

---

## Horváth Tamás – Kókay Péter

horvath.tamas@lib.pte.hu – kokay.peter@lib.pte.hu

### *„A kutatási adatok hatékony kezelése” Bevezetés a kutatási adatmenedzsmentbe*

---

#### **A tananyag címe**

„A kutatási adatok hatékony kezelése” – Bevezetés a kutatási adatmenedzsmentbe

#### **Cél**

A résztvevőket megismertetni az alapvető kutatási adatkezelési fogalmakkal, eljárásokkal. A résztvevők ismerjék az alapvető adatkezeléssel kapcsolatos eszközöket és azokat az apparátusokat, melyekhez a későbbiekben segítségért fordulhatnak.

#### **A megszerezhető kompetenciák**

A résztvevő a tananyag elvégzése után képessé válik az alábbiakra:

- ismeri és érti a FAIR kutatási irányelveket és alkalmazni is tudja őket a gyakorlatban,
- ismeri és tudja, mi az a kutatási adatkezelési terv és hogyan lehet alkalmazni ilyet a gyakorlatban,
- ismeri és tudja, mit jelentenek az Open Data, Open Science fogalmak egy kutató számára,
- ismeri a legjobb adatkezelési gyakorlatokat,
- ismeri és képes alkalmazni a számára releváns kutatási infrastruktúrát.

#### **A célközönség meghatározása**

Elsősorban PhD-s hallgatók és kutatók számára készül

---

\* A szövegben előforduló összes internetes hivatkozás letöltési és ellenőrzési dátuma 2021. május 31.

## **A tematika felépítése**

- 1. Definíciók, kifejezések és alapok**
  - a. Mi a kutatási adat? (A kutatási adat meghatározása, típusai)
  - b. Kutatási adat élelciklus
  - c. Mi az a kutatási adatmenedzsment (research data management) és miért van rá szükség?
  - d. FAIR adatok – FAIR adatkezelés, irányelvek
  
- 2. Tervezés, rendszerezés és adatkeresés**
  - a. A kutatás megtervezése
  - b. Kutatási adatkezelési terv (Data management Plan, DMP Tools, DMP példák, sablonok)
  - c. Adatgyűjtés – már meglévő adatok felkutatása
  - d. Metaadatok szerepe, fontossága
  - e. Adatszervezés, adatbiztonság
  
- 3. Elérés, közzététel és archiválás**
  - a. Hozzáférési lehetőségek
  - b. Adatok publikálása (adatok közzétételének előnyei)
  - c. Azonosítók szerepe (dokumentum- és szerzői azonosítók)
  - d. Adatrepozitóriumok
  - e. Re3data
  - f. Zenodo, Figshare, Dryad, Concorda
  - g. Hosszú távú megőrzés
  
- 4. Licenc és jogi kérdések**
  
- 5. Egyebek**

## 1. DEFINÍCIÓK, KIFEJEZÉSEK ÉS ALAPOK

### Bevezető gondolatok

A tananyag első fejezete a kutatási adatkezeléssel kapcsolatos főbb fogalmakat igyekszik összegyűjteni és tisztázni.

### Célok, megszerezhető kompetenciák

A résztvevő megismeri a kutatási adat fogalmát, illetve a kapcsolódó definíciókat, a kutatási adatok életciklusát, az adatmenedzsmenttel kapcsolatos fontosabb tudnivalókat és a FAIR irányelveket.

### Témakörök

- Mi a kutatási adat? (a kutatási adat meghatározása, típusai)
- Kutatási adat életciklus
- Mi az a kutatási adatmenedzsment és miért van rá szükség? (research data management)
- FAIR adatok – FAIR adatkezelés, irányelvek

### Irodalmak

- Research data – CESSDA TRAINING – <https://bit.ly/2Sou2mo>
- <https://openscience.hu/kutatasi-adatok>
- <https://instantscience.hu>
- <https://www.force11.org/group/fairgroup/fairprinciples>
- <https://www.dcc.ac.uk/guidance/how-guides/develop-data-plan>
- <https://www.go-fair.org/fair-principles/fairification-process>

#### *1.1. Mi a kutatási adat?*

##### *(A kutatási adat meghatározása, típusai)*

Mi is az adat? Az adatot többféle módon lehet meghatározni a tudományágtól és a kontextustól függően, de általánosságban azt mondhatjuk, nem más, mint rögzített és tárolt információk összessége.

##### *Mennyiben különbözik ettől a kutatási adat fogalma?*

A kutatási adatok a tudományos kutatási tevékenység során keletkező és rögzítésre kerülő tények, melyek általában szimulációk, kísérletek, megfigyelések során újonnan jönnek létre. Ezzel egyidejűleg kelet-

kezhetnek korábbi adatgyűjtés(ek)ből származó adatok szelektálásából, újrafeldolgozásból. A rögzített adatok igen sokfélék lehetnek, ha azt vesszük figyelembe, hogy milyen formátumban tároljuk őket. Alapvetően szöveges formában tárolt anyagok juthatnak rögvest eszünkbe:

- szöveges dokumentumok,
- táblázatok,
- adatbázisokból származó adatok,
- kérdőíves felmérések válaszai,
- jegyzetek és naplófájlok,
- gyakorlati- és terepnaplók,
- működési protokollok,
- munkafolyamatok,
- stb.

Ezek mellett nem szabad megfeledkeznünk a valamelyest rendezhetőbb formátumú kutatási adattípusokról sem:

- multimédiás formátumok: videó, hang, szöveg, képek, diaképek,
- színeképek és elemzések,
- a kutatás során előállított digitális objektumok, modellek,
- futtatási eredmények,
- metaadatok,
- stb.

A kutatási adatok keletkezésének körülményeit tekintve a következőkről beszélhetünk:

- *megfigyelésen alapuló adatok*: forrása a kutatók által végzett felmérések, illetve az éppen vizsgált tevékenységek és viselkedések megfigyelő munkája. Hangsúlyozandó, hogy a kutatók a megfigyelő tevékenység során nem avatkoznak be egyetlen folyamatba sem;
- *kísérletek során előálló adatok*: az előbbiekkal ellentétben a kutatók aktív munkája révén állnak elő az adatok, hiszen a kísérletek paraméterei gyakran változtatásra kerülnek annak érdekében, hogy összehasonlítható eredmények keletkezzenek;
- *szimulációk során keletkező adatok*: olyan rendszerek modellezése során keletkező adatok, melyek valós folyamatokat próbálnak leírni számítógépes környezet segítségével. A kí-

sérletekhez képest inkább egy prediktív jelleggel bíró eljárás során rögzített adatok összessége;

- *meglévő adatforrásokból keletkeztetett adatok*: korábban keletkezett adatforrásból gyűjtött, újrafelhasználás céljából válogatott adatok összessége.

Az egyes típusok más-más reprodukálási lehetőségekkel bírnak, hisz egy-egy kísérletet sokkal könnyebben reprodukálhatunk, mint emberi csoport(ok) megfigyelésén alapuló, több tényező együttes hatására kialakult reakciófolyamatot.

## 1.2. Kutatási adatok életrajza

A kutatási adatok élettartama hosszabb, mint a létrehozásuk pillanata. Az átláthatóság egyik módja az úgynevezett kutatási életrajz. De mi is ez az életrajz?

Az életrajz modell betekintést nyújt abba, hogy a kutatás életszakaszai és a kutatási adatok hogyan illeszkednek egymáshoz, és az egyik fázisban hozott döntések hogyan befolyásolják az adatminőséget a másokban. Így a perspektívát rövid távról hosszú távra terelheti: mit akarunk ezekkel a kutatási adatokkal? Sok életrajz van forgalomban egy adott felhasználásra vagy egy bizonyos felhasználói csoportra összpontosítva.



Forrás: <https://openscience.hu/kutatasi-adatok-nyomán>

A kutatási adatok életciklusának hat főbb állomása van, melyek további részállomásokra oszthatók.

1. Létrehozás
2. Feldolgozás
3. Analizálás
4. Tárolás
5. A hozzáférés beállítása
6. Újrafelhasználás

Természetesen ez a ciklus ennél azért sokkal bonyolultabb, és nem feltétlen körkörös a teljes folyamat, de szemléltetni jól lehet ezzel az ábrával. Amit érdemes kiemelni, az az hogy a létrehozási fázisban már megjelenik az adatmenedzsment megtervezése, így a kutatási adatkezelési terv megírását is például erre a pontra lehet helyezni.

Az alábbi ábrákon láthatóak az egyes állomások és a hozzájuk kötődő részfeladatok:



*A képet készítette Petró Leonárd (CC-BY)*

### ***1.3. Mi az a kutatási adatmenedzsment és miért van rá szükség? (research data management)***

A kutatási adatok kezelése, más néven adatmenedzsment, a kutatás életciklusa során meghozott valamennyi döntést lefedi a projekt tervezési szakaszától egészen az adatok hosszú távú tárolásáig.

*Miért fontos, hogy a kutatási adatainkat jól kezeljük?*

Az adatok kezelésének és megosztásának számos előnye van:

- a kutatási eredmények így igazolhatók,
- adatvesztés elkerülése,
- időhatékonyság,
- az adataink biztonságban vannak,
- megkönnyítjük az adatok újrafelhasználását,
- növelhetjük cikkünk láthatóságát, és ezzel impaktját.

Azonkívül, hogy így igazolhatóak az eredményeink, elkerüljük vagy redukálhatjuk az adatvesztés esélyét. Időhatékonyságról azért beszélhetünk, mert az a kutató, aki előállította az adatokat, fogja leghamarabb újra elővenni és használni ezeket, így neki nagyon fontos, hogy ne órákat kelljen keresnie azért, hogy megtalálja egy egy-két évvel vagy éppen öt évvel ezelőtti projektben létrehozott táblázatot vagy fájlt. Ezzel tudhatjuk legnagyobb biztonságban az adatokat, külsős személyek számára pedig megkönnyítjük az újrafelhasználást, ezzel egyszerűen növelhetjük a saját munkánk impaktját.

