatcsomag tartalmát leíró tárgyszavak, kifejezések
4.	Leírás	Az adatfájl/adatcsomag tartalmának ismertetése, összegző leírása
5.	Kiadó	Az adatfájl/adatcsomag nyilvánosság számára történő közvetítéséért felelős entitás
6.	Közreműködő	Az adatfájl/adatcsomag szellemi tartalmának, vagy annak egy részének létrehozásában részt vevő személy(ek) vagy intézmény(ek)
7.	Dátum	Az adatfájl/adatcsomag keletkezésének időpontja
8.	Típus	Az adatfájl/adatcsomag típusa, műfaja, általános sajátosságai
9.	Formátum	Az adatfájl/adatcsomag fizikai vagy digitális megjelenési formája, kiterjesztése
10.	Azonosító	Az adatfájllra/adatcsomagra utaló egyértelmű hivatkozás, az adatfájlt/adatcsomagot egyértelműen azonosító kód
11.	Forrás	Hivatkozás arra az eredeti információforrásra, amelyből a jelen adatfájl/adatcsomag származik
12.	Nyelv	Az adatfájl/adatcsomag tartalmának nyelve
13.	Kapcsolat	Hivatkozás az adatfájllal/adatcsomaggal kapcsolatban lévő más információforrásra (pl. publikációra)
14.	Térbeli és időbeli lefedettség	Az adatfájl/adatcsomag tartalmának térbeli vagy időbeli vonatkozásai, kiterjedése
15.	Jogok	Információk az adatfájl/adatcsomag felhasználásával kapcsolatos jogokról

A különböző tudományterületek más és más, a területen jellemzően létrejövő kutatási adatok sajátosságainak megfelelő, specifikusan kialakított metaadatszabványt használhatnak. Megfelelő szabvány híján az egyes intézmények, projektek, kutatócsoportok saját szabványt hozhatnak létre. Új séma kialakításánál fontos szem előtt tartani, hogy a sémák a FAIR *Szabványos* elvének megfeleljenek, azaz adott tudományágon belül egymással megegyezők, egymás között átjárhatóak legyenek. Jelenleg több nemzetközi szervezet (EOSC, RDA, GOFAIR) dolgozik azon, hogy olyan kereteket alakítsanak ki, amelyek iránymutatást adnak a metaadatszabványok létrehozásában.

Az egyes repozitóriumnak gyakran saját metaadatsémáik vannak, ezek azonban általában átjárhatóak egymás között. A kutatási adatokat leíró metaadatsémát, illetve az abban felsorolt metaadatlistát az adatrepozitóriumok a kutatási adatok feltöltésekor ajánlják fel. Az adatrepozitóriumok rendszere a megadott metaadatokat a kutatási adatokhoz csatolja, többek között ezzel is biztosítja utóbbiak megfelelő kereshetőségét.

Kutatásiadat-kezelési terv (Data Management Plan, DMP)

A kutatásiadat-kezelés írásos, rögzített formája a kutatásiadat-kezelési terv. A kutatásiadat-kezelési vagy adatmenedzsment terv néhány oldalas összefoglaló dokumentum, amelyben a kutató vagy kutatócsoport rögzíti az adatkezeléssel kapcsolatos döntéseket és tevékenységeket, biztosítva az adatok szabályos és etikus gyűjtését és kezelését, biztonságos tárolását a kutatás során és annak lezárását követően.

Elkészítése a kutató, kutatócsoport saját érdeke is, de egyre általánosabb, hogy a kutatások finanszírozói és a pályázató szervezetek már a kutatási koncepció benyújtásakor kérik az adatkezelési terv első (kezdeti) verziójának csatolását, előmozdítva ezzel az átgondolt adatkezelést, az elérhető, fenntartható és újrahasznosítható adatok létrejöttét, valamint a tudásmegosztás és a nyílt tudomány alapelveinek minél hatékonyabb alkalmazását.

A kutatásiadat-kezelés átgondolása, valamint a kutatásiadat-kezelési terv elkészítése

- támogatja a kutatást;
- segíti a kutatási adatok tudatos kezelését, hosszú távú megőrzését, későbbi újrahasznosulását;
- hozzájárul az adatvesztés elkerüléséhez és
- amennyiben van ilyen irányú intézményi vagy finanszírozói elvárás, megfelel annak.

