



Száldobágyi Ádám

Kutatási adatkezelés

Let's be Open! – Nyílt tudomány: lehetőségek és kihívások



Az adat az új olaj koncepció

- “Data is the new oil. It’s valuable, but if unrefined it cannot really be used. It has to be changed into gas, plastic, chemicals, etc to create a valuable entity that drives profitable activity; so must data be broken down, analyzed for it to have value.”
- 2006, Clive Humbley

Kutatási adat

Photo by [Scott Graham](#) on [Unsplash](#)

MATE Let's Be Open – 2024.október 16.



Mit nevezünk kutatási adatnak?

Kutatási adatnak nevezünk minden olyan tényszerű adatot, amelyet a kutatók közössége hoz létre és tárol azért, hogy a kutatási eredményeiket elfogadják¹.

Típusai:

- Szöveges dokumentum
- Táblázat
- Labor jegyzőkönyv
- Hang vagy videó felvétel
- Fénykép
- Vizsgálati eredmény
- Kísérleti eredmény
- Modell, algoritmus vagy script
- Metaadat
- stb.



Kép forrása: <https://www.salesforce.com/hub/analytics/how-to-use-big-data-analytics-tools/>



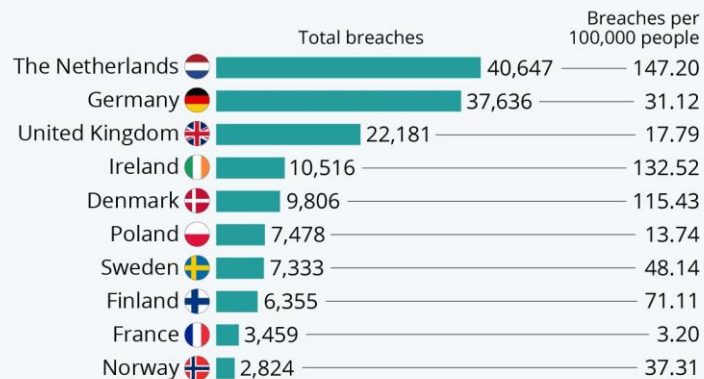
Miért fontos, hogy a kutatási adatainkat jól kezeljük²?

- Kutatási eredmények így igazolhatók
- Adatvesztés elkerülése
- Időhatékonyság
- Az adataink biztonságban legyenek
- Megkönnyítsük az adatok újra felhasználását
- Növelhetjük cikkünk láthatóságát ezzel impaktját

Kép forrása: <https://www.swissarmylibrarian.net/2012/01/24/lost-and-found-flash-drives/>

The Countries With The Most GDPR Data Breaches

Personal data breaches notified per EEA jurisdiction
(May 25, 2018 to Jan 27, 2020)*



* EEA - European Economic Area (EU-28 + Norway, Iceland, Liechtenstein).
Source: DLA Piper

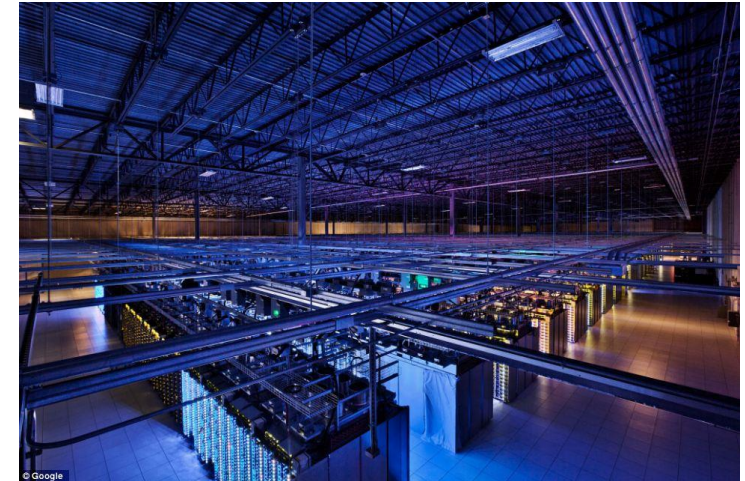


Adatok tárolása⁵

- Rövid távú:
 - Saját meghajtó
 - Flash drive
 - Felhő
- Hosszú távú:
 - Saját weboldal
 - Intézményi repozitórium, pl.: DEA
 - Diszciplináris repozitórium, pl.: NCBI Gene
 - Interdiszciplináris adatrepozitórium, pl.: Dataverse, Zenodo

Érdemes három helyen őriznünk biztonsági mentést az adatainkról: here, near and far.

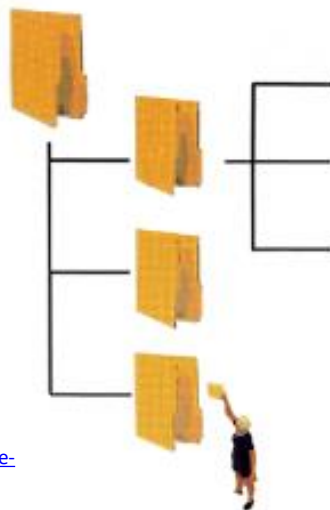
Kép forrása: <https://whatsabyte.com/internet/best-cloud-storage-services/>



