

Kutatások archiválása Adatmenedzsment Open science és adatok

Gárdos Judit és Meiszterics Enikő
Társadalomtudományi Kutatóközpont
2022 február

https://www.youtube.com/watch?v=66oNv_DJuPc

Mi a KDK, hol és hogy érhető el?

- A KDK – a Társadalomtudományi Kutatóközpont Kutatási Dokumentációs Központja - intézményi repozitórium, 2013 óta működik.
- A TK 4 intézetének kutatási anyagait gyűjti és teszi elérhetővé.
- Honlapja a kdk.tk.hu, repozitóriuma az openarchive.tk.mta.hu oldalon érhető el.

Milyen kutatási anyagok, milyen módon és kik számára érhetőek el a KDK-ban?

- A KDK-ban kvalitatív és kvantitatív módszerekkel készült kutatások anyagai egyaránt megtalálhatóak.
- A repozitóriumban sokféle kutatási adat (interjúk felvételei, leiratai, vezérfonalai; kérdőíves felmérések kérdőívei, módszertani leírásai, adatbázisai; terepnaplók, megfigyelések jegyzőkönyvei stb.) érhető el, különböző (szöveg, kép, videó, archiv stb.) formátumokban.
- A ***metaadatokhoz*** minden esetben bárki szabadon hozzáférhet.

- A ***letétbe helyezett kutatási adatokhoz*** a letétbe helyezővel kötött, a Letéti Nyilatkozatban rögzített megállapodásnak megfelelően
 - bárki szabadon hozzáférhet,
 - a TK kutatói szabadon, külsős kutatók egy email-es regisztráció után férhetnek hozzá,
 - az adminisztrátorokon kívül mindenki más csak kutatói jóváhagyással férhet hozzá.

Mi a 20. Század Hangja Archívum és Kutatóműhely, és hol és hogy érhető el?

- A 20. Század Hangja Archívum és Kutatóműhely (Voices) a Magyar Tudományos Akadémia Szociológiai Kutatóintézetében alakult meg 2009 tavaszán.
- Egy évtizeden át OTKA támogatással működött, 2019 tavasza óta a KDK-val közös szervezeti egységben dolgozik.
- Az archívum a 20szazadhangja.tk.hu oldalon érhető el.

Szeretnénk megosztani az adatainkat?

- Nem: titkos adatok / nem elég jó adatok / csak mi szeretnénk felhasználni
- Igen: Közös publikációk; kutatások terjesztése; pénz- és időspórolás; nemzetközi kapcsolatok
- Igen: Fontos anyagok megmentése

Ha szeretnénk: Megtalálják, amit keresnek?

- Hol tároljuk?
- Milyen nyelven?
- Milyen információkat adunk meg róla?
- Hogyan érjük el, hogy minél többen lássák?

Kötelező adatmegosztás

- Számos nemzetközi természettudományos lapnál
- Számos nemzetközi társadalomtudományos lapnál
- Számos külföldi/nemzetközi kutatásfinanszírozónál
- OTKA-nál is?

- **Biztonságos tárolás**
 - Ne csak egy helyen legyen
 - Hosszú távon legyen elérhető
 - Más is tudja, hol van (és mi is)
 - Verziókontroll (melyik kutatótárs mikor, hova, mit ment)
- **Megbízható tárolás és mások számára látható**
 - Digital Object Identifier (DOI) – adatokra is, cikkekre is

Nyílt tudomány (Open Science)

- A nyílt tudomány arra törekszik, hogy az általa előállított tudományos kutatásokat és adatokat mindenki számára elérhetővé tegye.
- Egy olyan mozgalom, amelyben a kutatások együttműködésen alapulnak, átláthatóak és elérhetőek.
- Kulcsterülete a kutatási adatok kezelése.

FAIR irányelvek

A FAIR kutatási adatok

Findable (meg tudjuk keresni/találni)

Accessible (hozzá tudunk férni, le tudjuk tölteni)

Interoperable (meg tudjuk érteni) és

Reusable (fel tudjuk használni)

Ugyanakkor az adatok nyilvánosságra hozásának vannak korlátai is (pl. személyes adatok védelme, védett madarak fészkelési helyének vagy régészeti objektumok pontos koordinátáinak megóvása), így a kutatási adatkezelés javasolt elve a „nyílt, amennyire lehetséges; zárt, amennyire szükséges”.