A kutatásiadat-kezelési terv strukturált, többnyire vázlatos formában megírt, gyakran táblázatszerűen rendezett dokumentum, amelynek kidolgozottsága, részletessége a kutató döntésétől és/vagy a finanszírozó, pályázató, kutatási intézmény elvárásaitól függ. Alapvetően az alábbi főbb pontokat tartalmazza, amelyek azonban kiegészíthetők további, adatkezeléssel kapcsolatos fontos információkkal is:

- Adminisztratív információk
 - Projekt/kutatás címe
 - Projekt/kutatás rövid leírása
 - Résztvevő kutató(k) neve(i)
 - Adatfelelős neve

- Kutatási adatok
 - Adatok gyűjtésének, létrehozásának módja
 - Keletkező adatok típusa, jellemzői
 - Adatok feldolgozásának módja
 - Metaadatok, metaadatsémák megadása, leírása
- Adattárolás
 - Adatok kiválasztása
 - Adattárolás és -megosztás kutatás közben
 - Adattárolás és -megosztás a kutatás lezárultát követően
- Etikai és jogi megfelelés
- Költségek és források
- Egyéb, tudományterület-specifikus információk

Bizonyos esetekben pályázat, megbízás esetén a kiíró, megbízó, finanszírozó biztosítja a kutatásiadat-kezelési terv sablonját vagy részletes útmutatóját (pl. OTKA Kutatásiadat-kezelési terv, Template HORIZON 2020 Data Management Plan), de a kutatási intézmények is kialakíthatnak saját sablont vagy elkészítési útmutatót. Ezek hiányában a kutató maga is összeállíthatja kutatásiadat-kezelési tervét, amelyhez javasolt segítséget kérni adatgazdász szakembertől.

Az adatkezelési terv a kutatás során változtatható, a kutatási szakaszokhoz alkalmazkodó dokumentum. Javasolt a kutatás során rendszeres időközönként felülvizsgálni, ha kell, módosítani rajta.

Hasznos linkek DMP készítéshez:

- <https://zenodo.org/record/6424769>
- https://www.scienceurope.org/media/nsxdyvqn/se_guidance_document_rdmeps.pdf
- <https://zenodo.org/record/4915862>
- <https://snd.gu.se/en/manage-data/guides/dmp-checklist>

Online elérhető kutatásiadat-kezelési terv készítő oldalak:

- DMPonline: <https://dmponline.dcc.ac.uk/>
- ARGOS: <https://argos.openaire.eu/>
- OpenDMP: <https://gitlab.eudat.eu/dmp/OpenAIRE-EUDAT-DMP-service-pilot>
- Data Stewardship Wizard: <https://ds-wizard.org>
- RDMO - RDM Organiser: <https://rdmorganiser.github.io/en/>
- DMPTool: <https://dmptool.org>
- ezDMP: <https://ezdmp.org>
- easyDMP: <https://easydmp.eudat.eu>
- DMP ARIADNEplus: <https://vast-lab.org/dmp/ariadneplus/form/>

Adatok tárolása

A kutatási adatokról mindig készüljön biztonsági mentés, lehetőleg több különböző helyen. Érdemes a „here-near-far” elvet követni, azaz egy példány legyen a saját számítógépen, egy másik lokális adathordozón (pl. külső merevlemez, intézeti szerver), egy pedig távoli szerveren (pl. repozitórium).

Az adattárolás módjának, eszközének kiválasztásakor figyelembe kell venni, hogy

- mennyi ideig szeretnénk vagy kell tárolni az adatokat;
- mekkora adatmennyiségről van szó;
- milyen körrel, hány emberrel, milyen eléréssel szeretnénk megosztani az adatokat;
- keletkeznek-e szenzitív adatok;
- mekkora erőforrás áll rendelkezésre,
- vannak-e intézményi vagy egyéb előírások az adatok tárolásával kapcsolatban.

