

Intézményi, támogatói szabályozások kialakítása a nyílt tudomány kutatási adatkezelési céljaihoz

Holl András
MTA Könyvtár és Információs Központ
holl.andras@konyvtar.mta.hu
ORCID: [0000-0002-6873-3425](https://orcid.org/0000-0002-6873-3425)

A nyílt tudomány (Open Science) kulcsterülete a kutatási adatok kezelése. Az immár a kutatási gyakorlat részévé vált nyílt hozzáférés (Open Access) tekintetében léteznek hazai szabályozások, továbbá rengeteg nemzetközi ajánlás is segíti ezek finomítását, bővítését, azonban a kutatási adatkezelés területén még előttünk áll a szabályozások kialakítása.

Kulcsszavak: nyílt tudomány, kutatási adatok, FAIR, szabályozások

0.) A FAIR kritériumrendszer

A kutatási adatkezelés ma elfogadott paradigmáját a FAIR mozaikszóval jellemzik: Findable (megtalálható), Accessible (hozzáférhető), Interoperable (szabványos) és Reusable (újrafelhasználható)¹. A kutatási adatok nyilvánosságának korlátait felismerve (pl. személyes adatok védelme) a kutatási adatkezelés javasolt megközelítése a „nyílt, amennyire lehetséges; zárt, amennyire szükséges”.

A jól dokumentált, rigorózus adatkezelésen és archiváláson túl két alapvető oka van a kutatási adatok FAIR kezelésének: egyfelől égetően szükség van a kutatási eredmények reprodukálhatóságára, másfelől az adatok újra felhasználásával időt és pénzt takaríthatunk meg.

Két kulcs-elemet emelünk ki az adatok megfelelő kezeléséhez: az egyedi azonosítók (DOI, ORCID, projekt- és intézményazonosítók), valamint a széles körben elfogadott sémák, szótárak, ontológiák használatát.

1.) Kötelezettség szabályozások kialakítására

Jól működő kutatási adatkezelési gyakorlat csak megfelelő szabályozások mellett képzelhető el. Mivel minden tudományterület képviselői jelentősen eltérő adatokkal dolgoznak, a magasabb szintű – országos, egyetemi, kutatóhálózati, finanszírozói – szabályozásoknak szükségszerűen általánosnak kell lenniük: gyakran csak deklarálják a FAIR alapelvek betartásának szükségességét, az említett „nyílt, amennyire lehetséges...” alapelveket és előírják az alacsonyabb szintű szabályozások (angolul mandate) elkészítését. Ilyen lehet az intézményi, kutatócsoportos szabályozások vagy az adatkezelési tervek megkövetelése. A tudományos folyóiratok is szabályozhatják – és egyre gyakrabban szabályozzák is – a publikált kutatásokat alátámasztó adatok kezelését.

¹ FAIR: <https://www.force11.org/group/fairgroup/fairprinciples>



Minden szabályozási szinten érdemes figyelembe venni a releváns nemzetközi mandátumokat (H2020², ERC³). Az európai országos szintű szabályozásokat a SPARC Europe és a Digital Curation Centre anyaga mutatja be (2020). Magyarországon az NKFIH Open Science Szakértői Bizottsága⁴ készített ajánlást a szabályozásra. Az adatok nyílt elérhetőségének biztosítására ajánlani kell minősített repozitóriumok használatát. Amennyiben a szabályozást készítő szervezetnek létezik nyílt hozzáférésre vonatkozó szabályozása, a kutatási adatkezelésre vonatkozó szabályozás beépíthető ebbe, vagy egységes nyílt tudomány szabályozás készítése javasolt.

2.) Adatkezelési tervek

Adatkezelési tervek témakörben több nemzetközi minta, ajánlás⁵ szakcikk és útmutató is található (Michener, 2015; Jones, 2011). Magyarországon az NKFIH pályázatok kiírói megkövetelik a nyertes pályázatokhoz az adatkezelési tervek készítését⁶.

Az adatkezelési terv élő dokumentum, amit a projekt során karban kell tartani, és a projekt befejezése után csatolni kell a projekt beszámolóhoz a keletkezett adatok listájával és az elérésükre is használható azonosítóval (DOI).