### ***1.4. FAIR adatok – FAIR adatkezelés, irányelvek***

*Nyílt adatkezelés előfutára: a F.A.I.R.*

Az ún. FAIR alapelveket az adatközpontú és a gépi eszközökkel támogatott nyitott tudomány (open science) szem előtt tartásával dolgozták ki. Az elképzelés legfőbb alapelve az volt, hogy mind a gépek, mind az emberek találják meg, férjenek hozzá, működjenek együtt, és ezáltal újra használják fel egymás kutatási tárgyait.

Az eredeti FAIR-tanulmányban – ‘The FAIR Guiding Principles for scientific data management and stewardship’<sup>1</sup> (Wilkinson, 2016) –

---

<sup>1</sup> <http://www.nature.com/articles/sdata201618>

fogalmazták meg ezen elveket; tulajdonképpen iránymutatásokat nyújt a digitális eszközök megtalálhatóságának, hozzáférhetőségének, átjárhatóságának és újrafelhasználásának javítására – de (tudatosan) nem tartalmaz részleteket a megvalósítási lehetőségekkel kapcsolatban.

A FAIR alapelvek ma már világszerte az adatkezelési környezet szerves részét képezik, több nagy intézmény/hálózat is átvette már (CERN, RDA, AGU, hogy csak néhány példát említsünk), és az Európai Nyílt Tudományos Felhő (European Open Science Cloud) építési tervének alapját képezik.

*Mit jelent maga a kifejezés?<sup>2</sup>*

- F, mint Findable, azaz megtalálható,
- A, mint Accessible, azaz elérhető,
- I, mint Interoperable, azaz együttműködő,
- R, mint Reusable, azaz újrafelhasználható.

FAIR Principles	Compliance
 <b>Findability</b> Resource and its metadata are easy to find by both, humans and computer systems. Basic machine readable descriptive metadata allows the discovery of interesting data sets and services.	<ul style="list-style-type: none"><li>✓ F1. Resource is uploaded to a public repository.</li><li>✓ F2. Metadata are assigned a globally unique and persistent identifier.</li></ul>
 <b>Accessibility</b> Resource and metadata are stored for the long term such that they can be easily accessed and downloaded or locally used by humans and ideally also machines using standard communication protocols.	<ul style="list-style-type: none"><li>✓ A1. Resource is accessible for download or manipulation by humans and is ideally also machine readable.</li><li>✓ A2. Publications and data repositories have contingency plans to assure that metadata remain accessible, even when the resource or the repository are no longer available.</li></ul>
 <b>Interoperability</b> Metadata should be ready to be exchanged, interpreted and combined in a (semi)automated way with other data sets by humans as well as computer systems.	<ul style="list-style-type: none"><li>✓ I1. Resource is uploaded to a repository that is interoperable with other platforms.</li><li>✓ I2. Repository meta- data schema maps to or implements the CG Core metadata schema.</li><li>✓ I3. Metadata use standard vocabularies and/or ontologies.</li></ul>
 <b>Reusability</b> Data and metadata are sufficiently well-described to allow data to be reused in future research, allowing for integration with other compatible data sources. Proper citation must be facilitated, and the conditions under which the data can be used should be clear to machines and humans.	<ul style="list-style-type: none"><li>✓ R1. Metadata are released with a clear and accessible usage license.</li><li>✓ R2. Metadata about data and datasets are richly described with a plurality of accurate and relevant attributes.</li></ul>

*A kép forrása: <https://bit.ly/3wSYmEz>*

<sup>2</sup> <https://www.force11.org/group/fairgroup/fairprinciples>



### Megtalálható:

- emberek és gépek számára egyaránt,
- egyedi azonosítóval rendelkezzen,
- metaadatokkal gazdagon ellátott,
- metaadatai repozitóriumban legyenek tárolva.

### Elérhető:

- emberek és gépek számára egyaránt könnyű hozzáférési, le-töltési lehetőség,
- sztenderd kommunikációs protokoll segítse az adatok elérhe-tőségét,
- a metaadatok legyenek elérhetők az adatok megsemmisítése után is.

### Együttműködő:

- lehetőséget ad az adatok változtatására, cseréjére,
- (meta)adatok a FAIR alapelveit követő szókészleteket hasz-nálnak,
- (meta)adatok tartalmazzák a más (meta)adatokra vonatkozó minősített hivatkozásokat.

### Újrafelhasználható:

- lehetőséget ad az adatok jövőbeni felhasználására,
- a (meta)adatok egyértelmű és hozzáférhető adathasználati en-gedélyekkel kerülnek kiadásra,
- a (meta)adatok megfelelnek a domain-releváns közösségi szabványoknak.



Készítette: *Petró Leonárd (CC-BY)*

### *A F.A.I.R. lényegi megfontolásai*

A kutatási adatokban rejlő lehetőségek teljes körű kihasználása érdekében azokat a kutatási ökoszisztémába a lehető legjobban *megtalálható, hozzáférhető, interoperábilis és újrafelhasználható* módon kell bevonni. A FAIR alapelvek 15 tényezőbből állnak.

- Irányelv: adatok nyitottsága, hosszú távú tárolás
- Nemcsak az emberek, de a gépek számára is
- „As open as possible, as closed as necessary”
- Metaadat
- Adatok osztályozása



A kép forrása: <https://bit.ly/3jbyS1d>

Kisokos:

*Mi a különbség a FAIR data és az Open data között?  
Egáltalán van-e különbség?*

A FAIR nem egyenlő a *nyílttal*. A FAIR nevében az A, mint accessible jelentése: elérhető, de csak **jól meghatározott feltételek** mellett.

*Mi nem a FAIR?*

- Nem szabvány, bár a rövidítést gyakran használják ebben az összefüggésben.
- A FAIR nem egyenértékű a *nyílttal* (és a nyílt nem egyenértékű a *szabaddal*): Sok oka lehet annak, hogy az adatok nem nyíltak, és csak bizonyos feltételek mellett állnak rendelkezésre bizonyos felhasználók számára (beleértve a gépeket is). Amíg a hozzáférhetőségi/elérhetőségi feltételeket megfelelően leírják, a nem nyílt adatok teljes mértékben FAIR-ek lehetnek. Ugyanakkor a teljesen nyílt és a korlátozás nélküli adatok nagyon alacsony pontszámot érhetnek el a FAIR-mutatókban, mivel például a gépeknél nem működtethetők.
- A FAIR alapelvek önmagukban nem fedik le a belső adatminőségi vagy az alapvető etikai szempontokat.

## 2. TERVEZÉS, RENDSZEREZÉS ÉS ADATKERESÉS

### Bevezető gondolatok

A fejezet áttekintést ad a kutatási adatokkal kapcsolatos főbb teendőkről a kutatás tervezési szakaszában. Példákat mutat be adatkezelési terv létrehozásának megkönnyítése érdekében, ismerteti a metaadatok szerepét és fontosságát, és megismerteti az olvasót az adatszervezés, valamint az adatbiztonság fontosságával.

### Témakörök

- A kutatás megtervezése
- Kutatási adatkezelési terv (Data management Plan, DMP Tools, DMP példák, sablonok)
- Adatgyűjtés – már meglévő adatok felkutatása
- Metaadatok szerepe, fontossága
- Adatszervezés, adatbiztonság

### Irodalmak

- <https://opencscience.hu/kutatasi-adatkezelesi-terv>
- <https://www.dcc.ac.uk/guidance/how-guides/develop-data-plan>
- <https://bit.ly/2Uy4jIU>
- <https://gyires.inf.unideb.hu/GyBITT/11/ch05s02.html>
- <https://bit.ly/2U02vrD>
- <https://mersz.hu/blog/igy-keszits-meno-kutatasi-tervet/>
- <https://bit.ly/3wRc7DZ>
- <https://webarchivum.oszk.hu/mediawiki/index.php/OAIS>
- <https://bit.ly/3xLs2DB>
- <https://bit.ly/3j6giYh>
- <https://tmt.omikk.bme.hu/tmt/article/download/5/10453/>

#### 2.1. A kutatás megtervezése

A kutatás megtervezése, következetes lépések megállapítása a kutatás megkezdésének esszenciális feltétele. Tervezés nélkül a kutatás, a kutatáshoz tartozó tevékenységek tudatos és következetes véghezvitele nagy valószínűséggel problémákba fog ütközni, ezért szükséges kutatási tervet létrehozni, mely tartalmazza az alábbi szempontokat:

- Milyen kérdést igyekszik megválaszolni a kutatás; mi a kutatás célja?
- Alátámasztás: Milyen előzmények köthetők a kutatásunkhoz, milyen a téma szakirodalmi háttere?
- Milyen relevanciái vannak a kutatásnak? Melyek a megfogalmazott hipotézisek?
- A kutatásban használt fogalmak meghatározása
- Az említett fogalmakhoz milyen eljárásokat, mérőszámokat társítunk?
- Milyen mintavételi módokat használunk? Melyek a mintavétel körülményei?
- Kérdőív és interjú forgatókönyv összeállítása
- Felhasznált szakirodalom összegyűjtése
- Ütemezés/ütemterv készítése
- Esetleges/várható költségek számbavétele

## ***2.2. Kutatási adatkezelési terv (Data management Plan, DMP Tools, DMP példák, sablonok)***

A kutatási adatkezelési terv (Data management Plan, rövidítve: DMP) nem más, mint egy rövid (egy-két oldalas) és könnyen áttekinthető leírás arról, hogy hogyan kell kezelni a kutatás során összegyűjtött adatokat. Elkészítése egyfelől a kutató saját érdeke, másfelől pedig egyre gyakrabban követeli meg a kutatás finanszírozója az adatkezelési terv benyújtását a támogatási kérelmek mellé. Fontos szem előtt tartani azt, hogy sok esetben az adatkezelési terv nincs azzal letudva, hogy a kutatás tervezési szakaszában elkészítettük és be is nyújtottuk a finanszírozó felé. A dokumentumra célszerű egy folyamatosan változó, a kutatás szakaszaihoz, azok változó jellegéhez igazodó leírásra tekinteni. Ennek fényében tehát érdemes a dokumentumot időről időre frissíteni. A dokumentum első verzióját legkésőbb a kutatás hatodik hónapja alatt célszerű elkészíteni.