A kutatási adatok tárolására leggyakrabban használt megoldások:

- személyes helyi adattárolási megoldások (pl. saját számítógép, külső merevlemez)
- intézményi helyi adattárolási megoldások (pl. intézményi számítógép, intézményi szerver, külső merevlemez)
- személyes felhőalapú tárolási szolgáltatások (pl. OneDrive)
- intézményi felhőalapú tárolási szolgáltatások (pl. intézeti felhő)
- Git típusú repozitóriumok
- személyes weboldal
- intézeti/projekt weboldal
- adatbázisok
- adatrepozitóriumok

Az adattárolás két típusa:

- kutatás alatti adattárolás
- kutatás lezárulta utáni hosszú távú adattárolás (archiválás)

A két típusú adattárolás módját, helyét alaposan átgondolva, a fenti szempontokat és lehetőségeket mérlegelve kell meghatározni. Az adatokat már a kutatás során is lehet repozitóriumban tárolni, de a repozitóriumok (egyelőre) elsősorban a kutatás lezárulta utáni adattárolásra szolgálnak. Jelenleg az adatrepozitóriumok tárolás a kutatási adatok hosszú távú tárolásának legbiztonságosabb módja.

Fájlstruktúra, fájlok elnevezése

A mappákat és adatfájlokat úgy rendszerezzük és nevezzük el, hogy megfeleljenek az adott kutatás logikájának és mások is értelmezni tudják azokat. Törekedjünk arra, hogy a projekten dolgozó összes munkatárs azonos szabályokon alapuló fájlnevezéseket használjon. Az adatkezelési tervben tüntessük fel a metódust, amely alapján a fájl- és mappastruktúrák és az elnevezések létrejönnek.

A fájlok nevébe felvehető, azok megkülönböztetését elősegítő információ lehet:

- a kutatás / projekt / kísérlet címe vagy betűszava
- hely / térbeli koordináták
- kutató neve / kezdőbetűi
- keletkezés dátuma
- adatok típusa
- verziószám

Példa a fájlstruktúra- és fájlnevezési rendszerre



Forrás: <https://www.wur.nl/en/Value-Creation-Cooperation/Collaborating-with-WUR-1/Organising-files-and-folders.htm> / magyarra fordítás, grafikai megjelenítés a HUN-REN ARP-projekt keretében

Adatrepozitóriumok

A kutatási adatok hosszú távú tárolása különböző módokon történhet. Az archiválás legbiztonságosabb helye az adatrepozitórium, amely digitális kutatási adatok biztonságos és hosszú távú tárolására, archiválására, valamint közzétételére, megosztására és hozzáférhetővé tételére szolgáló komplex infrastruktúra.

Miért jó adatrepozitóriumban elhelyezni a kutatási adatokat?

Az adatrepozitóriumban való adatelhelyezés elsődleges célja az adatok hosszú távú tárolásának és hozzáférhetőségének biztosítása.

Az adatrepozitórium-használat előnyei

- a biztonságos adattárolás- és megőrzés,
- a biztonságos (nyílt vagy korlátozott) adatmegosztás lehetősége,
- az adat-újrafelhasználhatóság biztosítása,
- a kutatás láthatóságának növelése,
- a pályázatok, intézményi előírások, folyóiratok elvárásainak való megfelelés.

Mit repozitáljunk?

A kutatási adatok adatrepozitóriumban történő elhelyezése elsősorban a kutatás során gyűjtött, mért, keletkezett, felhasznált és származtatott kutatási adatokat érinti. A kutatási adatok azonban önmagukban nem vagy nem megfelelően értelmezhetők, így ugyanannyira fontos az azokat kiegészítő, megmagyarázó, kontextusba helyező, illetve az azok megtalálhatóságát biztosító információk, információforrások repozitóriumba helyezése is.

A repozitóriumba helyezendő csomag tartalma:

- kutatási adat
- kutatási dokumentáció
- kutatási algoritmus, szoftver, modell (ha van)
- readme fájl
- metaadatok

Annak érdekében, hogy a kutatás később is érthető, áttekinthető, újrafelhasználható legyen, kiemelten javasolt readme fájl készítése és mellékelése a repozitóriumban való elhelyezéskor. A readme fájl összefoglalóan ismerteti a kutatás körülményeit, illetve a legfontosabb tudnivalókat (a kutatás rövid leírását, hátterét, módszertanát, az adatcsomag tartalmát, a fájlok leírását stb.) ahhoz, hogy a kutatási adat és dokumentáció a felhasználó számára érthető, használható legyen. A readme fájl lehet vázlatos vagy összefüggő szöveges dokumentum, és alapulhat a kutatás bármely leírásán (pl. a kutatási terven, pályázati anyagon).