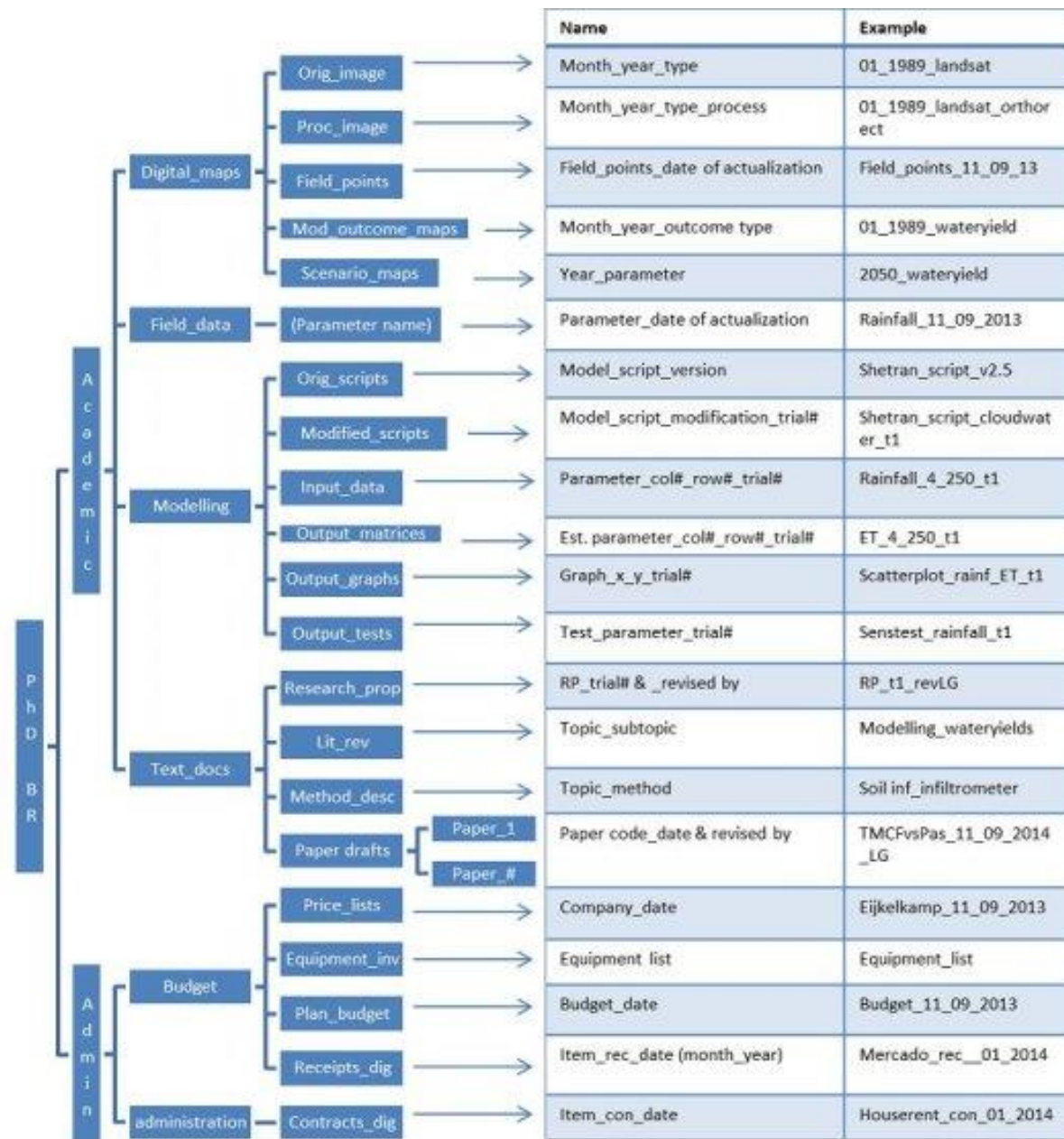
Kép forrása: <https://www.kcl.ac.uk/library/researchsupport/research-data-management/preserve/deposit-your-data.aspx>

Fájlstruktúra választása

- A mappákat és adatfájlokat úgy rendszerezünk és nevezzük el, hogy mások is értelmezni tudják azokat
- Fájl formátumokat úgy válasszuk meg, hogy mások is hozzá tudjanak férni



Kép forrása: <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-and-folder-structure>



Example of a proposed folder arrangement and file naming practice for a PhD project in Earth System Science. <https://www.wur.nl/en/Value-Creation-Cooperation/WDCC/Data-Management-WDCC/Doing/Organising-files-and-folders.htm>

Metaadat

„Adat az adatról”

Az adatok leírására szolgáló dokumentáció

Metaadat szabványok:

- Dublin Core,
- Darwin Core,
- Data Documentation Initiative,
- Metadata Encoding and Transmission Standard

Magyarországon a legelterjedtebb szabvány a Dublin Core.

A legtöbb repozitáriumnak megvan a saját szabályzata a metaadatokra, de ezek általában átjárhatók egymás között vagy kiegészíthetők a kívánt tartalommal.

Ezzel az azonosítóval hivatkozhat erre a dokumentumra idézés vagy hiperhivatkozás esetén:
<http://hdl.handle.net/2437/102285>

dc.contributor.author	Lakatos, Ákos
dc.contributor.author	Csik, Attila
dc.contributor.author	Langer, Gábor Antal
dc.contributor.author	Erdélyi, Gábor
dc.contributor.author	Katona, Gábor
dc.contributor.author	Daróczy, Lajos
dc.contributor.author	Vad, Kálmán
dc.contributor.author	Tóth, József
dc.contributor.author	Beke, Dezső László
dc.date.accessioned	2011-02-10T11:57:11Z
dc.date.available	2011-02-10T11:57:11Z
dc.date.issued	2009
dc.identifier.citation	Vacuum. -84 : 1 (2009), p. 130-133. -Vacuum. - 0042-207X
dc.identifier.issn	0042-207X
dc.identifier.uri	http://hdl.handle.net/2437/102285
dc.language	eng
dc.title	Investigation of failure mechanisms at Ta and TaO diffusion barriers by secondary neutral mass spectrometry
dc.type	folyóiratcikk
dc.rights.access	open access pre-print és post-print (24h embargó)
dc.identifier.opac	http://webpac.lib.unideb.hu:8082/ebib/CorvinaWeb?action=cclfind&resultview=long&ccltext=idno+BIBFORM013101
dc.subject.mab	Természettudományok
dc.subject.mab	Fizikai tudományok
dc.rights.owner	Elsevier
dc.identifier.doi	http://dx.doi.org/10.1016/j.vacuum.2009.06.007
dc.date.updated	2017-12-01T09:03:09Z
dc.date.pasync	2017-12-01T09:03:09Z

A tétel áttekintő adatai

Keresés a DEA-ban

OK

- ☒ Keresés a DEA-ban
☐ Ez a gyűjtemény

[Összetett keresés](#)