*Mikor kell frissíteni az adatkezelési tervet?*

Általánosságban elmondható, hogy akkor kell frissíteni, ha valamilyen jelentős változás következik be a kutatási folyamat során. Ilyen esetben célszerű revízió alá venni a DMP-t és ha a körülmények indukálják, frissíteni kell azt.

Az alábbiakat kell tárgyalnia egy adatkezelési tervnek:

- ki hozza létre az adatokat,
- milyen típusú adatok jönnek létre,
- hol tárolják ezeket az adatokat,
- hogyan tárolják az adatokat,
- ki férhet hozzá az adatokhoz a projekt végeztével,
- a projekt lezárulását követően meddig lehet hozzáférni az adatokhoz,
- kiegészítő jelleggel tartalmazhatja a terv az alábbi elemeket: biztonságra, fenntarthatóságra és az adatok minőségére vonatkozó passzusok.

*Mit szokott tartalmazni egy kutatási adatkezelési terv:*

- Adminisztratív adatokra vonatkozó rész:
  1. Projekt neve
  2. Projekt rövid jellemzése
  3. A fő kutató(k) neve(i)
- Adatgyűjtéssel kapcsolatos rész:
  1. Milyen típusú adatokat fogunk gyűjteni?
  2. Hogyan fogjuk létrehozni az adatokat?
  3. Milyen metaadat-szabványokat fogunk használni és miért?
- Etikai és jogi megfelelés
- Tárolás és mentés
- Kiválasztás és megőrzés
- Adatmegosztás
- Feladatok és források



Készítette: Petró Leonárd (CC-BY)

## *Hol található DMP segédleteket vagy sablonokat?*

1. DMPonline – <https://dmponline.dcc.ac.uk>
2. ARGOS – <https://argos.openaire.eu/home>
3. DMPTool – <https://dmptool.org>
4. Data Stewardship Wizard – <https://ds-wizard.org/features.html>
5. Horizon 2020 DMP minta – <https://bit.ly/3vXnkla>
6. Finanszírozói adatkezelési minták – a teljesség igénye nélkül:
  - a. Gazdasági és Társadalmi Kutatási Tanács DMP (Egyesült Királyság) – <https://bit.ly/3wUdl13>
  - b. Svéd Kutatói Tanács mintája – <https://bit.ly/3wWEZdS>
  - c. Natural Environment Research Council DMP – <https://bit.ly/3zUcYFw>
7. Egyetemi és más intézményi minták:
  - a. Checklist for Data Management Plan – University of Gothenburg – <https://bit.ly/3d5gX8y>
  - b. MIT Libraries Data Management – <https://bit.ly/3dq6rZZ>
  - c. DMP – University of Strathclyde – Glasgow – <https://bit.ly/2SsqR4q>

## *A legismertebbek*

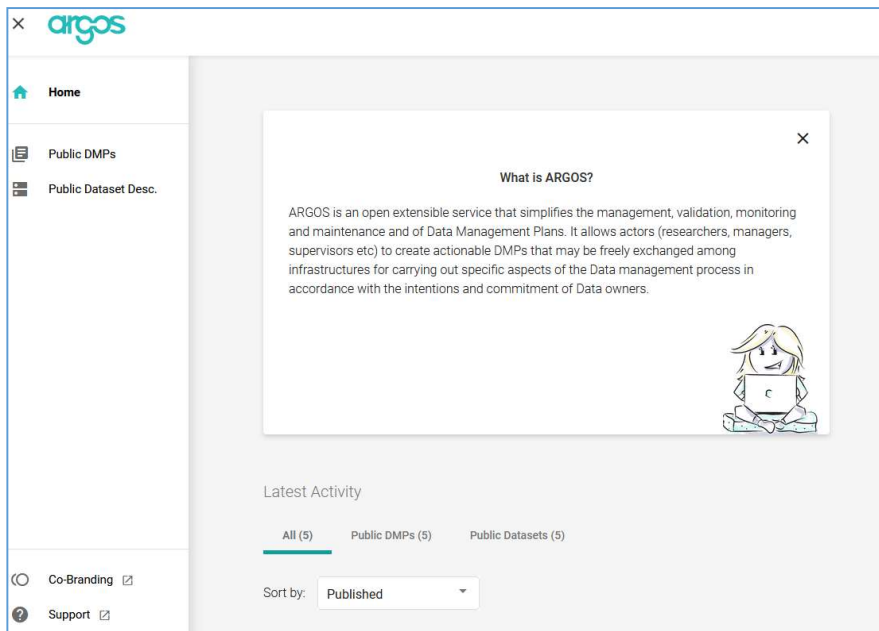
- DMPonline<sup>3</sup>: regisztráció után a felhasználók egyéni adatmenedzsment terveket készíthetnek, és akár több egyetem és kutatásfinanszírozó szervezet előre elkészített sablonját is használhatják.

DMPonline helps you to create, review, and share data management plans that meet institutional and funder requirements. It is provided by the Digital Curation Centre (DCC).

59,972 Users	314 Organisations	65,212 Plans
--------------	-------------------	--------------

<sup>3</sup> <https://dmponline.dcc.ac.uk/>

- ARGOS<sup>4</sup>: DMP-k kollaboratív megírására alkalmas szolgáltatás.



### 2.3. Adatgyűjtés – már meglévő adatok felkutatása

*Miért is fontos a már meglévő adatok felkutatása és felhasználása?*

Elsősorban időt és pénzt takarítunk meg, hiszen egyes kísérletek és azok későbbi reprodukálása igen költséges és időigényes tevékenységek tudnak lenni. Éppen ezért kulcsfontosságú a kutatások során keletkezett adatok pontos rögzítése és megfelelő metaadatokkal, leírásokkal való ellátása, hogy az ilyen minőségi adatok újra felhasználhatóak legyenek a későbbiek során.

### 2.4. Mi a metaadat és miért fontos?

A digitális tudomány fejlődésének egyik alapvető eleme a (digitális) adatok megosztása, hozzáférhetővé tétele. A kutatások során keletkező nagy mennyiségű adat fontos befolyásoló tényezője a kutatás minőségének, sikerességének és általában a kutatási folyamat zökkenőmentes haladásának. Külön figyelmet igényel tehát az értékes kutatási

<sup>4</sup> <https://argos.openaire.eu/explore-plans>



adatok megfelelő leírása (metaadatokkal való ellátása) és kezelése, mely nagyban hozzájárulhat ahhoz, hogy a kutatás kiemelkedjen a többi közül, és persze ahhoz is, hogy a keletkezett adatok később is felhasználhatóak legyenek.

A metaadatok a tudományos kommunikáció egyik leghatékonyabb eszközei, segítik a felfedezhetőséget, visszakereshetőséget és hozzáférést, valamint szerepet játszanak a hibák elkerülésében.

### *Definíció*

A népszerű definíció szerint a metaadat (meta data) adat az adatról. Elsődleges feladata az, hogy a digitális információtömegben segítse a fontos, releváns információk megtalálását adott szempont szerint. Közkeletű példa szokott lenni a digitális fényképek esete:

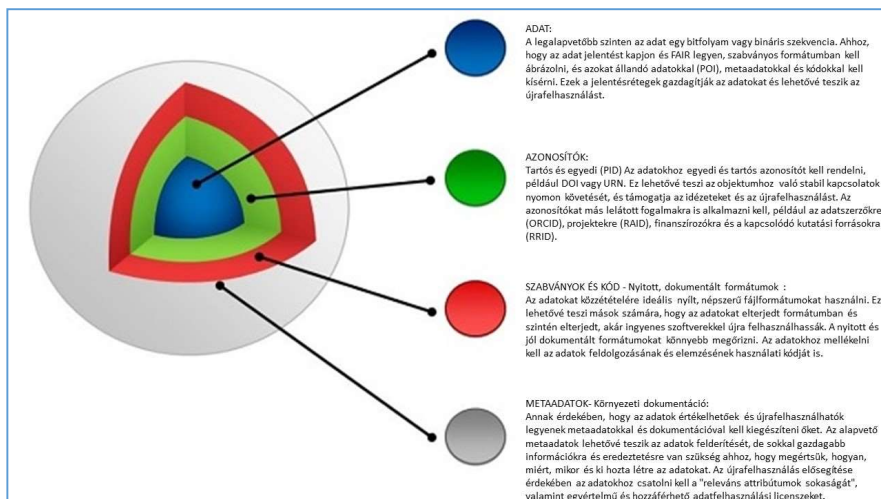
Készítünk egy fényképet az okostelefonunkkal. Ennek eredményeként keletkezik maga a digitális kép, ami a számunkra kedves vagy fontos személyt, jelenetet, helyet stb. rögzíti. Emellett automatikusan keletkeznek leíró adatok is:

- milyen fájlformátumban tárolódik a digitális fénykép,
- mikor készült a felvétel,
- mekkora a felvétel mérete,
- milyen készülék készítette a felvételt,
- mekkora a felvétel felbontása,
- adott esetben: hol készült a felvétel.

Ezek a strukturált információk az általunk készített képről nem mások, mint metaadatok. Természetesen az itt felsorolt leíró adatok korántsem teljeseek, már csak azért sem, mert mi magunk is ki tudjuk egészíteni őket:

- címkézzük/tárgyszavazzuk a képeket,
- megjelöljük azt, hogy ki van a képen,
- felhasználásról szóló licencet adhatunk hozzá...

A metaadatok önmagukban is értelmezhetőek adatként, így akár róluk is keletkezhetnek további metaadatok!