A metaadatok megadása általában nem külön fájlban, hanem az adatrepozitórium felületén történik, azok a feltöltéskor megjelenő mezők kitöltésével kerülnek az adatok mellé.

Hogyan repozitáljunk?

Ahhoz, hogy az adatrepozitóriumba feltöltött kutatási anyag később olvasható és újrafelhasználható legyen, a következő szempontokat kell figyelembe venni:

1. Megfelelő kutatási adatok és kutatási anyagok (pl.: dokumentáció, algoritmus, szoftver) kiválasztása, az egy adatcsomagba tartozó fájlok meghatározása: nem kell minden, a kutatás során keletkezett fájl adatrepozitóriumba helyezni, gondos odafigyeléssel kell kiválasztani a kutatás eredményeinek alátámasztására használt vagy az adatgyűjtés során keletkezett egyéb értékes, később felhasználható adatokat és dokumentációkat és egyéb kutatási anyagokat

2. Adattisztítás, anonimizálás: szükség esetén az adatok tisztítása, anonimizálása, megfelelő előkészítése a további felhasználásra
3. Az adatok körültekintő címkézése: a kutatási adatok értelemeszerű címkékkel, leírással, magyarázattal való, a tudományterület elvárásaival és szokásaival egyező ellátása az adatok értelmezhetőségének biztosítására
4. Megfelelő fájlformátum: amennyiben lehetséges, a tudományterületen bevett, elterjedt, nyíltan elérhető szoftvert igénylő, mások számára olvasható fájlformátum használata,
5. Érthető fájlnevezés: a fájlok lehetőleg érthető, következetes, a tartalomra utaló elnevezése
6. Átgondolt, áttekinthető adat- és fájlstruktúra: az adatok és fájlok olyan, a tudományterületnek és a kutatás menetének megfelelő rendszerezése, amely lehetővé teszi azok későbbi átláthatóságát
7. Verziókövetés jelölése: amennyiben egy adott kutatási adat vagy dokumentáció több verziója is bekerül a repozitóriumba, azok egyértelmű megkülönböztetése a változás nyomonkövethetőségének biztosítására
8. Readme fájl: a kutatás és az adatok értelmezésének és kontextusba helyezésének biztosítása érdekében kiemelten javasolt readme fájl elhelyezése az adatokkal együtt
9. Részletes metaadatolás: az adatok értelmezhetősége és kereshetősége biztosítására a tudományterületen és az adatrepozitóriumban használt metaadatsémának megfelelően
10. Jogok, etikai szempontok, licencek: csak azon kutatási adatok és dokumentációk repozitálása, amelyek a jogi és etikai feltételeknek megfelelnek, illetve olyan hozzáférés beállítása, amely biztosítja azok betartását

Hol repozitáljunk?

A megfelelő adatrepozitórium kiválasztása során általános és kutatásfüggő szempontokat érdemes figyelembe venni.

Általános szempontrendszerrel dolgozott ki 2020-ban egy kutatócsoport TRUST Alapelvek (TRUST Principles) néven (<https://www.nature.com/articles/s41597-020-0486-7>). A betűszó a Transparency (átláthatóság), Responsibility (felelősség), User focus (felhasználóorientáltság), Sustainability (fenntarthatóság) és Technology (technológia) kifejezések kezdőbetűiből áll.

A TRUST Alapelvek szerint a megbízható repozitórium ismérvei:

- Átláthatóság: legyen átlátható és a felhasználók számára elérhető szolgáltatási és adattárolási politikája
- Felelősség: vállaljon felelősséget az adatok hitelességéért és integritásáért, valamint az adatrepozitóriumi szolgáltatás megbízhatóságáért
- Felhasználóorientáltság: biztosítson a felhasználói igényeknek megfelelő adatkezelést
- Fenntarthatóság: garantáljon hosszú távú adatkezelést és –tárolást
- Technológia: rendelkezzen biztonságos, állandó és megbízható szolgáltatásokat nyújtó infrastruktúrával

További fontos általános szempontok lehetnek a megfelelő repozitórium kiválasztásánál:

