

Ajánlások a kutatási adatok kezeléséhez

1. Szempontok az adatkezelési terv elkészítéséhez

Kvalitatív és kvantitatív tudományos kutatásaink során jelentős mennyiségű *kutatási adat*¹ keletkezik, amelyet a kutató szeretne biztonságban tudni és amelyek biztonságáért, tárolásáért, megőrzéséért egyben felelősséggel is tartozik. Ezért szükséges elkészíteni az adatkezelési tervet, amely a kutatás különböző szakaszaihoz illeszkedő 1-2 oldalas dokumentum, és amely részletesen leírja, hogy a kutatási folyamat során keletkezett adatok kezelése hogyan történik.

Kutatási tervek készítésekor egyre gyakoribb, hogy a kutatások finanszírozói ill. a pályáztató szervezetek (pl. OTKA, Élvonal pályázatok) már a kutatási koncepció benyújtásával egyidejűleg kérik az adatkezelési terv csatolását, ezzel is előmozdítva az átgondolt adatkezelést, az elérhető, fenntartható és újrahasznosítható adatokat, a tudásmegosztást, és a nyitott tudományokban való minél aktívabb részvételt. Az adatok, az adatbázisok elérésére vonatkozóan az „*as open as possible, as closed as necessary*” irányelvet kell átgondolni és figyelembe venni.

Az adatkezelési terv elkészítésére valójában már a kutatás kezdeti szakaszában, a tervezéskor szükség van. Ennek a tervnek tartalmaznia kell az adatkezeléssel kapcsolatos legfontosabb információkat:

- hogyan, és milyen célból történik az adatgyűjtés,
- ki lesz a felelős az adatokért (vagyis ki lesz az adatgazda),
- milyen típusú adatok keletkeznek,
- mekkora méretű lesz az adatbázisunk, milyen és mekkora tárhelyre lesz szükség,
- hogyan kezelik, hol tárolják őket,
- hogyan biztosítható az adatok védelme, biztonságos tárolása,
- mi történik az adatokkal a kutatás befejezését követően,
- kinek/kiknek a számára lehetnek fontosak az adataink az újrahasznosítás szempontjából,
- szabványok/standardok és metaadatok: az adatgyűjtésre vonatkozó szabványok, módszerek, amelyeket az adatgyűjtéskor és az adatok kezelésekor használnak, valamint indoklás, hogy miért ezeket választottuk
- ki, mikor, hogyan, milyen jogokkal férhet majd hozzájuk,
- szükség szerint biztosítani kell a személyes adatok védelmét, anonimitását is.

Bár az adatkezelési tervnek vannak fontosabb pontjai, kérdései (egyes pályázatok esetében ezek külön nyomtatványon rögzített kérdések, mint pl. a már említett OTKA pályázatnál), amelyeket a kutatónak mindenképpen meg kell válaszolnia, azonban mégis egymástól jelentősen különbözőek lehetnek, hiszen a kutatási projekt specifikus tulajdonságaitól

¹ Kutatási adatokon a kutatás során keletkező nem-digitális (papír alapú) vagy digitális formátumban gyűjtött, felvett, ill. rögzített adatokat értjük.

függenek és ezekre reflektálnak.

Kutatási adatainkat a **FAIR** alapelveknek² megfelelően kell kezelnünk. Az archivált adataink legyenek:

F: findable/megtalálhatóak mások számára is, ezért metaadatokkal megfelelően ellátottak

A: accessible/elérhetőek könnyű letöltési, hozzáférési szolgáltatáson keresztül

I: Interoperable/együtműködőek

R: reusable/újrafelhasználhatóak mások számára, megfelelő adathasználati engedélyek alkalmazásával

2. Adatbiztonság, adatrepozitóriumok

Az adatok nyílt megosztásával (különösen az *adatrepozitóriumok* használatával) a kutatási eredmények sokkal gyorsabban, eredményesebben és tágabb körben hasznosulhatnak, hiszen a kutatótársadalom más tagjai is hozzáférhetnek (természetesen egy előre le szabályozott - a kutató ill. finanszírozó által kontrollált - módon), ami elősegíti a tudásmegosztást.

A kutatási adatok tárolásának egyben a legbiztonságosabb módja, ha feltöltjük őket egy *adatrepozitóriumba*, amely számos adatbiztonsági követelmény biztosításával segíti, hogy az adataink hosszú távon biztonságba legyenek.