RDA (Research Data Alliance)

- Az RDA-t egy közösségi kezdeményezésként indította útjára 2013-ban a Európai Bizottság, az Egyesült Államok Nemzeti Tudományos Alapítványa és az ausztrál kormány innovációs minisztériuma.
- Célja az adatok nyílt megosztását és újrafelhasználását lehetővé tevő társadalmi és technikai infrastruktúra kiépítése.
- A tevékenységében való részvétel mindenki számára nyitott, 2020-ban már 145 országból voltak tagjai.
- 2019-ben megalakult a globális szervezet magyar tagja, a Research Data Alliance Hungarian National Node (HRDA), tagjai között van többek között az ELKH, a TK, a Wigner, a SZTAKI.

European Open Science Cloud (EOSC)

- Az EOSC nem egy klasszikus felhő, hanem egy európai szintű együttműködés, ezt tükrözi a magyar neve, az Európai Nyílt Tudományos Együttműködés is.
- Missziója egy közös, egyesített, európai keretrendszer kialakítása a kutatási adatok nyílt megosztása és a szolgáltatások elérése céljából.

Az NKFIH és a nyílt tudomány

- Az NKFIH célja az Európai Bizottság ajánlásainak megfelelő és a nemzetközi kezdeményezésekhez illeszkedő magyar Open Science infrastruktúra támogatása
- 2021. október 18-án kezdeményező szervezetként az NKFIH kiadott egy „Állásfoglalást a nyílt tudományról”, amelyhez aláírásukkal csatlakozhatnak a támogató intézmények.
- Először az 2020-ban a nyertes Élvonal pályázatoknál kérték adatkezelési terv készítését, majd a 2021-es az OTKA pályázatok dokumentációja is bővült egy új adatlappal, adatkezelési tervvel.

Mi is az a kutatási adatkezelési terv?

A kutatási adatkezelési terv egy rövid, összefoglaló dokumentáció arról, hogy

- milyen adatok és hogyan keletkeznek az adott kutatásban,
- ezeket hogyan tárolják és ki férhet hozzájuk a kutatás során, illetve
- a kutatás lezárulta után.

A kutatási adatkezelési terv változhat a kutatás közben, ilyenkor - a különböző verziókat megőrizve - frissíteni kell.

Iránymutatás, javaslatok az 2022-es kutatásiadat-kezelési terv kitöltéséhez

2022-ben a kutatásiadat-kezelési terv vonatkozásában az előző évhez képest nem változott az OTKA pályázatok pályázati kiírása

- ugyanaz az adatkezelési terv, mint 2021-ben és
- akkor kell (pdf formátumban) feltölteni, amennyiben a projekt megvalósítás során várhatóan keletkeznek a kutatási adatok.

Ebben az esetben azt kéri a pályázótól, hogy

„erről az „Előlap” menüpont „Kutatási adatok és infrastruktúra” pontja alatt nyilatkozzon, és röviden foglalja össze a keletkező adatok jellegét, mennyiségét, az esetleges (pl. etikai vagy az adatok személyes jellegéből adódó) kockázatok tervezett kezelését, az adatok hosszú távú megőrzésére, hozzáférhetőségére, további felhasználhatóságára vonatkozó terveit, valamint a pályázatához töltsse fel a kitöltött Kutatási adatkezelési tervet. A Kutatási Adatkezelési terv formanyomtatványa a PÁLYÁZATI CSOMAG részeként letölthető az NKFI Hivatal honlapjáról.”

Keletkeznek-e adatok a kutatás során?

Milyen típusú adatok keletkeznek a kutatás során?

A kutatási adat meghatározása: ¹

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

Létrejöhetnek megfigyelések, kísérletek, szimulációk eredményeképpen – egy konkrét kutatás számára előállítva – vagy korábban gyűjtött adatok összegyűjtésével, válogatásával, feldolgozásával.