Az adatkezelési tervnek ki kell térnie az alábbiakra:

- adatok jellege, formátuma, mennyisége;
- adatok leírása;
- alkalmazott adatfeldolgozási eljárások;
- hosszú távú megőrzés, egyedi azonosítók használata;
- kockázatok (etikai, személyes adatok kezeléséből adódó, stb.);
- hozzáférhetőség, felhasználási licenc;
- adatkezelés költségei;
- adatok felhasználhatósága, esetleges értékesíthetősége;
- nyilvános, projekten kívül keletkezett adatok használata;
- ... és minden egyéb releváns körülményre.

Az adatkezelési terv alternatívája a szakterületi adatkezelési protokoll (Domain Data Protocol, DDP)⁷. Az intézményi szabályozások hivatkozhatnak a DDP-re, vagy – a kutatócsoportok munkáját megkönnyítendő – adhatnak intézményi alap/keret adatkezelési tervet is.

2 H2020: <https://www.openaire.eu/how-to-comply-to-h2020-mandates-for-data>

3 ERC: https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf

4 NKFIH: <https://nkfi.gov.hu/hivatalrol/open-science-szakertoi/open-science-szakertoi>

5 H2020: DOI: [10.5281/zenodo.2635768](https://doi.org/10.5281/zenodo.2635768); ERC: <https://erc.europa.eu/content/erc-data-management-plan-template>

6 OTKA: <https://nkfi.gov.hu/palyazoknak/nkfi-alap/kutatasi-temapalyazat-k21/palyazati-csomag/kutatasiadat-kezesi-terv-k21-fk21-pd21-ann21-snn21>

7 DDP: https://www.scienceurope.org/media/nsxdyqvn/se_guidance_document_rdmps.pdf

Az adatkezelési tervek szabályozásához javasoljuk – az amúgy igen különböző filozófiájú – ERC és HorizontEurope adatkezelési terv űrlapok⁸ tanulmányozását, de léteznek online eszközök is, amik az adatkezelési terv elkészítésében jelentős segítséget nyújtanak⁹.

3.) Publikus adatok hivatkozása, jelentése, nyilvántartása

A publikus adatok hivatkozása, jelentése, nyilvántartása tekintetében az eddigieknél jóval kevesebb támpontot találhatunk. Alapvetés, hogy a kutatási adatokat el kell látni állandó azonosítóval (pl. DOI), a nyilvános adatok felhasználásakor pedig a szükséges hivatkozásoknak az adatállomány azonosítóját is tartalmaznia kell. Amennyiben az adatokra alapuló publikációt közlő folyóiratnak van erre vonatkozó előírása, azt kell követni, egyéb esetben a cikkben hivatkozni kell a más kutatásokból, adatbázisokból felhasznált adatokra, amennyiben van, DOI azonosítóval. Ha a kutatáshoz felhasznált adatok egy dinamikus frissülő publikus adatbázisból való SQL lekérdezéssel keletkeztek, meg lehet fontolni a lekérdezett adatok archiválását és DOI-val való ellátását. Ez a kutatás reprodukálhatósága miatt is lényeges.

A saját kutatási adatoknál hivatkozni kell a nyilvános repozitóriumban, adatbankban elhelyezett adatok DOI azonosítójára. Mind a saját, mind az újra felhasznált adatoknál javasoljuk a hivatkozások elhelyezését a cikk irodalomjegyzékében, ahonnan a hivatkozáskapcsolatok gyűjtése könnyebben megvalósítható, mint a szövegtestből, ráadásul az irodalomjegyzékek kezelése sok adatbázisban automatikus. Más kutatók adatainak felhasználásánál általános gyakorlat az adatokra épülő cikkekre való hivatkozás, pedig a közvetlen adathivatkozások használatának vannak előnyei.

A kutatási jelentésekben a publikációk mellett a közzétett kutatási adatok elérhetőségét is fel kell sorolni a DOI azonosítókkal együtt. A felsorolást hazai beszámolási kötelezettség esetén célszerű az MTMT-ben tárolt adatokra való hivatkozással megtenni, amire az MTMT lehetőséget is ad: a publikus kutatási adatok felvihetők a rájuk kapott hivatkozásokkal együtt, majd az egyéni kutatók összesítő táblázatában ezek az adatok is megjelennek.