*FAIR Principles & Digital Objects: The role of METADATA nyomán*  
<https://bit.ly/2TVd3sb>

## *Metaadat típusai*

### 1. Leíró

Adott dokumentum keletkezésével, módosításával kapcsolatos adatokat írják le és biztosítják az információforrás visszakereshetőségét. Klasszikus példaként a könyvtári katalógusokban lévő bibliográfiai rekordok által tárolt adatokat szokás említeni (szerző, cím, tárgyszavak, kiadási adatok stb.).

### 2. Adminisztrációs

A digitális tartalom menedzseléséhez szükséges információkat tartalmazzák:

- technikai leírások,
- felhasználási adatok,
- jogok/jogosultságok kezelésről szóló adatok,
- származási adatok.

Általában valamely digitális gyűjteményt menedzselő szakemberek munkáját segítik, így a felhasználó személyeknek kevésbé hasznosak, mivel előlük többnyire rejtve maradnak.

### 3. Szerkezeti

Komplex digitális dokumentumok belső szerkezetét leíró és összefűző kötelék, mely leírja a forrás egyes elemei közti kapcsolatokat. A korábbiaktól eltérően ezek a metaadatok nem a felhasználó személy számára, hanem elsősorban a dokumentumokat megjelenítő szoftver számára értelmezhető módon tartalmaznak leíró adatokat. Egy automatikus tartalomjegyzék létrehozásához például szerkezeti metaadatok szükségesek.

### 4. Megőrzési

Forrásokot és az azokat alkotó fájlok keletkezésére és további történetére vonatkozó adatokat mutatnak be, továbbá a fájlok hosszú távú megőrzéséhez szolgáltatnak információt:

- használathoz szükséges technológia leírása,
- újrafelhasználási paraméterek,
- jogkezelés,
- újabb fájlformátumokba történő migráció dokumentálása.

#### *Metaadatszabványok*

A metaadatszabványok létrejötte a digitális dokumentumok és az azokat leíró metaadatok sokféleségének köszönhető. Különböző digitális gyűjtemények jöttek létre, bennük különféle fájl típusokkal és jogosultságokkal, különböző célokkal, felhasználói célcsoportokkal. Tulajdonképpen a sokféleségből fakadó különbözőség egyfajta kényszerként hívta életre a metaadatszabványokat. Mi tehát egy metaadatszabvány? Nemzetközileg elfogadott metaadatelemek meghatározott, folyamatosan fejlesztett és karbantartott készlete.

Egy metaadatszabvány meghatározza:

- a felhasználható metaadatelemeket,
- a felhasználható elemek definícióját,
- az elemekre vonatkozó szabályokat, például kötelező-e vagy opcionális egy metaadatelemet használni,
- tartalmi szabályokat, például milyen formátumban adjuk meg a dátum értékét,
- szintaktika: a szabvány értelmezhetőségét biztosítja különféle szoftverek számára.

## A leginkább elterjedtebb/népszerűbb szabványok listája

- Dublin Core (DC) – <http://www.dublincore.org>
- Encoded Archival Description (EAD)  
<http://lov.gov/ead/ead.html>
- Machine Readable Cataloguing (MARC)  
<http://loc.gov/standards/mets>
- Metadata Encoding and Transmission Standard (METS)  
<http://loc.gov/standards/mets>
- Metadata Object Description Schema (MODS)  
<http://loc.gov/standards/mods>
- Preservation Metadata Implementation Strategies (PREMIS)  
<http://www.oclc.org/research/projects/pmwg>
- Visual Resources Association (VRA)  
<http://www.loc.gov/standards/vracore/>

Ugyanakkor nemcsak nemzetközileg elfogadott szabványokról beszélhetünk, hanem helyi metaadat-alkalmazási profilokról is. Ezek többnyire csak egy adott intézményi digitális kollekcióna jellemző belső szabványok.

### *Dublin Core metaadatséma*

Népszerűségének oka, hogy bármely tudományterületre alkalmazható. Ezt a leíró elemeket tartalmazó szabványt már 1995 óta fejlesztik, ekkor került meghatározásra egy olyan alap elemkészlet, mely alkalmas elektronikus információk leírására. Egyszerű, átlátható és rugalmas szabvány, emellett rendkívül gyorsan reagál az új igényekre. Első körben 13, majd 15 alapelemet határoztak meg:

1. Title (cím)
2. Creator (alkotó/szerző)
3. Subject and Keywords (téma és kulcsszavak)
4. Description (leírás)
5. Publisher (kiadó)
6. Contributor (közreműködő)
7. Date (dátum)
8. Type (dokumentumtípus)
9. Format (fájl formátum)
10. Resource Identifier (azonosító a hálózaton)
11. Source (származási hely)

12. Language (nyelv)
13. Relation (kapcsolatok)
14. Coverage (hatókör)
15. Rights (jogok)

A felsorolt elemek használata opcionális, nincsenek kötelező elemek megadva, az egyes elemek tetszőlegesen ismételhetőek, sorrendjük sem kötött. Ezt a 15 elemből álló szabványt egyszerű Dublin Core-nak nevezik.

A minősített Dublin Core-t ezen a 15 elemen kívül az jellemzi, hogy az alap elemkészlet tovább finomítható és részletezhető. Ezeket a jelentéspontosító eszközöket „minősítőnek” nevezzük. Használatuk nem kötelező, de jó szolgálatot tudnak tenni, ha következetesek akarunk lenni a rendelkezésre álló digitális források leírásában.

Egy példa:

DC alapelem: Date  
kapcsolódó minősítők:

- created: létrehozás dátuma,
- valid: forrás érvényességének dátuma,
- issued: a forrás kibocsátásának dátuma,
- available: a forrás első elérhetőségének dátuma,
- modified: a forrás módosításának dátuma,
- accepted: a forrás elfogadva ekkortól,
- copyrighted: a forrás szerzői jog által védett ekkortól,
- submitted: a forrás feltöltésének dátuma.

Hogy néz ki a valóságban egy dátumelem-érték?

Pl.: DC.DATE.CREATED 2021-02-03

## ***2.5. Adatszervezés, adatbiztonság***

Ahogy azt az 1. fejezetben láthattuk, a kutatási adatok sok esetben rendhagyó formában és formátumban keletkeznek, illetve rögzülnek. Ezek az adattípusok pedig nagy változatosságot tudnak mutatni egy-egy kutatási területen, tudományterületen belül is. Az adatok hozzá-

férhetőségét biztosítani kell a jövő számára, ám ez több problémás szempontot is felvet:

1. **Hozzáférhetőség:** a pályázati pénzekből finanszírozott kutató-sok szinte kivétel nélkül megkövetelik a kutatás során keletkezett adatok közzétételét. Manapság ezt a hozzáférhetőséget egyértelműen az IT infrastruktúra biztosítja. Korábban a pályázatkíró, illetékes intézmény webtárhelye, fizetett tárhelyek, Cloud-tárhelyek (Google Drive, Dropbox, OneDrive stb.), hálózati meghajtók szolgáltak erőforrásként az adatok disszeminációjához. Manapság ezt a szerepet intézményi, illetve központi (például egyetemi, akadémiai, nemzetközi) adatrepozitóriumok veszik át, amelyeket kifejezetten kutatási adatok tárolására, megosztására és menedzselésére fejlesztettek ki.
2. **Kompatibilitás:** Ahogy az említésre került, a kutatás sok esetben időigényes folyamat. Itt éveken, évtizeden vagy évtizedeken át történő adatgyűjtésre is gondolnunk kell. Az idő előrehaladtával a technológiai és IT sztenderdek, főbb formátumok, protokollok is változnak. A gondos tervezés része tehát, hogy lehetőleg olyan szabványos formátumban tároljuk adatainkat, melyek a kutatás ideje alatt is mainstreamnek számítanak, és számíthatunk rá, hogy évek múltán is beolvashatóak maradnak.
3. **Megőrzés és tárolás:** Elsődleges feladat az adatok integritásának és épségének megőrzése annak érdekében, hogy a jövőben bármikor felhasználhatóak legyenek. A hosszú távú tárolás technikai feltételei adottak (biztonsági másolatokkal, szinkronizációval), ám sok esetben csak addig voltak biztosítva, míg az erre a célra elkülönített pénzforrás fedezte a tárolási költségeket. Szerencsére manapság több ingyenes alternatíva is rendelkezésre áll, melyek általában megfelelnek a pályázatkíró feltételeinek.

### 3. ELÉRÉS, KÖZZÉTÉTEL ÉS ARCHIVÁLÁS

#### Bevezető gondolatok

A fejezet áttekintést ad a kutatási adatok publikálásával kapcsolatos főbb teendőkről, a közzététel előnyeiről. Példákon keresztül ismerteti a különböző dokumentum és szerzői azonosítókat, megismerteti az olvasót az adatrepozitórium fogalmával, és néhány fajtája bemutatására kerül. A kutatási adatok tárolásával és hosszú távú megőrzésének egyes aspektusai is ismertetésre kerülnek.

#### Témakörök

- Adatok publikálása (az adatok közzétételének előnyei)
- Azonosítók szerepe (dokumentum- és szerzői azonosítók)
- Adatrepozitóriumok
- Re3data
- Zenodo, Figshare, Dryad, CONCORDA
- Tárolás, hosszú távú megőrzés

#### Irodalmak

- [https://mersz.hu/dokumentum/matud\\_\\_255](https://mersz.hu/dokumentum/matud__255)
- <https://openscience.hu/nyilt-hozzaferesu-adatrepozitoriumok/>
- <https://openaccess.mtak.hu/index.php/kiadoknak/doi>
- <https://bit.ly/35LYQR1>
- <https://bit.ly/3gVODXy>
- <https://bit.ly/3vQBxAi>
- <https://bit.ly/3dae5qW>
- [https://n2t.net/e/ark\\_ids.html](https://n2t.net/e/ark_ids.html)

#### 3.1. Azonosítók szerepe (dokumentum- és szerzői azonosítók)

A világháló térhódítása előtt is jól azonosítható igényként jelent meg a dokumentumok és különféle időszaki kiadványok egyértelmű beazonosításának szükségessége. Ennek érdekében lett életre hívva az International Standard Book Number (ISBN) és az International Standard Serial Number (ISSN) azonosító.