Böngészés

Teljes DEA

Kategóriák és gyűjtemények
Megjelenés dátuma
Szerző
Cím
Tárgyszó

Ez a gyűjtemény

Megjelenés dátuma
Szerző
Cím
Tárgyszó

Felhasználói adataim



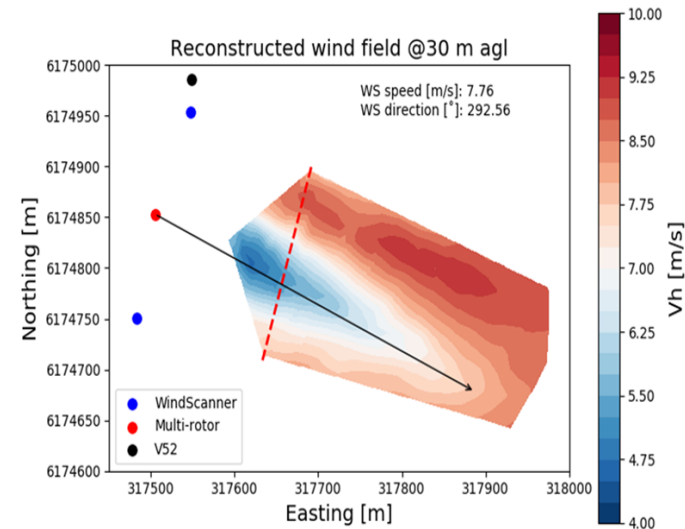
Kijelentkezés
Bejelentkezve
Feltöltések

iDEa Tudóstér

- Publikációim

Metadata type:

Title:	Wake plot of wind turbine
Keyword:	wake plot, multi-rotor, wind
Creator:	Nikola Vasiljević
Location:	Risø, Denmark
Date:	14 August 2018
Wind turbine:	Vestas Multi-rotor
Instrument:	Long-range windScanner
Usage rights:	Creative Commons
Project ID:	Vestas_DTU_multi_2018
Project period:	01 November 2017 – 31 July 2019



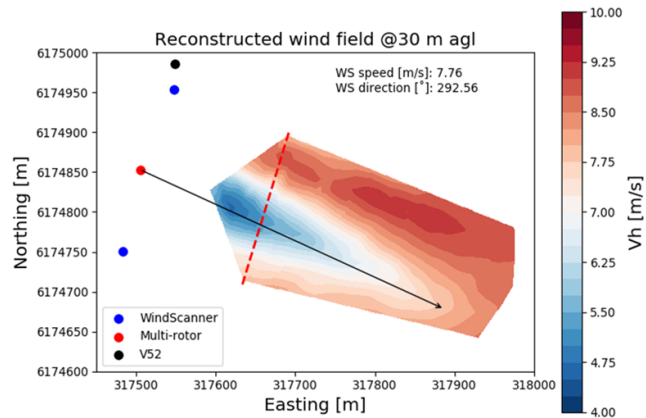
Metadata type:

Administrative

Title: Wake plot of wind turbine
 Keyword: wake plot, multi-rotor, wind
 Creator: Nikola Vasiljević

Location: Risø, Denmark
 Date: 14 August 2018
 Wind turbine: Vestas Multi-rotor
 Instrument: Long-range windScanner

Usage rights: Creative Commons
 Project ID: Vestas_DTU_multi_2018
 Project period: 01 November 2017 – 31 July 2019



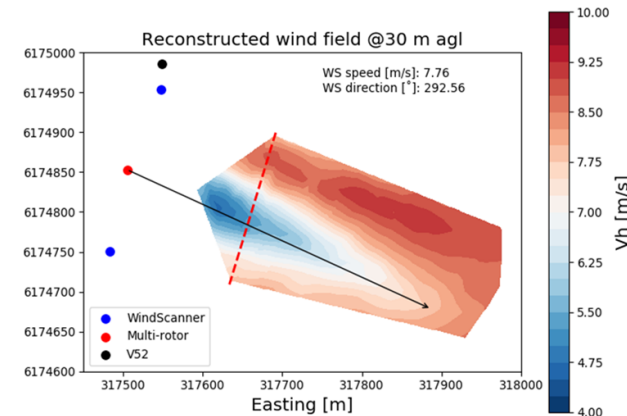
Metadata type:

Descriptive

Title: Wake plot of wind turbine
 Keyword: wake plot, multi-rotor, wind
 Creator: Nikola Vasiljević

Location: Risø, Denmark
 Date: 14 August 2018
 Wind turbine: Vestas Multi-rotor
 Instrument: Long-range windScanner

Usage rights: Creative Commons
 Project ID: Vestas_DTU_multi_2018
 Project period: 01 November 2017 – 31 July 2019



Fájl elnevezések

Törekedjünk arra, hogy a projekten dolgozó összes munkatárs azonos fájl elnevezéseket használjon.

Az adatkezelési tervben tüntessük fel a metódust.

Fontolja meg az alábbi információk némelyikének felvételét a fájlneveibe, de felvehet bármely olyan információt, amely lehetővé teszi a fájlok megkülönböztetését:

- Projekt vagy kísérlet neve vagy betűszava
- Hely / térbeli koordináták
- Kutató neve / kezdőbetűi
- A kísérlet dátuma vagy dátumtartománya (YYYYMMDD or YYMMDD)
- Az adatok típusa
- A fájl verziószáma
- Hárombetűs fájlkiterjesztés az alkalmazás-specifikus fájlokhoz

Különleges karakterek, például ~! @ # \$% ^ & * () ` ; <>? , [] {} "és | kerülni kell!