Példák kutatási adattípusokra

- adatbázis-tartalmak (videó, hang, szöveg, képek)
- alkalmazások működési adatai (naplófájlok tartalma, alkalmazások futási eredményei,
- szimulációk eredményei)
- audio/video anyagok, fotók, filmek, diapozitívok, hangfelvételek
- dokumentumok (szöveges vagy táblázatos formátumban)
- fehérje vagy genetikai szekvenciák
- kérdőívek, szövegátiratok, kódkönyvek
- kérdőíves felmérések válaszai, teszteredmények
- kutatási folyamat során megszerzett és/vagy előállított digitális tárgyak
- laboratóriumi jegyzetek, terepgyakorlati naplók, feljegyzések
- modellek, algoritmusok, szkriptek
- módszerek és munkafolyamatok
- szabványos működési eljárások és protokollok
- színeképek, színeképelemzések stb.

Társadalomtudományos kutatásokban a leggyakrabban a következő kutatási adatok keletkeznek:

kérdőív, adatfájl, kódkönyv, interjú vezérfonal, hangzó interjú, interjú leirat, terepnapló, módszerek és munkafolyamatok leírása, fénykép, film, alkalmazások működési leírása, modellek, algoritmusok

„Ki felel az adatkezelésért?”

Minden kutatásban ki kell nevezni egy adatfelelőst.

„Milyen eljárással végzik az adatgyűjtést?”

Nem csak kutatási adat van sokféle, hanem eljárás is, amelyekkel ezeket az adatokat összegyűjtik. Jellemző megoldások a(z)

- mérés,
- megfigyelés,
- elemzés, másodelemzés,
- esettanulmány,
- kísérlet,
- kérdezés,
- tesztelés.

A társadalomtudományok területén jellemző eljárás az interjúzás, csoportos interjúzás, terepmunka, megfigyelés, kérdőíves felvétel, másodelemzés, szövegelemzés, képelemzés, kísérlet, big data elemzés.

„Hol és hogyan tárolják az adatokat, dokumentációkat és gépi kódokat a kutatás során?”

A kutatók a kutatási adatokat a kutatás során több helyen (akár párhuzamosan is) tárolhatják, ilyenek a(z)

- személyes helyi adattárolási megoldások (pl. saját számítógép, pendrive, külső merevlemez)
- intézményi helyi adattárolási megoldások (pl. intézményi szerver, pendrive, külső merevlemez)
- személyes felhőalapú tárolási szolgáltatások (pl. OneDrive, google account)
- intézményi felhőalapú tárolási szolgáltatások (pl. intézeti felhő, google account)
- személyes weboldal
- intézeti/projekt weboldal
- Git típusú repozitóriumok (pl. GitHub)
- kutatási adattárak, adatbázisok (pl. OpenBioMaps)

Az adatokat már a kutatás során is lehet repozitóriumban tárolni, de a repozitóriumok elsősorban inkább a kutatás lezárulta utáni adattárolásra színhelyei.

„Milyen hozzáférést biztosítanak ezekhez az adatokhoz?”

A kutatás során jellemzően vagy maga a kutató vagy a kutatás résztvevői (intézményen belülről vagy kívülről) férnek hozzá az adatokhoz, de van, amikor már ebben a fázisban is szélesebb körnek adnak hozzáférést.

Nemcsak a hozzáférők körét, hanem a hozzáférés szintjét is lehet szabályozni, jogosultság beállításával (csak olvashat; olvashat és szerkeszthet; olvashat, szerkeszthet és feltölthet is pl.).

A hozzáférés beállítását meghatározhatják egyéb szempontok is (pl. GDPR, finanszírozó elvárásai), elképzelhető pl., hogy egy kutatás személyes adataihoz szűkebb kör fér hozzá, mint a kutatásban keletkező egyéb adatokhoz, de akár az is, hogy a finanszírozó már a kutatás során is repozitóriumi elhelyezést és hozzáférést kér.

A TK-s kutatások során személyes adatokat csak a TK-s felhőben lehet tárolni, egyéb kutatási adatokat, dokumentációkat a felhőben, a kutatók jelszóval védett munkahelyi és saját számítógépein, az ELKH felhőben vagy a Github-on, úgy, hogy az arra jogosultak férjenek hozzájuk.

„Hol és hogyan tárolják az adatokat, dokumentációkat és gépi kódokat a kutatás lezárulta után?”