A publikációk és egyéb erőforrások interneten való elhelyezése és megosztása további kihívás elé állította a világot, hiszen a publiká-

ciók robbanásszerűen növekvő tömege minden eddigénél jobban megköveteli a dokumentumok és a szerzők egyértelmű beazonosítását. Mi több, mindez újfajta azonosítási szabványok létrejöttét indukálta, hogy biztosított legyen a gyors fellelhetőség és az egyértelmű, egyszerű hivatkozhatóság.

### *Dokumentumazonosítási megoldások*

Elektronikus dokumentumok azonosítására leginkább elterjedt megoldásként a *DOI (Digital Object Identifier)* szolgál. Ez egy olyan azonosító, amely online objektumokat (például dokumentumokat) azonosít egyértelműen és hosszú távon. A DOI azonosító egy prefixből és egy suffixből áll, melyeket egy „/” választ el.

Például: 10.15170/DIKE.2020.04.01.02.

A példaként felhozott DOI azonosító mögött az alábbi webcím bújjik meg: <https://bit.ly/35MtpWN>. Amennyiben ez a webcím valamely oknál fogva megváltozik, úgy a DOI igénylőjének kötelessége a változás tényét bejelenteni annak a szolgáltatónak (például CrossRef, DataCite), amelytől korábban a DOI azonosítót igényelte. (Egyetemenk a CrossRef nevű szolgáltatóval áll szerződésben jelenleg, mely elsősorban publikációk azonosítására szakosodott, a DataCite nevű szolgáltató szakterülete a kutatási adatok azonosítása.)

*URI (Uniform Resource Identifier)*: helyeket, neveket vagy épp helyeket és neveket egyszerre azonosító karakterlánc. Legyakoribb előfordulása a webcím vagy URL.

Felépítése: sémanév.sémaspecifikus név, pl.: <https://lib.pte.hu>

*Handle*: olyan azonosító, amely a repozitóriumokban elhelyezett objektumokat azonosítja. A dokumentumok feltöltése során automatikusan generálódik, és állandó hivatkozásként biztosítja az adott dokumentum elérését még akkor is, ha átrendezik a szerveret.

### *Mi okozhat zavart egy szerző azonosításában?*

- névazonosság (pl.: más egyetemen kutató/oktató személy neve megegyezik a saját nevével),
- kevésbé egyedi nevek (a szerző neve gyakori vezeté- és keresztnévből áll),
- névváltozás (pl.: házasságkötés után a szerző neve megváltozik, és úgy dönt, hogy nem hajdonkori nevéen publikál a továbbiakban),



- kéttagú vezetéknev/kéttagú keresztnév következtelen használata,
- keleti névsorrend vs. nyugati névsorrend.

Világosan látható tehát, hogy szükség van olyan azonosítók bevezetésére, melyek segítenek a szerzők (illetve az szerzői affiliációk) mindenkori, egyértelmű beazonosításában. Milyen lehetőségek állnak rendelkezésre?

- *MTMT szerzői azonosító:*  
A Magyar Tudományos Művek Tára<sup>5</sup> segítségével gyorsan és egyszerűen hozható létre személyes tudományos bibliográfia. Ez a személyes bibliográfia a szerzőt egyértelműen azonosító és reprezentáló MTMT azonosító alá tagozódik be. A szerző személye mellett az MTMT azt az intézményt is azonosítja, mely intézmény aktuálisan foglalkoztatja a szerzőt. A továbbiakban felsorolt azonosító típusoktól eltérően az MTMT és ezzel együtt az MTMT azonosító használata jelen pillanatban magyarországi viszonylatban számottevő tényező MTA, ODT, MAB, OTKA kompatibilitás). Ugyanakkor az MTMT szerzői azonosítójához társíthatóak jelen felsorolás további elemei.
- *Google Scholar profil:*  
A Google Scholar webszolgáltatás weboldalán<sup>6</sup> ingyenesen létrehozható profil. Saját nevünk, e-mail-címünk és az érdeklődési területek mellett meg kell adni az aktuális intézmény nevét, mely égisze alatt publikálunk. Ezek után megkezdhetjük a Google Scholar által már indexelt publikációkból a profilunk alá rendezni a ténylegesen általunk írt műveket. Ez történhet a Google Scholar automatikus ajánló rendszerének segítségével, vagy kézi kereséssel.
- *Author ID (Scopus):*  
A Scopus<sup>7</sup> szerzői profil tartalmazza a szerző nevét, illetve a szerzői névváltozatokat, intézményi affiliáció(ka)t, kutatási területet, publikációs/hivatkozási adatokat, valamint Hirsch-indexet és a szerzői azonosítót. Ez a szerzői azonosító automatikusan jön létre, feltétele csupán annyi, hogy egy Scopus által indexelt folyóiratban jelenjen meg publikációnk. Mivel a

---

<sup>5</sup> <https://www.mtmt.hu/>

<sup>6</sup> <https://scholar.google.com/>

<sup>7</sup> <http://www.scopus.com/home.url>

Scopus által erre a célra használt algoritmus nem tévedhetetlen, fontos a szerzői profilok tisztítása annak érdekében, hogy ténylegesen saját publikációink jelenjenek meg a saját Author ID-nk alatt. A szerzői profil összekapcsolható ORCID azonosítóval.

- *ResearcherID (Web of Science/Publons):*  
A szerzői profil azonosítására hivatott egyedi karakterlánc. Egyrészt adott szerző Web of Science<sup>8</sup> Core Collection-ben indexelt publikációit gyűjti egybe, másfelől pedig a Publons<sup>9</sup> felületén létrehozott publikációs listát rendeli egyértelműen egy szerzőhöz. (Ez utóbbi esetben bármilyen kiadványban megjelent publikációt rögzíthetünk szerzői profilunk alá, nincs Web of Science Core Collection-re korlátozva.) A szerzői profilok a Scopushoz hasonlóan egy algoritmus tevékenysége során automatikusan keletkeznek. Szintén összekapcsolható ORCID azonosítóval.
- *ORCID:*  
Egyedi, nemzetközileg is általánosan elismert szerzőazonosító kód, mely ingyenesen igényelhető az ORCID honlapján<sup>10</sup>. ORCID-hoz tartozó profilunkban szintén beállíthatunk intézményi affiliácót. Egyre gyakrabban szükséges publikációk megjelentetése folyamán ORCID azonosító megadása, illetve pályázati kiírásokban is egyre gyakrabban követelik meg az azonosító meglétét.

### 3.2. Adatrepozitóriumok

A kutatási adatok tárolására és közzétételére szakosodott rendszereket adatrepozitóriumnak nevezzük. Általánosságban beszélhetünk intézményi adatrepozitóriumokról, témaspecifikus adatrepozitóriumokról és általános adatrepozitóriumokról. A publikációs (full text szöveget tartalmazó) repozitóriumoktól eltérően az adatrepozitóriumok elsődleges szerepe az, hogy a kutatás, illetve kísérletezés során keletkezett adatokat, illetve az adatok előállításának folyamatát és körülményeit rögzítse. Az adatrepozitóriumok feladata a fent felsoroltak mellett az is, hogy könnyen kereshetővé tegye a benne tárolt digitális ob-

---

<sup>8</sup> <http://webofknowledge.com/UA>

<sup>9</sup> <https://publons.com/>

<sup>10</sup> <https://orcid.org/>

jektumokat, valamint biztosítsa azok hosszú távú megőrzését. A repozitóriumnak a tárolt adatokat egyértelműen és tartósan kell tudnia azonosítani, amelyhez a főbb eszközök a következők:

- *URI (Uniform Resource Identifier)*
- *Handle*
- *DOI (Digital Object Identifier)*

A megszokott dokumentumtípusok (például: Excel-táblázat, PDF, XML-állomány stb.) mellett az adatrepozitóriumok egyik fő képessége az, hogy több komponensből álló csomagokat (RO-research object) is képes kezelni. Egy research object tartalma általában:

- a kutatás során létrejött adatok/a kutatáshoz felhasznált adatok,
- az adatok előállításának módszere, analízis és feldolgozási módszerek,
- az eredet leírása: az adatok keletkezésének körülményei, adatforrások metaadatai, a feldolgozás lépésének metaadatai,
- a kutatás résztvevőinek neve, affiliációja,
- egyéb, értelmezést segítő metaadatok,
- keletkezett publikációk.

A tárolás, megosztás és azonosítás mellett további fontos funkciója az adatrepozitóriumoknak, hogy szabályozni tudják a bennük tárolt objektumok felhasználását. Ez az elhelyezett objektumokhoz társított hozzáférési beállítások és licencek mentén történik. A repozitórium szoftvere naplózza a repozitóriummal kapcsolatos tevékenységeket (belépések, feltöltések, módosítások stb.), ám ami ennél is fontosabb, hogy lehetőséget biztosít arra, hogy jelentős számú objektum leíró adatait adja át olyan központi adattáraknak, melyek közös keresőfelületet nyújtanak több, földrajzilag elkülönült helyen található repozitóriumban lévő rekordtömeghez.

### 3.3. *Re3data*

A kutatási adatokhoz való állandó hozzáférés kihívást jelent a tudományos közösség minden érdekeltje számára. Az elmúlt évek során egyre több kutatóközpont, egyetem és szervezet hozott létre kutatási adatrepozitóriumot. A Re3data<sup>11</sup> nem más, mint ennek a folyamatosan gyarapodó adattárhalmaznak a globális nyilvántartó rendszere, mely 2012 óta nyújt segítséget a repozitóriumok közti eligazodásban: böngészhetünk benne ország, tárgy vagy tartalomtípus szerint, illetve direkt módon is kereshetünk megfelelő adatrepozitóriumok után, a találati listánkat pedig szűrőkkel finomíthatjuk. Tehát a Re3data nem a különböző repozitóriumokban elhelyezett adatkészletekben keres, hanem világszerte (jelenleg) 3580 adatrepozitóriumot próbál meg jellemezni, és összesíti az adattárakhoz kötődő fontos adatokat (például a repozitórium típusa, elérhetősége, gyűjtött adatok jellemzése tárgyszavakkal, kontaktinformációk, repozitórium típusa, nyelvek, üzemeltető intézmény adatai, alkalmazott protokollok, minősítési megfelelések, DOI használat, Open Access lehetőség, stb.).