Every Presenter in the world



new.pptx

newfinal.pptx

newfinalfinal.pptx



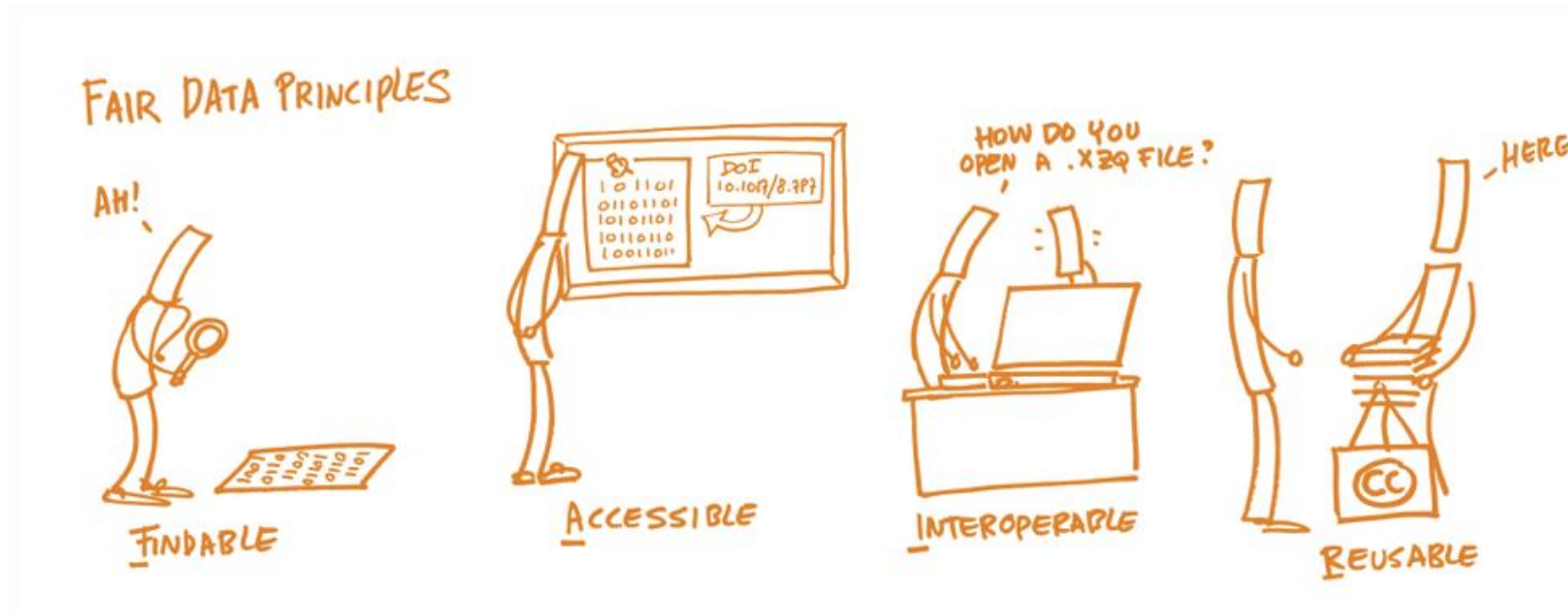
newfinalest
final.pptx

newfinalest
finalforsure.pptx

newfinalscrewit
thiswilldo.pptx

<https://smoothav.com.au/file-names-for-your-presentation/>

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



FAIR

kutatási adatkezelés

Az adatok nyílt hozzáférését EU-s törvények is szabályozzák, amelyek szerint az EU-országok adattermésére – függetlenül attól, hogy rendelkezik-e szerzői jogvédelemmel vagy nem – ún. sui generis (hagyományos oltalmi formák közé nem beilleszthető) szerzői jogi védelmet határoz meg, ezért ezek az adatok (legalábbis az EU-országokban) nem használhatók a szerzői jog tulajdonosának engedélye nélkül.

F

FINDABLE

Megtalálható



Emberek és gépek számára egyaránt
Egyedi azonosítóval rendelkezzen
Metaadatokkal gazdagon ellátott
Metaadatai repozitóriumban legyenek tárolva

A

ACCESSIBLE

Elérhető



Emberek és gépek számára könnyű hozzáférési vagy letöltési lehetőség.
Sztenderd kommunikációs protokoll segíti az adatok elérhetőségét.
A metaadatok elérhetők az adatok megsemmisítése után is.

I

INTEROPERABLE

Együttműködő



Lehetőséget ad az adatok változtatására, cseréjére
A (meta)adatok a FAIR alapelveit követő szókészleteket használnak
A (meta)adatok tartalmazzák a más (meta) adatokra vonatkozó minősített hivatkozásokat.