4.) Tudományértékelési szerep, mutatók

Amíg a kutatási adatok (általában a publikációkon túlmenő egyéb, nyilvánosságra hozott kutatási termékek) mennyisége és hatásmutatói (hivatkozások) nem számítanak a tudományos tevékenység értékelésében, addig a kutatók nem lesznek ösztönözve ezen termékek (esetünkben az adatok) nyilvánosságra hozatalára (bár a finanszírozók és a folyóiratok esetleges ilyen irányú követelményei segíthetnek). Mind a kutatási alapoknak, mind a folyóiratoknak célszerű a kutatási adatok kezelésének kérdését szabályozniuk, hozzájárulva a publikált kutatási adatok és a rájuk kapott hivatkozások értékének megalapozásához.

8 ERC: <https://erc.europa.eu/content/erc-data-management-plan-template>
H2020: https://zenodo.org/record/2635768#.YEsVEP4o_0o

9 Mint a Digital Curation Centre által készített DMPTool: <https://dmptool.org/>



Célszerű mind egyéni, projekt és intézményi szintű értékeléseknél figyelembe venni az MTMT-ből kinyerhető teljesítménymutatókat, továbbá a kutatási adatokra vonatkozó hivatkozásokat ugyanúgy értékelni, mint a publikációkét. Az értékelési rendszerről az Európai Bizottság dokumentumot készített (2017).

5.) Infrastruktúrák biztosítása

A kutatási adatok megőrzése és szolgáltatása hosszú távra biztosítandó feladat. Infrastruktúrákat és szervezeti egységeket kell létrehozni, amik hosszú távú működését megfelelő szabályozással biztosítani kell. A létrejött adatrepozitóriumokat el kell helyezni az intézményi struktúrában és költségvetésben. Működésüket szabályozni kell: a kutatók csak akkor tudják a FAIR követelményeket teljesíteni, amennyiben ennek megvannak az intézményes feltételei, amik feltételei nyilvánosak, szabályozottak. Az adatrepozitórium használata szerződés létrejöttével jár a kutató és a repozitóriumot működtető intézmény között, még ha konkrét aláírás nem is történik. A repozitóriumnak közzé kell tennie a megőrzésre vonatkozó vállalásait, az adatelhelyező számára kötelező nyilatkozatot, melyet az adatrepozitórium használatának kezdetekor el kell fogadni.

A repozitóriumok felé való bizalom megteremtéséhez/fenntartásához szükség van a repozitórium hosszú távú stabilitásának biztosítására, ami megjelenhet a fenntartó intézmény szabályozásaiban, pl. az SZMSZ-ben és egy külső, minőségbiztosító szervezet által kiadott tanúsítványban. Ilyen tanúsítványt Magyarországon az MTMT Repozitórium-minősítő bizottság¹⁰ ad, általánosan elfogadott nemzetközi tanúsítványt pedig a CoreTrustSeal¹¹.

6.) Kivételkezelés, embargók, fokozatosság

Minden szabályozás a kivételek, felmentések kezelésével lesz teljes. Lehetőséget kell teremteni a szabályozás alól való felmentésre indokolt esetben, megfelelő nyilvánossági embargók alkalmazását kell biztosítani, adott esetben a nemzetközi gyakorlattól eltérő türelmi időszakok megadásával, és a szabályozásokat fokozatosan kell bevezetni.

Köszönetnyilvánítás: a szerző hálás Harnos Noéminek és a tanulmány másik bírálójának a szöveg javításáért.

¹⁰ MTMT: <https://www.mtmt.hu/repozitoriumminosito-szabizottsag>

¹¹ CoreTrustSeal tanúsítvány: <https://www.coretrustseal.org/>

Irodalom

EC „Evaluation of Research Careers fully acknowledging Open Science Practices; Rewards, incentives and/or recognition for researchers practicing Open Science”, 2017, DOI: [10.2777/75255](https://doi.org/10.2777/75255)

Jones, Sarah „How to Develop a Data Management and Sharing Plan.” DCC, 2011. <https://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How%20to%20Develop.pdf>

Michener, William K. „Ten Simple Rules for Creating a Good Data Management Plan.” PLoS computational biology vol. 11,10 e1004525. 22 Oct. 2015, DOI: [10.1371/journal.pcbi.1004525](https://doi.org/10.1371/journal.pcbi.1004525)

SPARC Europe & DCC „An Analysis of Open Science Policies in Europe.” v6. DOI: [10.5281/zenodo.4005612](https://doi.org/10.5281/zenodo.4005612)