### 3.4. *Zenodo, Figshare, Dryad, Concorda*

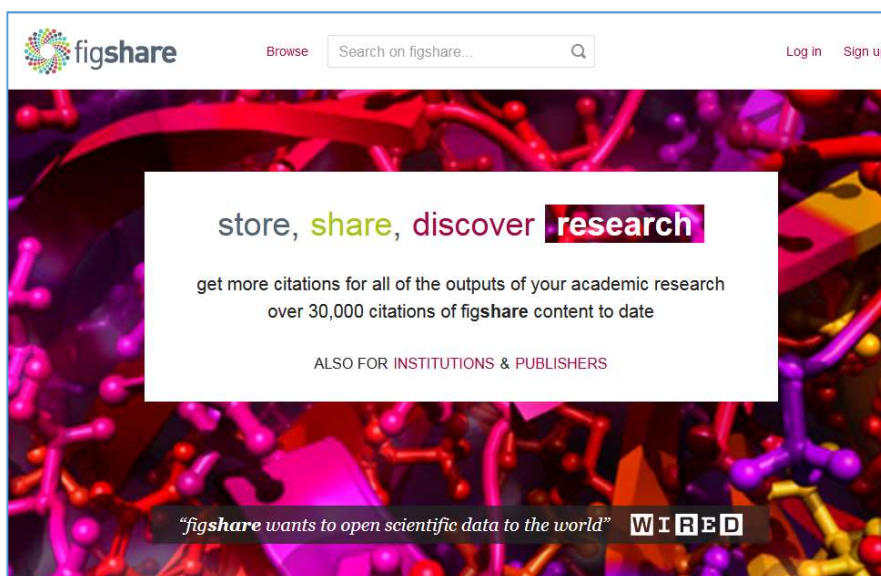
Az alábbiakban röviden ismertetünk néhány szabadon hozzáférhető adatrepozitóriumot. Az online, szabad hozzáférésű általános multidiszciplináris adatrepozitóriumok közül a két legnépszerűbb a

---

<sup>11</sup> <https://www.re3data.org/>

Zenodo és a Figshare. Közös jellemzőjük a felhőalapú technológián túl az, hogy lehetőséget biztosítanak a kutatók számára a kutatási adataik ingyenes elhelyezésére és megosztására, ezzel erősítve az adatok újrafelhasználásának folyamatát.

A Figshare<sup>12</sup> 2011-ben indult, majd 2012 óta az Egyesült Királyság-beli Digital Science nevű cég tulajdonában lévő szolgáltatás.



Jelenleg az alábbiakat biztosítja:

- hozzáférés a kutatási tevékenységünkhöz és a tárolt anyagokhoz, illetve adatokhoz bárhol, bármikor,
- 20 Gb személyes, ingyenes tárhely,
- bármilyen feltölthető fájl típus,
- 5 Gb/fájl maximális méret feltöltött fájlként,
- korlátlan tárhely a Figshare-en szabadon elérhetővé tett adatok számára,
- ingyenes DOI azonosítóval való ellátás (DataCite szolgáltatón keresztül), előzetes DOI azonosító lefoglalás,
- asztali feltöltőkliens,
- kollaborációs tér, ahol az általunk megbízhatónak vélt kutatótársakkal működhetünk együtt,

---

<sup>12</sup> <https://figshare.com/>

- gyűjtemények létrehozása a kutatások jobb átláthatósága érdekében,
- Figshare API a kutatási munkafolyamatok automatizálásához,
- egyedi díjszabás alapján intézményi repozitóriumszolgáltatás nyújtása (Figshare for institutions).

A Zenodo<sup>13</sup> az OpenAIRE és a CERN által létrehozott nyílt hozzáféréssű repozitórium, amely jelenleg lehetővé teszi:

- mind a publikációk, mind az adatok tárolását, valamint eszközök felajánlását ezek összekapcsolásához,
- ingyenes DOI azonosító igénylését a nyilvánosan elérhető objektumhoz; mi több a DOI verziószámozást is,
- akár 50 Gb méretű datasetek feltöltését,
- bármilyen fájl típus feltöltését,
- a rugalmas licenckezelést,
- közösségi gyűjtemények létrehozását és azonos érdeklődésű kutatók csoportokba való szerveződését.



A Dryad<sup>14</sup> egy teljesen nyílt forráskódú, közösségvezérelt projekt, mely 2009-ben indult és DSpace szoftveres alapokra támaszkodott, 2019-ben pedig összeolvadt a Dash-sel, mely egy – a University of

<sup>13</sup> <https://zenodo.org/>

<sup>14</sup> <https://datadryad.org/stash/>

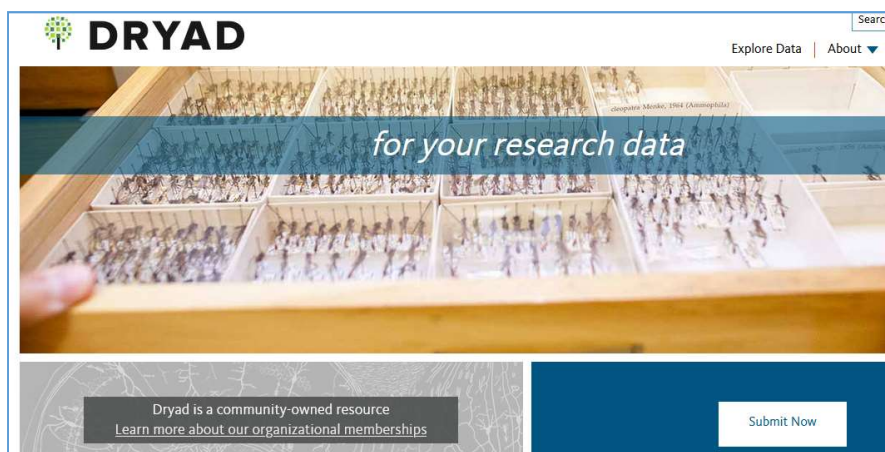
California Curation Center által fejlesztett – adatközzételi célú platform. A Dryad költségmentesen biztosít hosszú távú hozzáférést felhasználói számára az általa tárolt tartalmakhoz a tagok és az adatszolgáltatók pénzügyi támogatásának köszönhetően.

*2013 óta viszont számolni kell submission fee-vel a Dryad rendszerében:*

Amennyiben a Világbank által alacsony jövedelmű vagy alacsonyabb, közepes jövedelmű gazdaságokba sorolt országokból származó kutató nyújt be dataseteket, úgy a Dryad eltekint az ún. Data Publishing Charges-tól (DPC), minden más esetben 120 USD/beküldés DPC költséget számolhatnak fel a kurátorok. Szponzor, illetve DPC fizetési felmentés nélkül beküldött adatok esetében ez a 120 USD összeg 50 Gb-nyi adat beküldését fedezi, minden további 10 Gb-ért 50 USD díjat számolnak fel (50-60 Gb közti beküldés: 50 USD; 60-70 Gb közti beküldés: 100 USD; 70-80 Gb közti beküldés: 150 USD...).

A Dryad jelenlegi jellemzői:

- rugalmas adatformátum-kezelés,
- a kéziratbenyújtási folyamatba könnyen integrálódó rendszer,
- DOI-k biztosítása a felküldött adatállományokhoz,
- elősegíti az adatok láthatóságának növelését,
- professzionális kurátorcsoport biztosítja az adatok és a leíró információk ellenőrzését, javítva azok minőségét,
- keretrendszer biztosít ingyenesen letölthető és újrafelhasználható adatok számára (CC0 licenc),
- biztosítja az adatok hosszú távú megőrzését,
- nyílt forráskódú, szabványkövető technológiát alkalmaz.



### 3.5. *CONCORDA*

A CONCORDA<sup>15</sup> (Concentrated Cooperation on Research Data) a SZTAKI Elosztott Rendszerek Osztálya, a SZTAKI HBIT osztálya és a WIGNER Fizikai Kutatóközpont együttműködésének gyümölcse, a legelső magyar, általánosan a kutatási adatokra specializálódott rendszerek egyike. A rendszer a magyar kutatási szféra egészének rendelkezésére áll. A repozitórium Dataverse alapokon nyugszik, a betöltött adatok a „tárolók” hierarchikus struktúrájában helyezkednek el, így a kutatók, kutatócsoportok, intézetek saját tárolóikban kezelhetik saját adatsomagjaikat. Az új felhasználók fájlelhelyezési jogosultságot az általuk használni kívánt intézményi tároló adminisztrátorától kérhetnek az adott tárolón belül.

A CONCORDA jelenlegi jellemzői:

- regisztráció e-mail-cím, EduID vagy ORCID alapján,
- adatfájlok és metaadatok nyilvános vagy zártkörű megosztása,
- rugalmasan bővíthető leíró metaadatsémák,
- verziókövetés,
- API alapú hozzáférés,
- hivatkozások létrehozása adathalmazokra és fájlokra,
- EndNote XML, RIS és BibTeX formátumokkal való kompatibilitás,
- nemzetközi kutatási adatregiszterekhez kapcsolt rendszer,
- metaadatok nyilvánossá tétele más rendszerek felé (OpenAire, Google Dataset Search stb.),
- DOI igénylés lehetősége a feltöltött adatokra az MTA Könyvtár és Információs Központon keresztül.