A kutatás lezárulta utáni tárolás történhet a kutatás során történő tárolásnál már felsorolt tárhelyeken, illetve – elsősorban ekkor - repositóriumokban is.

Az adatrepositórium választásról nem mindig a kutató dönt, van amikor a projekt finanszírozója vagy a publikáció kiadójának előírása határozza meg, hogy melyik repositóriumba lehet az adatokat feltölteni.

Ha a kutató dönt, a repozitórium választásnál többféle szempontot lehet és kell is mérlegelnie, ilyenek pl.

- hazai vagy nemzetközi, illetve
- általános vagy témaspecifikus repozitóriumban szeretnék az adatainkat elhelyezni,
- megfelel-e a konkrét adatigénynek, illetve
- ingyenes vagy fizetős szolgáltatást nyújt az adott repozitórium.

Az <https://openscience.hu/> oldalon részletes információ található repozitóriumokról és kifejezetten adatrepozitóriumokról, tudományterületi bontás szerinti is.

A hazai repozitóriumok közül jó megoldás lehet az ELKH felhőjén futó Concorda, illetve később az ARP projektben megvalósuló repozitórium.

<https://science-data.hu/>

Csekély mennyiségű kutatási adat esetében előfordulhat az is, hogy az adatok a kutatási dokumentációban és a megjelent publikációkban érhetőek el. Ilyenkor a kutatási dokumentáció megőrzéséről illetve a publikáció elérhetőségéről kell nyilatkozni.

Intézményi szinten is előfordulnak vonatkozó irányelvek, előírások, van, ahol az egy kutatáshoz tartozó adatok, dokumentációk, gépi kódok összességéből létrejövő adatcsomag egy példányát intézményi tárhelyen, adattárban vagy repozitóriumban (is) el kell helyezni.

A TK-ban a kutatás lezárulta után az adatokat a TK-s felhőben is archiválni kell, és a 2021. január után induló kutatások adatait fel kell ajánlani a KDK repozitóriumban történő elhelyezésre. A KDK és a letétbe helyező kutató az összes felmerülő szempontot figyelembe véve közösen dönt az elhelyezésről és annak feltételeiről.

„Milyen hozzáférést biztosítanak ezekhez az adatokhoz?”

A kutatási adatokhoz

- egyéni érdeklődők számára, külön kérésre
- kijelölt személy, személyek számára
- kijelölt intézmények vagy kutatócsoportok tagjai számára
- publikációt elbírálók számára
- konkrét publikációt olvasók számára
- kutatási projekt tagjai számára
- kutatási szövetség, konzorcium tagjai számára
- adott tudományterület dedikált megosztási platformjának használói számára vagy
- teljesen nyíltan, általános platformon bárki számára adhatunk hozzáférést.

- Elképzelhető az is, hogy a hozzáférés a kutatás lezárása óta eltelt idő függvényében válik egyre inkább nyitottá.
- A repozitóriumokban kérhető egy embargós időszak is, amikor az adatok ott vannak már a repozitóriumban, láthatóak, de még nem hozzáférhetőek.
- A KDK-ban három hozzáférési szint - szabad hozzáférés, kutatói hozzáférés (ami a TK-s kutatóknak szabad, a külsős kutatóknak igénylés után szabad hozzáférést jelent) és zárt (csak egyedi engedéllyel) – van.
- A kutatás honlapján történő elhelyezésnél a hozzáférés lehet nyílt (pl. ESS Magyarország) vagy regisztrációhoz kötött (PTI CAP).

„Mely adatokat fogják nyíltan elérhetővé tenni?”

Abban az esetben, ha a projektben nincs erre vonatkozó (pl. finanszírozói, intézményi) előírás, akkor ez kutatói mérlegelés és – a korlátozó tényezők figyelembe vételével meghozott - döntés függvénye. Itt a korábban már említett „nyílt, amennyire lehetséges; zárt, amennyire szükséges” megfontolást érdemes figyelembe venni.

„Szükség van-e az adatmegosztás korlátozására?”
Amennyiben igen, célszerű röviden indokolni.

„Milyen metaadat szabványt használnak?”

Előfordulhat, hogy egy kutatás adatairól

- nem készül metaadat leírás, vagy
- készül ugyan, de nem szabvány szerint.