- Biztonság: biztosítsa az adatok megfelelő védelmét
- PID biztosítása: adjon perzisztens (egyedi) azonosítót (pl. DOI) a repozitált kutatási adatoknak, adatcsomagoknak
- Kereshetőség: kínáljon megfelelő metaadatsémát, amely alapján a repozitált adatok és metaadatok kereshetőek és megtalálhatóak
- Nyíltság: a repozitórium és ezzel a metaadatok legyenek mindenki számára elérhetőek
- Testreszabhatóság: az adatok elérhetősége (az adatokhoz hozzáférők körének meghatározása) legyen egyénileg megadható
- FAIR: biztosítsa a FAIR alapelveknek megfelelő adatkezelést

Megfontolandó szempontok lehetnek továbbá például:

- Széles körben való elfogadottság: feleljen meg a folyóirat/finanszírozó/kutatási intézmény követelményeinek, ha vannak ilyenek
- Ingyenesség: legyen ingyenes (a megfelelő affiliáció birtokában) mind a feltöltés, mind a használat
- Helpdesk biztosítása: legyen mögötte gyors és megbízható segítő személyzet

Előfordulhat, hogy adott kutatás esetében az adatok tárolásához specifikus adatrepozitórium kiválasztására van szükség. Ilyen eset például, ha adott folyóirat csak a saját repozitóriumát fogadja el az adatok publikációhoz való csatolásához, vagy ha egy adott műszerrel való mérés után az adatok csak a műszert üzemeltető intézmény repozitóriumában vagy adattárában tárolhatóak.

Alapvető elv, hogy ha elérhető, válasszunk tudományterületi vagy intézményi repozitóriumot. Amennyiben ilyen nem áll rendelkezésre, abban az esetben válasszunk általános repozitóriumot adataink tárolására. Ajánlott, hogy az itthon keletkezett kutatási adatokat hazai adatrepozitóriumba (is) töltsük fel, különösen, ha közpénzből finanszírozott kutatás adatairól van szó.

Intézményi adatrepozitóriumok a HUN-REN-ben

A HUN-REN kutatói számára jelenleg rendelkezésre álló adatrepozitóriumok

- a Kutatási Dokumentációs Központ (KDK) repozitóriuma, valamint
- a Concorda adatrepozitórium, amelynek továbbfejlesztett és kibővített változata 2024 elejétől a HUN-REN Adatrepozitórium Platform (HUN-REN ARP).

A [KDK repozitóriumában](#) a Társadalomtudományi Kutatóközpont (TK) négy intézetének kvalitatív és kvantitatív módszerekkel készült kutatásainak kutatási adatai és dokumentációja (interjúk felvételei, leíratai, vezérfonalaj; kérdőíves felmérések kérdőívei, módszertani leírásai, adatbázisai; terepnaplók, megfigyelések jegyzőkönyvei stb.) érhetőek el, különböző (szöveg, kép, videó stb.) formátumokban. A metaadatok minden esetben nyíltan hozzáférhetőek.

A [Concorda](#) (Concentrated Cooperation on Research Data) a SZTAKI (Számítástechnikai és Automatizálási Kutatóintézet) által működtetett Harvard Dataverse alapon kifejlesztett adatrepozitórium. A repozitórium a teljes kutatási hálózat számára nyitott, a tárolt adatok a „tárolók” (dataverse-ek) hierarchikus struktúrájában helyezkednek el. A kutatóhelyek, kutatócsoportok, kutatók intézményi tárolóikban kezelhetik adatcsomagjaikat.

A HUN-REN ARP a HUN-REN által fenntartott, a TK, a SZTAKI és a Wigner kutatóintézetek közreműködésében fejlesztett intézményi adatrepozitórium. Az adatrepozitórium minden HUN-REN kutatóhely számára elérhető, minden tudományterület adatait tárolni képes infrastruktúra. Az ARP repozitóriuma Harvard Dataverse alapú, azt azonban további funkciókkal kiegészítő többkomponensű rendszer.