---

<sup>15</sup> <https://concorda.sztaki.hu/>





The screenshot shows the CONCORDA web interface. At the top, there is a navigation bar with links for 'Keresés', 'Felhasználói útmutató', 'Támogatás', 'Magyar', 'Regisztráció', and 'Bejelentkezés'. The main header features the CONCORDA logo and the text 'Concentrated Cooperation on Research Data'. Below the header, there is a search bar with '38 Letöltés' and buttons for 'Kapcsolat' and 'Megosztás'. The main content area is titled 'A CONCORDA befoglaló tárolója'. It includes a search bar with 'Keresés' and 'Összetett keresés' buttons. On the left, there is a sidebar with filters for 'Tárolók (14)', 'Adatszomagok (20)', and 'Fájlok (107)'. The main area displays search results for 'Sztakotó rep', including a document titled 'Kormányzati Informatikai Fejlesztési Ügynökség (KIFÜ)' and a document titled 'Sztakotó rep' with a date of '2020. szeptember. 09.' and a file name 'Varburo, Gabor, 2020, "Sztakotó rep", CONCORDA, V1, UNF-6:P2obUN9czqalCU1dTRdwmg== [fileUNF]'. The interface is in Hungarian.

### 3.6. Tárolás, hosszú távú megőrzés

A hosszú távú megőrzés közkeletűen digitális tartalmak, gyűjtemények, adatvagyonok és a hozzájuk kapcsolódó metaadatok (esetleg csatolt dokumentációk) megőrzését, fenntartását jelenti olyan módon, hogy azok felhasználhatóságát a folyamatosan változó és fejlődő hardveres és szoftveres környezet ne befolyásolhassa.

Erre a célra dolgozták ki a 2002-ben szabványosított, majd 2012-ben átdolgozott Open Archival Information System (OAIS) referenciamodelt, mely eredetileg űrkutatási adatok tárolására jött létre, később viszont más területeken is elterjedt.

Fontos megjegyezni, hogy adott esetben nemcsak dokumentumok, fájlok megőrzésében kell gondolkodnunk, hanem a technológia megőrzésében (szoftverkönyezetek, hardverek), azok emulációjában és az adatok konverzióiban (más fájlformátumra történő alakításban) is.

### 3.7. Adatok tárolása

Az adatok tárolásáról mindenképp szót kell ejteni. Rövid és hosszú távon más-más tárolási módokat különböztetünk meg!

Rövid távú:

- Saját meghajtó (merevlemez)
- Flash Drive
- Optikai adattároló eszközök
- Felhő

Hosszú távú:

- Saját weboldal
- Intézményi repozitórium
- Diszciplináris repozitórium, például NCBI Gene
- Interdiszciplináris adatrepozitórium, például Zenodo, Figshare

zipcloud

box

justcloud.com

bitcasa  
INFINITE STORAGE

OneDrive

IDrive®

Dropbox

Google Drive

iCloud

MediaFire

What is the Best Backup  
**Local Storage** OR  
**in the Cloud ?**

OR

A rövid távú tárolási módokat a legtöbben ismerik és használják is, a hosszú távú tárolási módoknál kiemelendő, hogy egyre több adatrepozitóriumot hoznak létre. Ezek általában megbízhatóak, minősítési rendszer is társul hozzájuk, bár jelenleg kevés minősített adatrepozitórium van. Ha egy kutató a kutatási adatait adatrepozitóriumban szeretné tárolni – márpedig jó, ha ott tárolja őket –, akkor először egy diszciplináris adatrepozitóriumban való tárolást szokták javasolni, utána a saját intézmény által üzemeltetett adatrepozitóriumban érdemes tárolni az adatokat, és ha ezek közül egyik sem áll rendelkezésre, akkor egy interdiszciplináris adatrepozitóriumban, mint például a Zenodóban.

A kutatási adatokkal szemben támasztott tárolási és egyéb biztonsági szempontok egyik alapelve az úgynevezett „Here, near, far” elv. Ezt úgy szokás lefordítani, hogy minimum három helyen érdemes őrizni ezeket az adatokat, biztonsági mentést készíteni a kutatásainkról, hiszen egy tűzvész vagy hasonló katasztrófa bekövetkeztével így lehetünk biztosak abban, hogy tőlünk messze is megtalálhatóak ezek az adatok, és újra elő tudjuk őket keresni, ha szükség lenne rájuk.

## 4. LICENC ÉS JOGI KÉRDÉSEK

### Bevezető gondolatok

A fejezet áttekintést ad a szerzői jogok védelmével kapcsolatos főbb teendőkről, a kutatási adatok licencekkel való ellátásának fontosságáról. Röviden ismerteti a Creative Commons licenceket és a licencválasztás folyamatát.

### Témakörök

- Szerzői jogok védelme
- Creative Commons licencek
- Egyéb megoldások

### Irodalmak

- <https://creativecommons.org/choose>
- <https://openscience.hu/szerzoi-jogok/>

### 4.1. A szerzői jogok védelme

Az előzőekben ismertetett FAIR-elvek fontos része az, hogy hogyan lássuk el a kutatási adatainkat különböző licencekkel. Ezekben a licencekben meghatározott jogokat adhatunk más kutatóknak a saját munkánkhoz. A cél lehetővé tenni, hogy a kutatási adatok világszerte minden érdekelt fél számára újrafelhasználhatóak legyenek, jól meghatározott licencek érvénye alatt.

#### *Átláthatóság*

A kutatási adatokkal szemben támasztott átláthatósági kritérium abban is megnyilvánul, hogy elvárásként fogalmazódik meg velük szemben az, hogy minden kutató egyértelműen és gyorsan képes legyen eldönteni azt, hogy a mások számára megosztott kutatási adatokkal, adatcsomagokkal mihez kezdhet, felhasználhatja-e saját kutatásaihoz. Szintén fontos, hogy az erre vonatkozó információk gép által is olvashatóak legyenek: például a Google-höz hasonló keresőmotorok számára is láthatóak, érthetőek legyenek. A legegyszerűbb vagy legjobb megoldás az, ha Creative Commons (CC) licencet használ-

juk, mivel a Creative Commons pont úgy fejlesztette ki a licenccmodelljét, hogy ezeknek az elvárásoknak megfeleljen.



Kép forrása: <https://www.orj.co.uk/copyright-protection/>

#### **4.2. Amit a Creative Commons lincense-ekről tudni kell**

A Creative Commons (CC)<sup>16</sup> a szerzői jogok (copyright) egyik alternatívája. A CC széles skálán mozog a szerzői oltalom és a szabad felhasználás között.

Kétféle jogot biztosíthat:

- megoszthatóságot (másolható, megosztható, előadható, bemutatható),
- átdolgozhatóságot, mely alapján származékos művek készíthetők belőle.

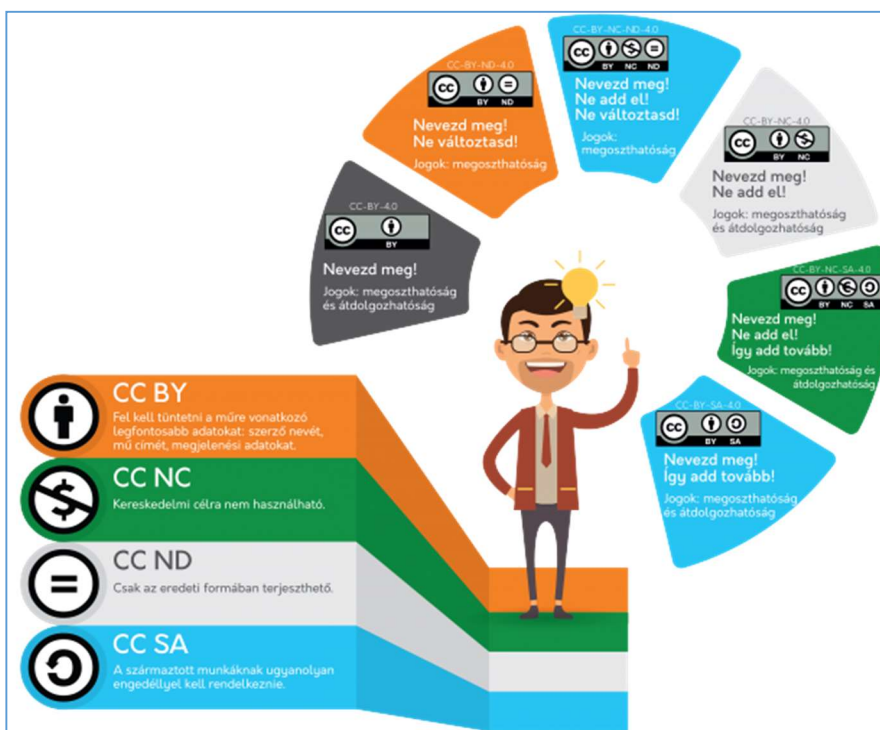
A különböző licenckek eltérően viszonyulnak ehhez a két lehetőséghez.

---

<sup>16</sup> <https://creativecommons.org>

Creative Commons 4.0 licencké nevei	Jogok	Korlátozások <sup>[6]</sup>	Megfelelés a Szabad kulturális művek feltételeinek
Attribution (CC-BY-4.0) <i>Nevezd meg!</i>			szabad licenc <sup>[5]</sup>
Attribution No Derivatives (CC-BY-ND-4.0) <i>Nevezd meg!-Ne változtasd!</i>			nem szabad licenc <sup>[5]</sup>
Attribution Non-Commercial No Derivatives (CC-BY-NC-ND-4.0) <i>Nevezd meg!-Ne add el!-Ne változtasd!</i>		 	nem szabad licenc
Attribution Non-Commercial (CC-BY-NC-4.0) <i>Nevezd meg!-Ne add el!</i>			nem szabad licenc
Attribution Non-Commercial Share Alike (CC-BY-NC-SA-4.0) <i>Nevezd meg!-Ne add el!-Így add tovább!</i>		 	nem szabad licenc
Attribution Share Alike (CC-BY-SA-4.0) <i>Nevezd meg!-Így add tovább!</i>			szabad licenc
Public Domain (CC0) <i>közkinccs<sup>[7]</sup></i>			szabad licenc

*A különböző CC-licencké (Forrás: Wikipedia)*



*A kép forrása: <https://openscience.hu/szerzoi-jogok>*

További előnye, hogy nagyon egyszerű a licencválasztás. A Creative Commons-nak van egy magyarul is elérhető licencválasztó felülete, amely ezeknek a döntéseknek a meghozatalát követően megmutatja, hogy melyik licenc passzol a leginkább az elképzeléseinkhez: <https://creativecommons.org/choose>. Ezek a speciális szimbólumokkal ellátott licencké könnyen dekódolható módon tudatják a döntésünket. A Creative Commons által egy feltöltőúrlapon csak be kell ikszelni, hogy milyen viszonyt szeretnénk létrehozni, és már használhatjuk is.