R

REUSABLE

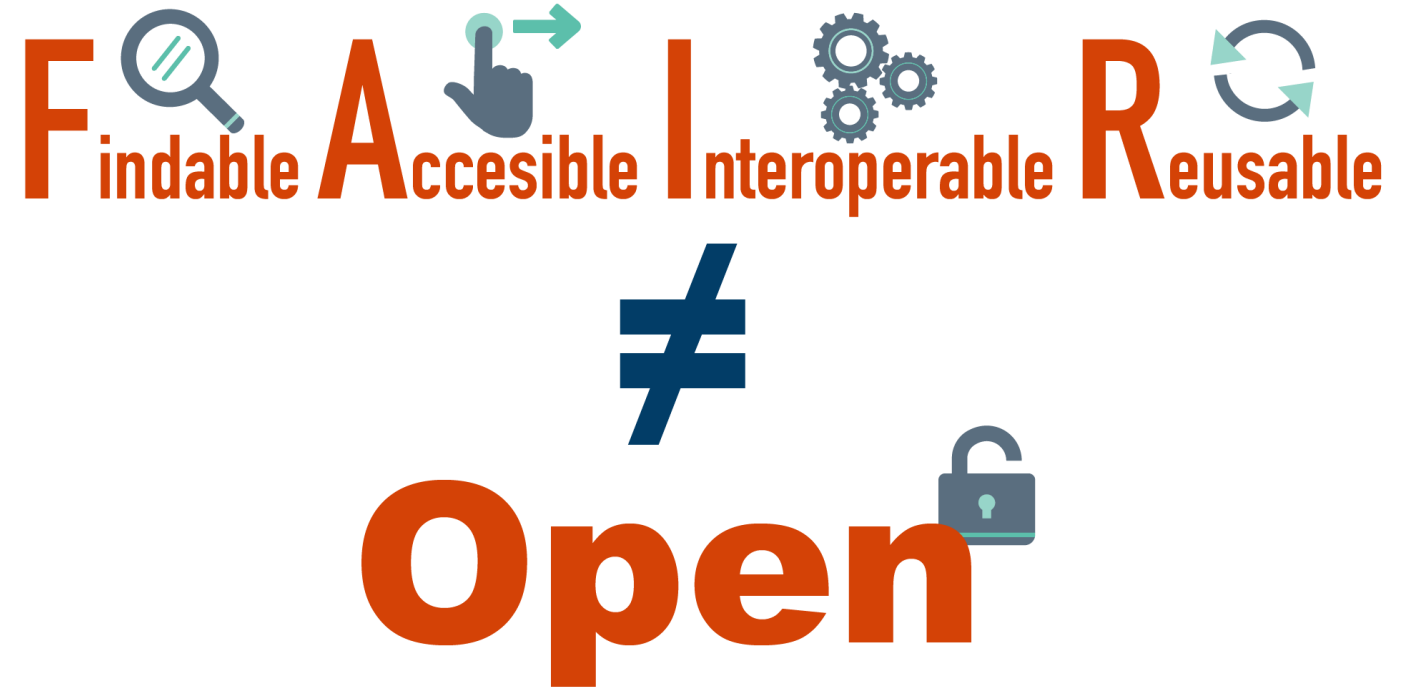
Újrafelhasználható



Lehetőséget ad az adatok jövőbeni felhasználására
A (meta)adatok egyértelmű és hozzáférhető adathasználati engedélyekkel kerülnek kiadásra.
A (meta)adatok megfelelnek a domain-specifikus közösségi szabványoknak

A FAIR leglényegesebb elemei

- Használjunk állandó azonosítókat
- Használjunk szabványosított metaadat formátumokat
- Töltsük fel az adatokat egy megbízható repozitóriumba
- Lássuk el licenszekkel az adatokat
- Használjunk egyszerű adatformátumokat



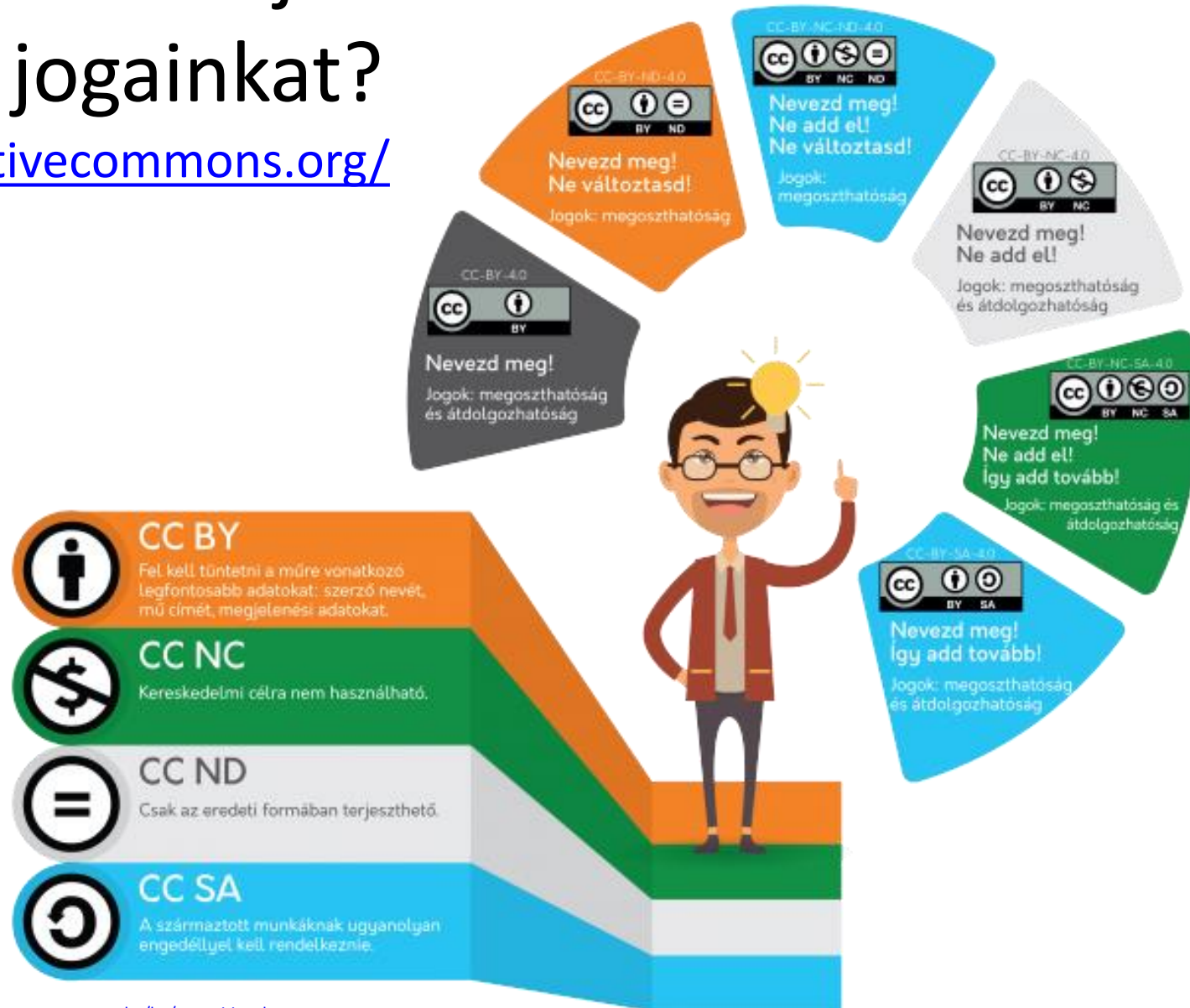
Forrás: <https://libraries.mit.edu/news/data-necessarily/31421/>

Szerzői jogok védelme



Hogy biztosíthatjuk a szerzői jogainkat?

- <https://creativecommons.org/>



Hogy válasszunk licenc-t?

LICENSE CHOOSER

Follow the steps to select the appropriate license for your work. This site does not store any information.

1 Do you know which license you need?

- ☐ Yes. I know the license I need.
- ☐ No. I need help selecting a license.

NEXT

2 Attribution

3 Commercial Use

4 Derivative Works

5 Sharing Requirements

6 Confirm that CC licensing is appropriate

7 Attribution Details

Kép forrása: <https://chooser-beta.creativecommons.org/>

License Clearance Tool



Our wizard guides you through the needed information for clearing your resources, facilitating openness and reproducibility in research.