Egyedi, állandó azonosító (persistent identifier, PID)

Egy megbízható adatrepozitóriumnak képes a benne őrzött digitális fellelhetőségét és azonosíthatóságát biztosítani. Ennek a képességnek fontos komponense, hogy a repozitált digitális objektumokhoz perzisztens azonosítók kapcsolódjanak, azok az objektumok metaadatai között szerepeljenek. A perzisztens (állandó) azonosítók a digitális objektumok hosszú távú, globális és egyértelmű azonosítására szolgálnak, általában számokból és betűkből álló, linkkel ellátott kódként generálják azokat. A perzisztens azonosítók fontos feladata, hogy konzisztens módon biztosítsák az adatok elérését, azaz abban az esetben is, ha azok tárolási helye megváltozik. A perzisztens azonosítók így az adatrepozitóriumtól független entitások kell, hogy legyenek. A leggyakoribb perzisztens azonosítók a DOI, az ARK, a Handle, az ORCID vagy a ROR. Az előbbieket adatok, adatcsomagok, az utóbbi kettő kutatók (szerzők), és (kutatással foglalkozó) intézmények azonosítására szolgálnak.

A DOI (Digital Object Identifier) a legelterjedtebben alkalmazott azonosító mind a tudományos publikációk, mind a kutatási adatok esetében. Előnye az elterjedtségen, szabványosságon túl a központi metaadattár (az egyes DOI ügynökségek, mint a DataCite és a CrossRef külön adattárat működtetnek), az adatáramlást elősegítő megoldások léte és terjedése (CrossRef - ORCID kapcsolat) és a scientometriai beágyazódás (DOI aratás a cikkekből).







A szabványos azonosítók közül léteznek lokális, helyi adatbázisokkal kiszolgálhatók, ilyen például az ARK (Archival Resource Key). Elterjedt perzisztens azonosító továbbá a Handle, amely nem kereskedelmi alapon fejlesztett decentralizált azonosítórendszer, globális (vagy lokális) bővebb metaadattára azonban nincsen.

Felhasználói engedélyek (licenck)

A kutatási adatok közzétételkor el kell dönteni, hogy a közzétett adatokkal mások mit kezdhetnek, hogyan használhatják fel azokat. Ennek meghatározására szolgálnak a licenck. A legelterjedtebb *Creative Commons (CC)* licenck kutatási adatok esetében is jól használhatók, mivel ismertek, átláthatók és a keresőmotorok számára is olvashatók.

A licenck tulajdonképpen engedélyekként értelmezhetők bizonyos feltételek teljesülése esetén.

Mielőtt licencket választanánk adataink megosztásához, érdemes utánajárni, hogy a kutatási adatok tulajdonosa/a kutatás finanszírozója tett-e bármilyen megkötést e tekintetben (akár a finanszírozáshoz, akár a helyi irányelvekhez kapcsolódóan). A *Creative Commons* licenck hét különböző típusát különböztetjük meg, amelyek két jog és négy korlátozó feltétel variációiból állnak össze. Ezek:

- Megosztási jog² 
- Feldolgozási jog³ 
- „Hivatkozd!” feltétel (BY)⁴ 
- „Ne üzletelj vele!” feltétel (NC)⁵ 
- „Ne változtasd!” feltétel (ND)⁶ 
- „Ugyanígy oszd meg!” feltétel (SA)⁷ 

A jogok és feltételek részletesen a következőket takarják:

Megosztási jog: lehetőséget nyújt arra, hogy a licencelt mű szabadon másolható, terjeszthető, bemutatható és előadható legyen.

Feldolgozási jog: lehetőséget nyújt arra, hogy a licencelt mű alapján származékos művek (feldolgozások) legyenek létrehozhatók.

„Hivatkozd!” feltétel: a licenct kibocsátót (azt kibocsátani joggal rendelkezőt) fel kell tüntetni az adatok használatakor, megosztásakor, egyéb jellegű megjelenítésekor.

„Ne üzletelj vele!” feltétel: célja, hogy megakadályozza az adatokkal való üzleti jellegű visszaélést; sok esetben ún. duális rendszerben kerül használatra, az üzleti felhasználást lehetővé tévő, fizetős licenccel párban.

„Ne változtasd!” feltétel: megengedi egy anyag másolását és terjesztését, de tiltja annak mindennemű változtatását, adaptálását, átformálását vagy lefordítását, vagyis csak az eredeti verzió használható és terjeszthető. Röviden: nem engedélyezett az eredeti anyag felhasználása más anyagok vagy adaptációk létrehozásához.

„Ugyanígy oszd meg!” feltétel: az adatokból származó, azokat felhasználó új munkák megjelentetésekor azokat a forrásadatokhoz tartozó licencekkel azonos licencekkel kell ellátni.