## A licenc tulajdonságai

Ezen a panelen tett választásaid frissítik a többi panelt ezen az oldalon.

Engedélyezed az átdolgozott műved megosztását?


Igen  Nem  Igen, és azt szeretném, ha a felhasználók is ugyanilyen feltételekkel tennék közzé az általuk létrehozott származékos műveket.

Engedélyezed a műved kereskedelmi célú felhasználását?


Igen  Nem

## Kiválasztott licenc

Nevezd meg! 4.0 Nemzetközi



Ez egy Szabad Kultúra Licenc!



A kép forrása: <https://creativecommons.org/choose>

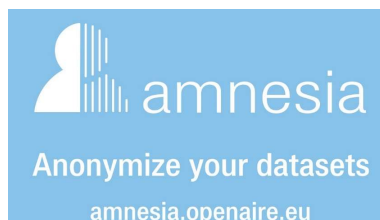
### Személyes adatok

Ki kell térni a személyes adatok kezelésére. Ez egy fontos aspektus, hiszen nagyon sok orvosi vagy egyéb olyan kutatás zajlik, ahol különböző személyes adatokat kezelünk. Személyes adatról akkor beszélünk, ha az információ természetes személyre utal, valamint a személy beazonosítható az adat alapján. Gondoskodni kell a megfelelő védelemről: ezeket az adatokat anonimízálni kell.

*Anonimizálás:* Olyan eljárás, amelynek következtében az érintett nem vagy többé nem azonosítható

Az anonimizálás népszerű online segédeszközei:

- Amnesia online program<sup>17</sup> segítségével.
- ARX Data Anonymization Tool: <https://arx.deidentifier.org>



<sup>17</sup> <https://amnesia.openaire.eu/amnesia>

## 5. AZ EGYETEMI OFFICE 365 ELŐFIZETÉS KIHASZNÁLÁSA A KUTATÁSI FOLYAMAT SORÁN

### Bevezető gondolatok

A tananyag jelen fejezete elsősorban arra igyekszik ösztönözni a kutatókat, hogy használják ki az Egyetem által rendelkezésre bocsátott Office 365 szolgáltatáscsomag elemeit, melyek megkönnyíthetik a kutatómunkát: elsősorban az online kooperatív munkavégzéshez, valamint az adatok és anyagok felhőben való tárolására biztosított infrastruktúrával.

### Célok, megszerzhető kompetenciák

A résztvevő felfedezi, használja/kihasználja a Pécsi Tudományegyetem polgárai számára elérhető Office 365 szolgáltatásban rejlő potenciálokat.

### Témakörök

- Fontosabb szolgáltatások
- Hátrányok és előnyök
- Adatvesztés ellen

#### 5.1. Fontosabb szolgáltatások

A Pécsi Tudományegyetem Office 365 előfizetésének<sup>18</sup> fontosabb szolgáltatásai:

- Dokumentumok (szövegfájlok, bemutatók, táblázatok, OneNote-fájlok) online, böngészőből indított és böngészőből történő szerkesztése,
- Office 365 irodai szoftvercsomag telepítése 5 eszközre (Word, Excel, PowerPoint, OneNote, Access, Publisher, Outlook-kliens, Skype, OneDrive-kliens),
- 1 Tb személyes OneDrive felhőtárhely,
- Kooperáció és online kommunikáció szabadon létrehozható TEAMS-csoportokkal,
- Forms űrlapkészítő alkalmazás,

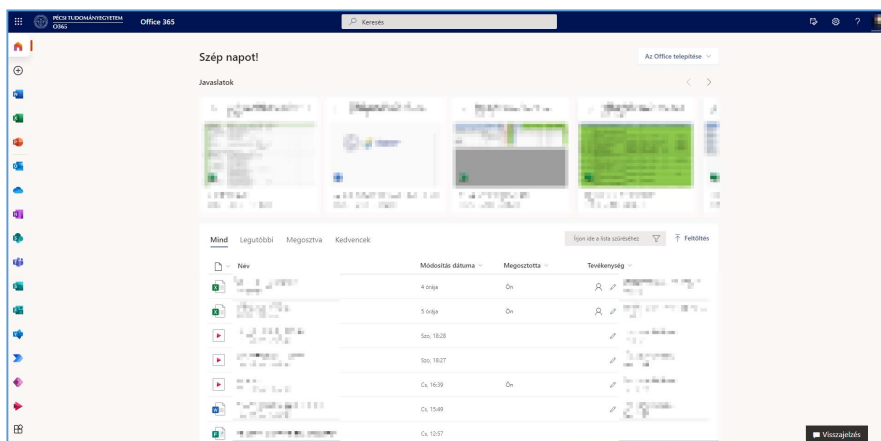
---

<sup>18</sup> <https://www.office.com/?auth=2>



- Power Automate platform eseményvezérelt vagy ütemezett folyamatok létrehozására,
- OneDrive-kliens<sup>19</sup> a személyes OneDrive-tárhely, illetve TEAMS-csapattárhelyek szinkronizálására,
- Megosztás/hozzáférés kezelési lehetőségek.

Bejelentkezési segédlet: <https://bit.ly/3wTrY4T>



*Az Office 365 felülete*

## **5.2. Hátrányok és előnyök**

A szolgáltatás hátrányai:

A OneDrive személyes tárhely fájllelési-megosztási rendszere tipikusan szervezetben belüli (PTE-s) munkatársak hozzáférési jogait engedni szabályozni. A PTE-n kívüli munkatársakkal fájlokat megosztani csak olyan TEAMS-csoportban lehetséges, ahová a külsős felek vendégként meghívást kaptak.

A szolgáltatás előnyei:

Egyrészt jogtiszt és karbantartott Office-szoftvereket telepíthetünk számítógépünkre/laptopunkra/okostelefonunkra, amelyekkel akár közösen szerkeszthetünk fájlokat. Másfelől pedig kihasználhatjuk a OneDrive-kliens által nyújtott lehetőségeket:

<sup>19</sup> <https://www.microsoft.com/hu-hu/microsoft-365/onedrive/download>

- személyes OneDrive felhőtárhelyünk tallózható meghajtóként jelenik meg a számítógép fájlkezelőjében;
- bármely olyan TEAMS-es csoport csatornájának tárhelyét szinkronizálhatjuk, melyeknek tagjai vagyunk: a szinkronizált TEAMS-tárhelyek szintén megjelennek tallózható háttértárként a számítógépünk fájlkezelőjében;
- mind a OneDrive, mind a TEAMS/SharePoint megosztott tárhelyeken elhelyezett Office-dokumentumok esetében elérhető a verziókezelés, mely megkönnyíti az esetleges szerkesztési hiba után elmentett fájl korábbi állapotára történő visszaállítását;
- biztonsági mentések készítése: a OneDrive-kliensben ezt a funkciót bekapcsolva lehetőségünk nyílik arra, hogy a helyi meghajtónkon lévő „Asztal”, „Dokumentumok” és „Képek” mappa tartalma, valamint azok módosításai automatikusan szinkronizálódjanak a OneDrive személyes tárhelyünkkel;
- fájlok mentése OneDrive-tárhelyre és fájlok megnyitása közvetlen OneDrive-tárhelyről.

### **5.3. Adatvesztés ellen**

Az adatvesztés elkerülése érdekében az iménti felsorolás utolsó két pontja igényel némi figyelmet. A „biztonsági mentés” révén az említett helyi mappákban létrehozott vagy módosított tartalmakat a rendszer automatikusan menti a háttérben a OneDrive személyes tárhelyen.

Az utolsó pontot röviden kifejtve pedig elmondhatjuk, hogy ha egy új fájlt hozunk létre az asztali környezetbe telepített irodai szoftverek egyikével, akkor lehetőségünk nyílik rögtön a személyes felhő tárhelyünkbe vagy adott TEAMS csapat megosztott tárhelyére menteni dokumentumokat. Asztali környezetben futtatott Office 365 Word vagy Excel program esetében is lehetőségünk nyílik a dokumentumok közös szerkesztésére, akár csak a böngészőből indított szerkesztés esetén is, noha a böngészős Word/Excel/PowerPoint-variánsok funkcionalitásban valamelyest limitáltabbak az asztali verziókhöz képest.

Ha a telepített Office 365-ös Word, PowerPoint és Excel programok esetében bekapcsoljuk az „Automatikus mentés” funkciót, akkor az éppen használt szerkesztőprogram néhány másodpercenként automatikusan elmenti a fájlokat munka közben: kiválaszthatjuk, hogy helyi tárhelyre vagy OneDrive-tárhelyre történjen az automatikus mentés. Ezzel nagyobb eséllyel kerülhetjük el az adatvesztést egy áramkimaradás vagy más hardveres probléma esetén. Természetesen

ennek feltétele a működő internetkapcsolat és a háttérben futó OneDrive-kliens is!

A legtöbb célszoftverrel – melyek nem a Microsoft termékei – ugyanúgy tudunk a felhőben lévő tárhelyünkre menteni, és onnan szerkesztésre fájlokat megnyitni, hiszen a OneDrive-kliens segítségével tállózhatóvá válik a saját számítógépünk meghajtói közt a Cloud/megosztott csoporttárhelyünk. Mindössze arra kell ügyelnünk, hogy adott célszoftver (például SPSS, LibreOffice, Notepad++ stb.) nem képes néhány másodpercenként automatikus mentésre: bizonyos időközönként saját magunknak kell megnyomni a mentés gombot vagy az azzal ekvivalens billentyűkombinációt.