[Login via NI4OS AAI](#)

[Explore the tool as a guest user](#)

Személyes adatok kezelése

Személyes adatról akkor beszélhetünk, ha az információ természetes személyre utal, és a személy beazonosítható az adat alapján⁶.

A megfelelő védelemről gondoskodni kell!

Anonimizálás:

Amnesia online program segítségével
<https://amnesia.openaire.eu/amnesia/>



Kép forrása: <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/5.-Protect/Ethics-and-data-protection>



Kép forrása: <https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/5.-Protect/Anonymisation>

Kutatási adatkezelési terv

ADAT MENEDZSMENT TERV (DMP)

MEGHATÁROZZA ÉS TARTALMAZZA

Egyszerű, 1-2 oldalas összefoglaló, amely leírást ad arról, hogy a kutatás során gyűjtött adatokat hogyan kell kezelni.
Egy folyamatosan változó, a **kutatási szakaszokhoz alkalmazkodó** vagyis „élő” dokumentum.
Általában az első verziót a kutatás első 6 hónapjában kell elkészíteni, és **folyamatosan frissíteni**, ha kutatás közben jelentős változás történik.

A kutatás alatt várhatóan létrejövő adatokat

Ki hozza létre és milyen típusú adatok keletkeznek a kutatás során?

A kezelés a megőrzés és a fenntarthatóság módját

Hogyan tárolják ezeket az adatokat, ki és milyen módon férhet hozzájuk?

A biztonság és a fenntarthatóság kérdése

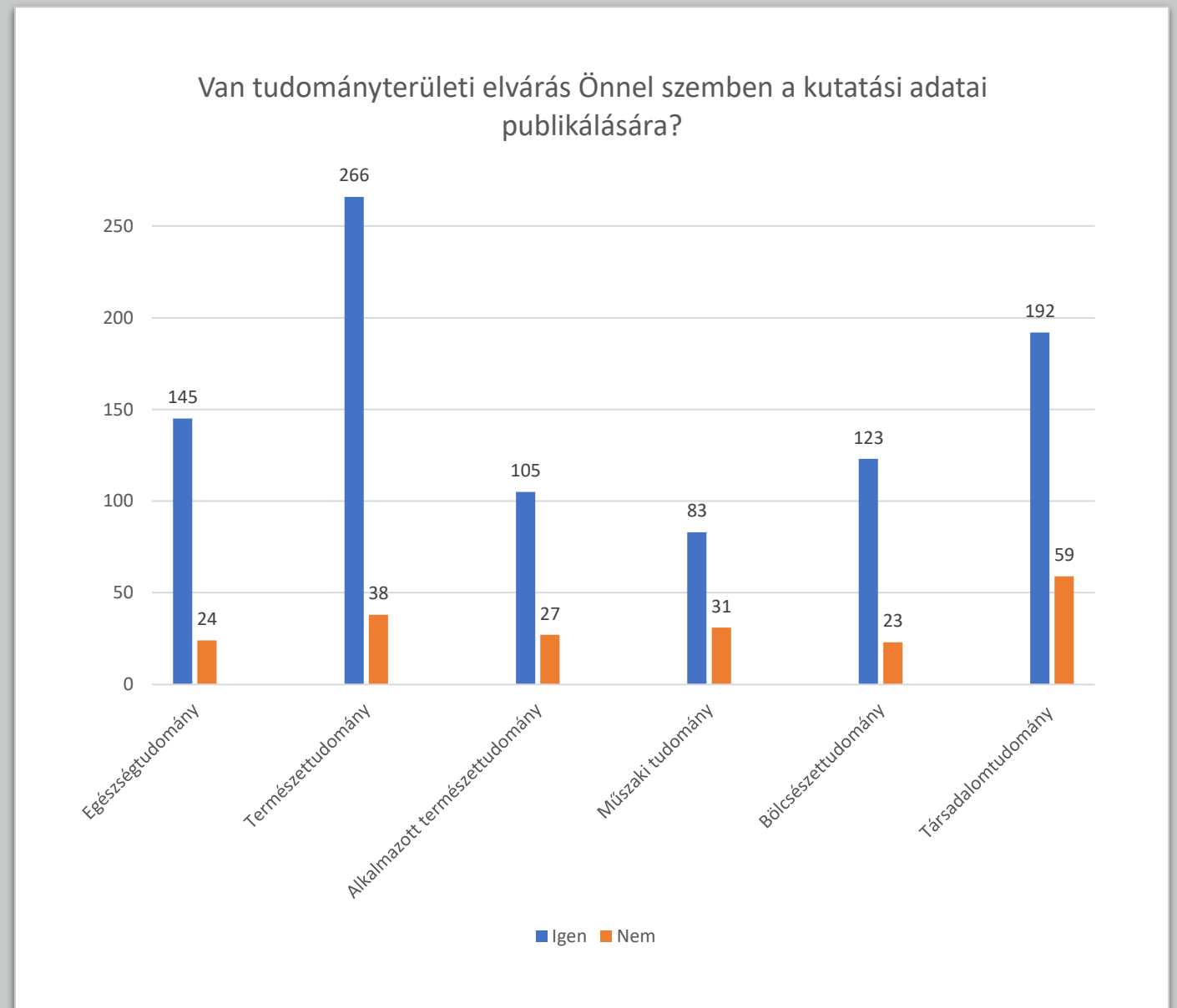
Az adatok mely része nyitott és hogyan, projekt után ki férhet hozzá az adatokhoz, milyen jogokkal és meddig?

Az adatmenedzsment költségeit, melyek elszámolhatók.

Az adatok minőségének leírását

Kutatási adatkezelési elvárások

- 2021-ben az OTKA és az Élvonal pályázatok kiírásában kritériumként jelent meg a kutatási adatkezelési terv.
- 2021. október 18-án a Nemzeti Kutatási Fejlesztési és Innovációs Hivatal kezdeményezésével, az Innovációs és Technológiai Minisztérium támogatásával egy állásfoglalás született, amelyben kiemelt szerepet kap a FAIR és a CARE adatkezelés, továbbá a hazai kutatási adat-kezelési infrastruktúra kiépítésének támogatása.
- 2024 Nemzeti Kutatási Kiválósági Program European Research Council DMP



Mit szokott tartalmazni a kutatási adatkezelési terv³?