² To share (right)

³ To remix (right)

⁴ Attribution requirement (BY)

⁵ Non-commercial requirement (NC)

⁶ No derivative works requirement (ND)

⁷ Share alike requirement (SA)

Creative Commons licenck

Creative Commons 4.0 licenck	Jogok		Korlátozó feltételek			A licenck korlátozottságának mértéke
Public Domain (CC0) <i>Közkinck</i>						szabad licenck
Attribution (CC-BY-4.0) <i>Hivatkozod!</i>						szabad licenck
Attribution – Share Alike (CC-BY-SA-4.0) <i>Hivatkozod! – Ugyanígy oszd meg!</i>						szabad licenck
Attribution – No Derivative Works (CC-BY-ND-4.0) <i>Hivatkozod! – Ne változtasd!</i>						korlátozott licenck
Attribution – Non Commercial (CC-BY-NC-4.0) <i>Hivatkozod! – Ne üzletelj vele!</i>						korlátozott licenck
Attribution – Non Commercial – No Derivative Works (CC-BY-NC-ND-4.0) <i>Hivatkozod! – Ne üzletelj vele! – Ne változtasd!</i>						korlátozott licenck
Attribution – Non Commercial – Share Alike (CC-BY-NC-SA-4.0) <i>Hivatkozod! – Ne üzletelj vele! – Ugyanígy oszd meg!</i>						korlátozott licenck

Forrás: https://hu.wikipedia.org/wiki/Creative_Commons, magyarra fordítás, grafikai megjelenítés a HUN-REN ARP-projekt keretében, <https://creativecommons.org/about/clicenses/>

A Creative Commons-nak van magyarul is elérhető licenckválasztó felülete, amely a fenti döntések meghozatalát követően megmutatja, melyik licenck illik leginkább az adott kutatáshoz: <https://creativecommons.org/choose>.

Részletesebb útmutatás (angol nyelven): <https://www.dcc.ac.uk/guidance/how-guides/license-research-data#x1-4000>

Etikai szempontok, adatvédelem

Egyes kutatások során személyes, bizalmas vagy titkos adatok is keletkezhetnek, amelyekre szükséges gondosan odafigyelni a kutatás során, a tárolási mód kiválasztásakor, és a hosszú távú megőrzés kialakításakor is.

A kutatások során az emberektől és az emberekről gyűjtött adatok személyes, érzékeny vagy bizalmas információkat is tartalmazhatnak. Etikai és adatvédelmi szempontok minden olyan kutatás esetében felmerülnek, ahol személyek a kutatás alanyai. Érzékeny vagy bizalmas adatok továbbá egyéb területen is keletkezhetnek, ilyen lehet például egy veszélyeztetett faj előfordulási helyét vizsgáló, vagy egy üzleti érdekeket, titkos információkat tartalmazó kutatás, de más speciális területen vagy kutatásban is keletkezhetnek hasonló adatok.

Sok esetben a személyes vagy érzékeny adatok is kezelhetők és megoszthatók az etikai és jogi követelmények betartásával, többek között az alábbi szempontok figyelembevételével:

- A kutatás előkészítése során:
 - A kutatási terv elkészítése során végig kell gondolni, és a kutatási tervben jelezni, hogy keletkeznek-e személyes/érzékeny adatok a kutatás folyamán
 - Tájékoztató és beleegyező nyilatkozatban kell rögzíteni, hogy a kutatás hogyan biztosítja a különböző szakaszokban – beleértve az adatok megosztásának, újrafelhasználásának lehetőségét is – a személyes adatok védelmét
- A kutatás során:
 - Különös figyelmet szükséges fordítani a személyes/érzékeny adatok tárolására, a hozzáférés, megosztás meghatározására és módjára (pl. felületére)
 - Kizárólag feltétlenül szükséges esetben, és a lehető legrövidebb ideig indokolt a személyes/érzékeny adatok kezelése és tárolása
 - A kutatásban való részvételnek a résztvevő szabad akaratából kell történnie
- A kutatás lezárultát követően
 - Ha kutatási szempontból már nem indokolt a személyes/érzékeny adatok megőrzése, gondoskodni kell azok biztonságos és teljes törléséről
 - Ha a személyes/érzékeny adatok nem törölhetők, álnevesítéssel, anonimizálással csökkenthető a személyes adatkezelés negatív következményeinek kockázata
 - A GDPR szerint anonimizált adatok esetében nem kell alkalmazni a személyes adatok védelmére vonatkozó szabályokat
 - Kutatási adatok repositálása esetén, személyes/érzékeny adatokat is tartalmazó kutatás esetén különös odafigyeléssel kell figyelembe venni a kiválasztott adatrepozitórium vagy adattár felhasználási feltételeiben leírtakat
 - A legtöbb esetben csak olyan kutatás kutatási adatai repositálhatók, amelyek nem tartalmaznak személyes/érzékeny adatot, vagy más, személyek beazonosítására alkalmas információt
 - Kivételt jelenthet, ha:
 - valamennyi érintett személy (pl. a kutatás finanszírozója, kutatásvezetője és valamennyi, személyes/érzékeny adattal érintett alanya) kifejezett írásos beleegyezését adta a személyes/érzékeny adatok tárolásához és nyilvánosságra hozatalához, és a feltöltő ezek birtokában, valamint a megfelelő hozzáférési szintek beállításával helyezi el azokat
 - valamennyi azonosított vagy azonosítható személy elhunyt, és nincs olyan körülmény vagy jogszabály, amely kifejezetten korlátozná a személyes adatok vagy információk kiadását
 - a személyes/érzékeny adatokat is tartalmazó kutatási adatokhoz történő repositóriummi hozzáférés engedélyköteles (a feltöltő szándékának és jogosultságának megfelelő beállításokkal) és az engedélykérés a rendszerben megoldott

Mesterséges intelligencia a kutatásban

Tudományterületfüggő, hogy milyen mértékben, de általános tendencia, hogy egyre több kutatásban használnak valamilyen módon mesterséges intelligenciát, mesterséges intelligencia alapú algoritmust, mesterséges intelligencia által támogatott kutatási eszközöket. Mesterséges intelligencia segítségével készült adat, adatelemzés, szöveg, tanulmány vagy egyéb kutatási anyag esetében fel kell tüntetni az alkalmazott módszert és eszközt, illetve megfelelő (visszakereshető) módon hivatkozni kell rá.

Adatrepozitóriumok használatával kapcsolatos ismeretek terjesztése, kutatásiadat-kezelési tanácsadás a HUN-REN ARP-projektben

A repozitóriumi infrastruktúra kialakítása mellett a HUN-REN Adatrepozitórium Platform projekt kiemelt feladata az adatrepozitóriumok használatával kapcsolatos ismeretek terjesztése, a szükséges HUN-REN- és intézményi szintű adatkezelési irányelvek és szabályozások létrehozása vagy annak támogatása, valamint a hazai és nemzetközi adat- és metaadat-kezelési, -tárolási szabványok és ajánlások, jógyakorlatok meghonosítása, ezzel is lehetővé téve a FAIR adatrepozitórium-kultúra megteremtését a HUN-REN intézményhálózatán belül.

Ennek érdekében

- A HUN-REN ARP honlapján folyamatosan frissülő ismeretanyag olvasható a kutatási adatokkal és azok kezelésével kapcsolatban
- A honlapon és különböző kommunikációs csatornákon keresztül folyamatos tájékoztatást kapnak a kutatók a közelgő rendezvényekről, eseményekről
- A HUN-REN ARP szakértői rendszeresen tartanak előadásokat, képzéseket, a korábbi események felvételei [megtekinthetők a HUN-REN ARP oldalán](#)
- A HUN-REN ARP Nagykövet Program keretében elkezdődött az intézményi adatgazdász hálózat kialakításának megalapozása, az intézményi jógyakorlatok kialakítása
- A HUN-REN ARP munkatársai rendelkezésre állnak bármilyen, a témával kapcsolatos kérdésben, probléma megoldásában

A HUN-REN kutatói információkat szerezhethetnek a HUN-REN ARP <https://science-research-data.hu/> honlapjáról, valamint az ARP szakembereitől. Az ARP szakemberei készséggel állnak rendelkezésre konzultációs lehetőséggel és támogatás nyújtásával a kutatás bármely szakaszában az adatkezeléssel és a repozitóriumhasználattal kapcsolatban felmerülő kérdések megválaszolására az info@science-research-data.hu és a kdk@tk.hu email címeken.