Az is lehet, hogy egy mérés metaadatait automatikusan létrehozza, esetleg tárolja is maga a méréshez használt műszer vagy a mérési adatokat tartalmazó fájl.

Amennyiben használnak sémát a metaadatok leírásra, akkor az lehet

- saját séma,
- kutatócsoport által kialakított séma
- intézményi séma,
- műszerek és programok saját szabványa is.

Az egyik legáltalánosabb leíró séma a DublinCore, a KDK-ban is ezt használjuk, de pl. a biológia területén elterjedt a már a DarwinCore.

„Hol és hogyan tárolják a metaadatokat a kutatás során?”

- A metaadatokat a kutatás során lehet a kutatási adatokkal egy helyen (ld. Korábban felsorolt tárolási megoldások) vagy azoktól elkülönítve is tárolni.
- Előfordulhat az is, hogy a metaadat leírást/leírásokat nem a kutató vagy a kutatócsoport valamelyik tagja, hanem egy külön erre kijelölt személy (pl. intézményi könyvtáros vagy data steward) készíti és egy, a kutatási adatokétól eltérő tárhelyen tárolja.

„Milyen hozzáférést biztosítanak a metaadatokhoz?”

- Általánosságban elmondható, hogy a metaadatokhoz nyíltabb a hozzáférés, mint magukhoz a kutatási adatokhoz.
- Ugyanakkor a tárolásuk helye is befolyásolja vagy meghatározza a hozzáférhetőségüket, ha a kutató gépén vannak, akkor kevésbé, ha egy honlapon, adatbázisban, repo-ban fellelhetőek, akkor inkább hozzáférhetőek.

A TK-ban a KDK-ban és az intézeti honlapokon, oldalakon található metaadatok nyílt hozzáférésűek.

„Használnak-e állandó és egyedi azonosítókat (DOI, ORCID)?”

Az állandó és egyedi azonosítók közül dokumentumok és adatcsomagok, gyűjtemények vonatkozásában jelenleg a DOI (digital object identifier), szerzőkre az ORCID a leginkább elfogadott, elterjedt.

A KDK-ban elhelyezett gyűjteményekhez a Datacite Ügynökségnél regisztrálunk DOI-t az MTA Könyvtár és Információs Központ DOI irodáján keresztül.
ORCID-ot a kutatók tudnak maguknak igényelni.

„Hol és hogyan tárolják a metaadatokat, dokumentációkat és gépi kódokat a kutatás lezárulta után?”

A kutatás lezárulta után a metaadatokat célszerű a kutatási adatokkal együtt, egy adatcsomagban tárolni, a korábban már felsorolt lehetséges tárolási helyeken.

A TK-ban az archivált kutatási adatokkal együtt a TK felhőben, ha a kutatásnak van honlapja, akkor – legalább is egy részüket - ott is, ha van gyűjteménye a KDK-ban, akkor ott is.

„A kutatási adatok kezelése megfelel-e a GDPR előírásainak?”

- A GDPR az egyik olyan szempont, amire az adatkezelés vonatkozásában a kutatás során végig tekintettel kell lenni.
- A kutatási adatkezelési terv készítése egy jó lehetőség annak átgondolására, hogy a projekt során esetleg felmerülő személyes adatokkal mit és hogyan tegyünk (pl. beleegyezési nyilatkozat kérése, benne az esetleges hivatkozások megoldásának meghatározásával; anonimizálás szükségességének, megoldásának átgondolása).
- A GDPR elvárásainak az adatkezelési megoldásoknak meg kell felelni, a „nem” itt nem túl jó válasz.

A TK-ban a GDPR előírásainak megfelelően történnek a kutatások.

Kutatói igényeket felmérő kérdőív az ARP projektben

Építsünk közösen kutatási adatarchívumot!

A mai eseményre szóló meghívóval együtt ment ki a kérdőív linkjét tartalmazó levél, illetve elérhető a KDK oldaláról is:
<https://kdk.tk.hu/>

Kérjük, hogy a kérdőív kitöltésével támogassa a projekt sikeres megvalósítását!

Ha szeretne többet megtudni a projektről, vagy szívesen venne részt a megvalósításban, írjon a kdk@tk.hu címre!

Köszönjük a figyelmet!