- Adminisztratív adatok:
 1. Projekt neve
 2. Projekt rövid jellemzése
 3. A fő kutató(k) neve(i)
- Adatgyűjtés:
 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



Kép forrása: <https://www.dreamstime.com/stock-photography-too-much-information-one-person-image30561332>

De lehet másképp is csinálni

Online kitölthető adatkezelési tervek:

- DMPonline: <https://dmponline.dcc.ac.uk/>
- ARGOS: <https://argos.openaire.eu/home>
- Data Stewardship Wizard: <https://ds-wizard.org/>



Test2

Project Details Plan overview **Write Plan** Share Download

expand all | collapse all

0/13 answered

Data Collection (0 / 2)



Documentation and Metadata (0 / 1)



Ethics and Legal Compliance (0 / 2)



Storage and Backup (0 / 2)



Selection and Preservation (0 / 2)

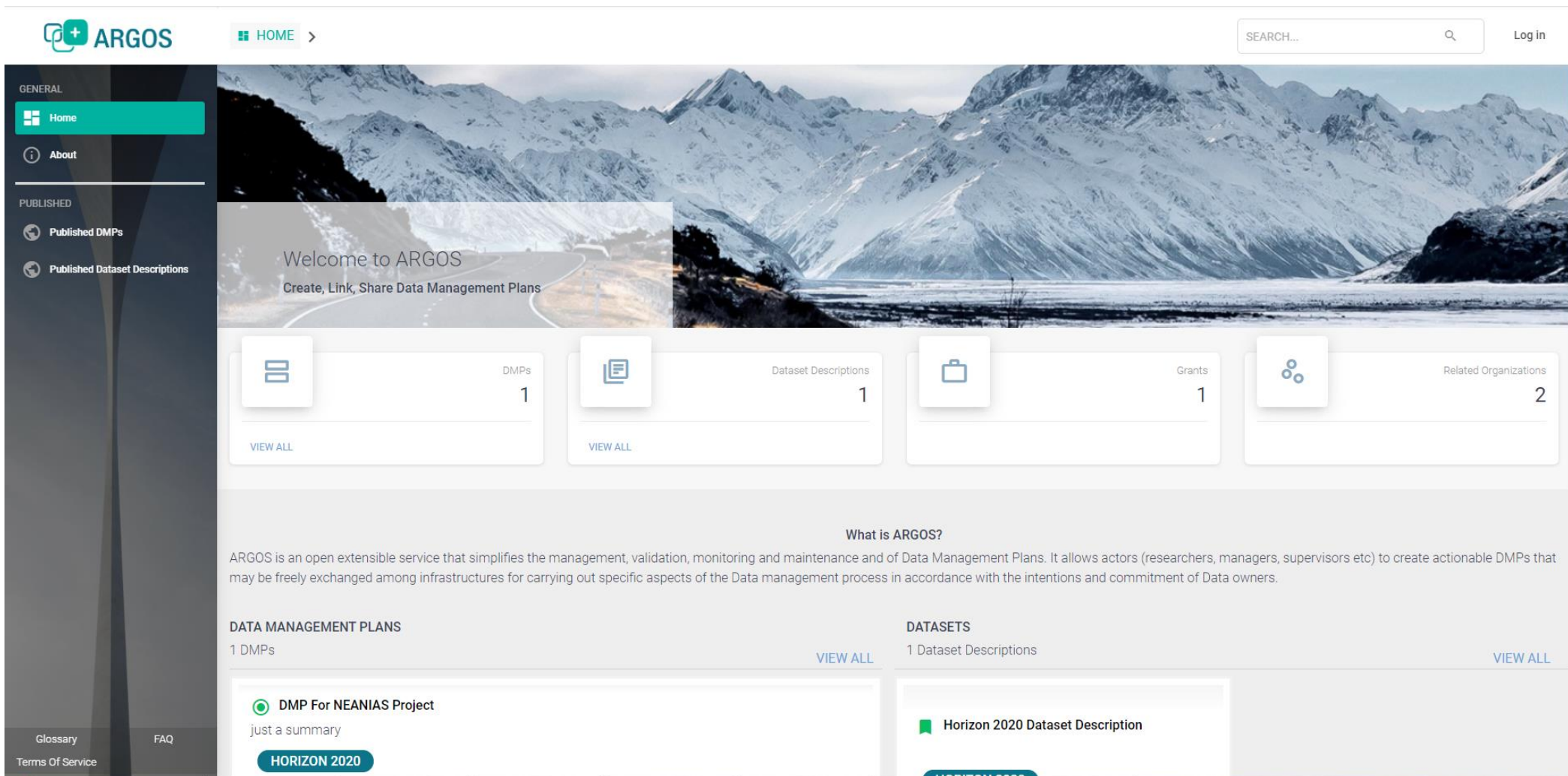


Data Sharing (0 / 2)



Responsibilities and Resources (0 / 2)





DMP terv készítése az ARGOS segítségével

 DS Wizard

Knowledge Models

Projects

Help

Ádám Száldobágyi

Collapse sidebar

Ádám

Questionnaire

TODOs

Metrics

Preview

Documents

Settings

Share

Current Phase

Before Submitting the Proposal

Chapters

I. Design of experiment3

II. Data design and planning5

III. Data Capture/Measurement3

IV. Data processing and curation1

V. Data integration1

VI. Data interpretation1

VII. Information and insight8

I. Design of experiment

Before you decide to embark on any new study, it is nowadays good practice to consider all options to keep the data generation part of your study as limited as possible. It is not because we can generate massive amounts of data that we always need to do so. Creating data with public money is bringing with it the responsibility to treat those data well and (if potentially useful) make them available for re-use by others.

1

Is there any pre-existing data?

Are there any data sets available in the world that are relevant to your planned research?