Ehhez végig kell gondolnunk, milyen stratégiát követünk az adataink kezelésekor, hogyan fogjuk az egész kutatás megvalósítása során kezelni az adatainkat, hogyan fogjuk fizikailag és virtuálisan tárolni, hogyan fogjuk védeni őket és biztosítani, hogy csak a megfelelő felhatalmazással rendelkező emberek férjenek hozzá, és a kutatás befejezésekor hogyan fogjuk ezeket megosztani a szélesebb tudományos társadalommal.

A három legfontosabb adatbiztonsági követelmény az adatok tárolási helyének kiválasztásánál, amelyet mindenképpen szem előtt kell tartanunk:

- **biztonság:** az adatokat és az információ tartalmukat csak *megfelelő jogosultsággal* rendelkező személyek ismerhessék meg, *akkor*, amikor ez számukra engedélyezve van (pl. egy adott időszak, embargo lejártá után), és csak azon *a szinten* láthassák, amely számukra általunk engedélyezett. Ugyanez vonatkozik az adatok felhasználására, ill. esetleges újrahasznosításukra is,
- **sértetlenség vagy integritás:** az adatokat ne módosíthassák illetéktelenek, biztosak lehessünk abban, hogy az eredeti állapotukban megmaradnak,

² A FAIR alapelvek az EU-s országok kutatási adatkezelésének, az adatok nyílt hozzáféréseinek az irányelveit jelentik.

- rendelkezésre állás: az adat, az információ és a kapcsolódó infrastruktúra ott és akkor rendelkezésre álljon, amikor arra szükség van.

2.1 Adattárolási lehetőségek

Adatainkat fontos legalább 2-3 egymástól fizikailag is távol lévő helyen lévő biztonságos adattárolón tárolnunk, az adatvesztés elkerülése érdekében. Legbiztonságosabb adatainkat egy adatrepozitóriumba is feltölteni.

A kutatás során keletkező adatok és információk hordozása, tárolása rendszerint az alábbi felületek, eszközök, csatornák valamelyikén történik, amelyekre szintén ki kell, hogy terjedjen az adatbiztonság:

- a papírra, ahová rögzítik,
- a hardverre, amin tárolják és feldolgozzák:
 - saját PC/notebook meghajtó (HDD, eMMC),
 - flash drive, DVD,
 - hálózati meghajtó, server,
- felhő szolgáltatásra, ahol tárolják (pl. Google cloud),
- adatrepozitóriumra (pl. Zenodo),
- a szoftverre, amivel feldolgozzák,
- az operációs rendszerre, ami a szoftver futtatja,
- a kommunikációs csatornára, amin az adatok áramlanak,
- az eljárásokra, amik a működést szabályozzák,
- és a felhasználókra, aki mindezeket kezelik.

2.2 Felhő alapú adatrepozitóriumok

Hazai repozitóriumok *(még kísérleti jelleggel működnek)*:

MTA SZTAKI adatrepozitóriuma (CONCORDA): <https://science-data.hu/> (régén: <https://concorda.sztaki.hu>) Eredetileg kísérleti jelleggel a Covid 19 kutatások összegyűjtésére jött létre, azonban jelenleg bármilyen adatbázis feltöltése lehetséges.

Debreceni Egyetem adatrepozitóriuma: <https://adattar.unideb.hu/>

Nemzetközi interdiszciplináris repozitóriumok:

Zenodo (OpenAIRE): <https://zenodo.org/>

Dataverse: <https://dataverse.org/>

Dryad: <https://datadryad.org/stash>

FigShare: <https://figshare.com>

Mendeley Data (Elsevier): <https://data.mendeley.com › research-data>

Open Science Framework (OSF): <https://osf.io>

Globus: <https://www.globus.org>

ELN'S (Labguru, Labarchives): <https://www.labarchives.com/eln-for-research/>

A fentiek közül kiemelnék a legnépszerűbb ingyenesen használható, és megfelelő adatbiztonsági követelményekkel rendelkező adattárat a *Zenodot*, amely egy általános célú, nyílt hozzáférésű adatrepozitórium, amelyet az Európai OpenAIRE program keretében fejlesztettek ki³, és ahova max. 50GB/adatok méretben tehetünk fel anyagokat. A kutatók, intézmények a publikációjukhoz kapcsolódóan feltölthetik a hozzájuk tartozó adatbázist is. Vannak más adatrepozitóriumi keresők is, amelyek kifejezetten abban nyújtanak segítséget, hogy céljaink és adataink számára a legmegfelelőbb adattárat választhassuk ki:

- <https://www.re3data.org/search>
- http://oad.simmons.edu/oadwiki/Data_repositories
- <https://www.nature.com/sdata/policies/repositories>
- <https://fairsharing.org/databases>
- <https://repositoryfinder.datacite.org>
- <https://explore.openaire.eu/search/find>

3. Adatvédelem, etikai szempontok

A személyes adatok védelme kiemelten fontos szempont. Minden adat, információ, amely egy bizonyos személy, család, háztartás beazonosításához vezet személyes adatnak⁴ minősül, így védelméről gondoskodnunk kell, hiszen a *magyar adatvédelmi rendelet*⁵ hatálya alá tartoznak, valamint 2018 óta meg kell felelnünk a *GDPR*⁶ (*Általános Adatvédelmi rendelet*) elvárásainak is.

A kutatási adatok kezelésének tervében:

- nyilatkozni kell, hogyan biztosítjuk az egész adatkezelési eljárás során a különböző szakaszokban az adataink védelmét, pl. dokumentum védelem: hogyan fogjuk tárolni fizikailag és virtuálisan az adatokat? Hogyan biztosítjuk, hogy mások számára ne legyenek hozzáférhetőek (pl. a kérdőíveket és a személyes adatokat fizikailag más helyiségben tároljuk, vagy páncélszekrényben tároljuk a személyes adatokat, vagy digitális adatainkat virtuális crypt kóddal látjuk el, stb.)?

³ A Zenodo adatrepozitóriumot a CERN Data Centre üzemelteti.

⁴ Példa személyes adatokra: vezetéknév és utónév, lakcím, személyi igazolvány szám, helymeghatározásra alkalmas adatok (a virtuális adatok is ide tartoznak, pl. IP-cím), mobil-telefonszám stb.

⁵ Az 1992. évi LXIII. törvény a személyes adatok védelméről és a közérdekű adatok nyilvánosságáról.

⁶ A GDPR (angolul: *General Data Protection Regulation*) az Európai Parlament és a Tanács 2016/679. rendelete, amely hazánkban 2018. májusában lépett érvénybe és a természetes személyek adatait védi, valamint rendelkezik a tagállamok közötti szabad információáramlásról.

- az adatainkat (visszafordítható módon) *anonimizálnunk* is kell (pl. valamilyen kódrendszer használatával)⁷. A szabályoknak megfelelő és egyszerű anonimizálást segíti ez az on-line elérhető anonimizáló algoritmus, az Amnesia: <https://amnesia.openaire.eu/amnesia>
- figyelembe kell vennünk a GDPR⁸ vonatkozó szabályait, pl. kérdőív készítése esetén, nyilatkozni kell a kérdőívben, hogy az adatokat bizalmasan kezeljük, és csak az eredeti kutatási célra használjuk fel
- nyilatkozni kell, hogyan biztosítjuk az adatfeldolgozás során, hogy a megkérdezett sem egyéni, sem családi, sem háztartási szinten ne lehessen beazonosítható (pl. egyedi azonosító használata, véletlen szám generátorral, aggregált adatok használata stb.)

4. Az adatkezelési terv elkészítése a gyakorlatban

Az adatkezelési terv elkészítéséhez on-line elérhető nyílt forráskódú eszközöket is találunk az Interneten, amelyek végig vezetnek minket a főbb pontokon és személyes igényekre szabható sablonnal segítik az adatkezelési terv elkészítését - amennyiben a kutatásfinanszírózó nem írja elő adott formanyomtatvány használatát. Ilyen hatékony eszközök például a következők:

- DMPTool⁹: <https://dmptool.org/>
- DMPOnline¹⁰: <http://dcc.ac.uk/dmponline/>

Az adatkezelési tervnek tartalmaznia kell:

- az adatokról információt: milyen típusú adatokat gyűjtünk, milyen célból, hogyan kapcsolódik ez a projektünk eredeti céljaihoz,
- mi az adatok tartalma (pl. táblázatok, nyilvántartások, képek),
- mekkora lesz az adatok mennyisége (pl. MB, TB, ZB méretű), mekkora és milyen tárhelyre lesz szükségünk (például átgondolni: papíron is tároljuk az adatainkat, vagy csak elektronikusan rögzítve?),
- hogyan kezeljük, tároljuk, hasznosítjuk az adatokat az adatgyűjtéstől kezdve a kutatás különböző fázisain keresztül, valamint a kutatás lezárulását követően,
- lesz-e kapcsolat más nyilvános adatbázisokkal/adattárakkal,
- másodlagos további felhasználásról információt: az elkészült adatbázis további tervezett felhasználása,
- tisztázza az adatok hozzáférhetőségét, újrahazsnosítását, megosztását, licencét: az adatok megosztásának tervét, az adatok rendelkezésre bocsátásának, mások általi felhasználhatóságának módját (pl. nyilvános adatrepozitóriumok (ld. 2.2 pont), vagy kérésre elküldési lehetőséget, szükség szerint a hozzáférési mechanizmusok rögzítését) Rögzíteni kell, hogy hol, hogyan, kinek, mikor lesz elérhető az adatbázisunk. Ez lehet a kutató, kutatócsoport saját döntése, de gyakran a megrendelő, finanszírózó, pályázatkíró és egyéb más szempontok (pl. a konzorciumi tagok állásfoglalása) is befolyásolhatják ezt. Ha egyáltalán nem kívánjuk megosztani az adatainkat, ennek indoklását.

⁷ Egyes esetekben szükséges lehet az érintett írásbeli hozzájárulása az adatok kezeléséhez (pl. egy több hullámból álló panelvizsgálat során a telefonszámát, címét újra felhasználva megint megkeressük).

⁸ A GDPR megfelelő adatkezelési gyakorlatáról szóló rendeletének szabályait és előírásait az EU 28 tagállamában minden kutatásnál, ahol adatok keletkeznek figyelembe kell venni.

⁹ A DMPToolt készítője a University of California Curation Center of the California Digital Library.

¹⁰ A DMPOnline az angol Digital Curation Centre terméke.

- hasonlóan tisztázni kell a metaadatokhoz való hozzáférést, mikortól, kik számára lesznek elérhetőek, valamint, hogy hol és hogyan tárolják ezeket,
- milyen metaadat¹¹ szabványt használnak? (DublinCore¹², Datacite¹³, stb.)
- adattulajdonosi jogok: az adatmegosztás bármilyen korlátozása az adatok, metaadatok védelmének szükségessége miatt, ki fog gondoskodni a repozitórium felügyeletéről, ha ez szükséges,
- időkeretet az adatok nyilvánossá tételére, vagy türelmi idő (embargó) rögzítését az adatok nyilvánossá válásának, közzétételének ütemezését,
- a végső adatbázis formátumát (pl. laborjegyzőkönyv, jpg, vagy kvantitatív társadalomtudományi kutatások esetén leggyakrabban ez egy cvs, sav, sys vagy dta, do kiterjesztésű fájl)
- gondoljuk át milyen költségei lehetnek az adataink hosszú távú tárolásának (pl. az anonimizálási eljárásnak, vagy DOI-k költsége, ezeket is írjuk le az adatkezelési, valamint a költségvetési tervbe,
- milyen állandó, illetve egyedi azonosítókat használunk majd (pl. DOI¹⁴ vagy ORCID¹⁵) az adatbázishoz,
- az adatkezelési tervet érdemes a kutatás során időnként újra elővenni¹⁶ és szükség szerint updatelni, hiszen változhatnak szempontjaink, új innovációs policyk, infokommunikációs technológiai elvárások, infrastruktúrák jelenhetnek meg.

¹¹ A **metaadat** „adat az adatról”, egy dokumentum vagy adatbázis esetében a legfontosabb leíró tulajdonságokat tartalmazza, szerző, kiadás éve, stb. Ezek az adatok nagymértékben segíthetik a dokumentumok, adatok megtalálását.

¹² A **Dublin Core Metadata Initiative** egy nemzetközileg elfogadott eljárás az on-line elérhető dokumentumok metaadatainak szabványosítására, és amely egy digitális „könyvtári kártya katalógus” létrehozásával megkönnyíti a keresőmotorok számára, hogy elérjék a dokumentumokat.

¹³ **DataCite** egy non-profit szervezet, amely célja, hogy tagjai számára DOI vagy más azonosítók használata révén elősegítse a kutatási adatok, eredmények on-line elérhetőségét, megtalálhatóságát.

¹⁴ A **CrossRef** ügynökségnél regisztrált **DOI** (*Digital Object Identifier*) egy egyedi azonosító, amely segíti az on-line formátumban megjelenő tudományos közlemények elérését.

¹⁵ Az **ORCID** (Open Research and Contributor Identifier) egy nemzetközi szerző-azonosító kód, amely a DOI azonosítók alapján összegyűjti a kutatók publikációit és segít a kutató beazonosításában.

¹⁶ Ehhez érdemes egy ütemtervet (time-table) is rögzítenünk az Adatkezelési tervben.