☒ Desirable: Before Submitting the Proposal

☒ Data Stewardship for Open Science: [atq](#)

☒ a. No

☐ b. Yes

Clear answer

You know that this is very unlikely? This question is not only about data sets that are similar to what you want to determine yourself, but also reference data or data that should be mined from the existing literature. Further, it is very likely that you will refer to related data, e.g. other databases where you usually "quickly look something up", but that could maybe be properly integrated, especially if you need to do such lookups multiple times.

2

Will reference data be created?

Will any of the data that you will be creating form a reference data set for future research (by others)?

☒ Desirable: Before Submitting the Proposal

☒ Data Stewardship for Open Science: [rbz](#)

☐ a. No

☐ b. Yes

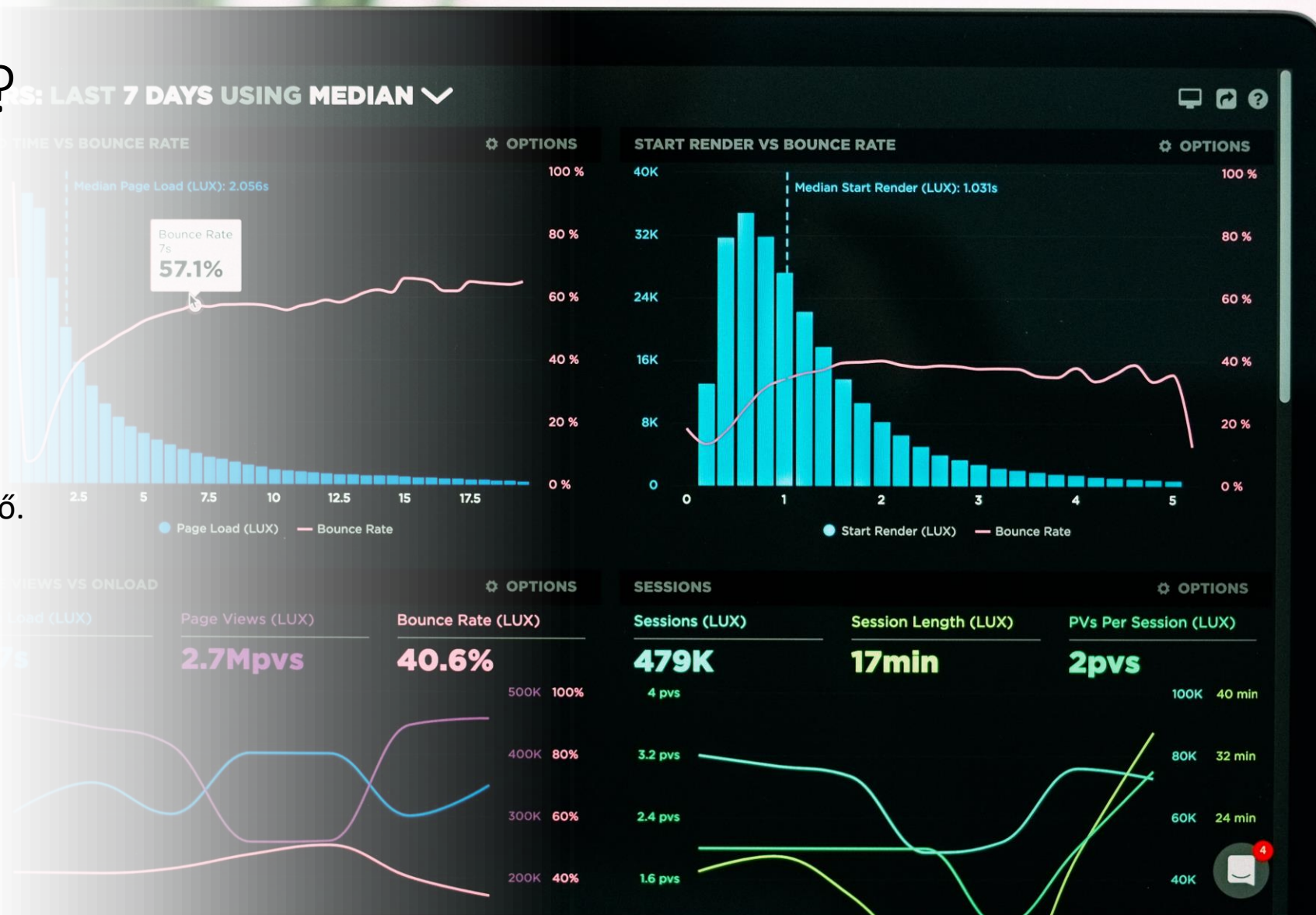
DMP terv készítése a DS Wizard segítségével

MATE Let's Be Open – 2024.október 16.

Adatrepozitórium

Hogyan válasszunk adatrepozitriumot?

- A tudományos kutatás során keletkezett publikálni kívánt kutatási adatok tárolására használt platformok.
- Első lépésként az adott tudományterületnek dedikált adattárat keressünk. Ebben segít re3data.org adatrepozitrium kereső.
- Ha nem találtunk ilyet, akkor az intézményünk által üzemeltetett használjuk.
- Használjunk interdiszciplináris repozitriumot, például: [Zenodo](https://zenodo.org), [Figshare](https://figshare.com), [DRYAD](https://dryad.org)



Köszönöm a figyelmet!

Száldobágyi Ádám
szaldobagyi.adam@lib.unideb.hu

Felhasznált irodalom

1. <https://www.CESSDA.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/1.-Plan/Research-data>
2. <http://www.dcc.ac.uk/resources/how-guides/develop-data-plan#Why%20develop>
3. <http://www.dcc.ac.uk/resources/how-guides/develop-data-plan#Why%20develop>
4. http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
5. <https://www.CESSDA.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/4.-Store/Storage>
6. https://openscience.hu/sites/default/files/oldal/2602/szemelyes_adatok_20181013_magyar.pdf
7. EOSC: NI4OS projekt
(https://training.ni4os.eu/pluginfile.php/3631/mod_resource/content/0/HU-NI4OS%20-%20EOSC.pdf)
8. Borítókép: Photo by [Luke Chesser](#) on [Unsplash](#), Photo by [Markus Winkler](#) on [Unsplash](#)