

Sajti Szilárd

2025-01-29

Kutatásiadat-kezelés, HUN-REN Adatrepozitórium Platform

Minden adat, amely kutatás eredményét támasztja alá:

- Táblázat,
- Szöveg, kép, hang, videó,
- Szoftver,
- Kézirat, napló,
- Labor jegyzőkönyv

A publikált kutatások jelentős része nem reprodukálható, mert nem áll rendelkezésre elegendő adat az ellenőrzéshez.

- Sokszor nem publikálják az adatokat
- Nem elérhetőek, vagy nehezen kezelhető formátumban
- Elérhető volt valamikor, de már csak egy üres link
- Nehéz értelmezni hiányos dokumentáció miatt

- Nyílt bíráló
- Nyílt forráskód
- Nyílt jegyzőkönyv
- Kutatási adatok repositálása, FAIR adatkezelés

FAIR adatkezelési elvek

Megtalálhatóság (emberek és számítógépek számára is):

- Sok metaadatra van szükség.
- Globális állandó azonosító (persistent identifier, PID):




- Mind az adatokhoz, mind a hozzájuk kapcsolódó metaadatokhoz szükség van egyedi azonosítókra
- A (meta)adatokat regisztrálni kell.



FAIR adatkezelési elvek

Hozzáférhetőség:

- Megfelelő szabványos, nyílt, ingyenes protokoll a (meta)adatok elérésére, ha szükséges autentikációval.
- A metaadatok akkor is elérhetőek kell legyenek, ha az adatok már nem

Findable 

Accessible 

Interoperable 

Reusable 

- Szabványosság, Együttműködési képesség, együtthasználhatóság
- Az adatokat gyakran más adatokkal, programokkal, munkafolyamatokkal használjuk fel, azokhoz kapcsolódnak.
 - Szabványos formátum, leírási mód, nyelv
(mit tartalmaz az adatcsomag).
 - (Kontrollált) szótárak, Ontológiák



Újrafelhasználhatóság

(Ez lenne a főcél)

- Alapos és pontos metaadatokat kell kapcsolni az adatokhoz.
- Világos adatfelhasználási engedély (licenc)
- Az adatok létrejöttének pontos leírása metaadatokkal, ellenőrizhetőség, visszakövethetőség.



A kutatási adatokra vonatkozó tények és információk összessége, adatok az adat(ok)ról.

Legáltalánosabb (Dublin Core  <https://www.dublincore.org>):

- adatcsomag neve, eredet, dátum,
- hozzáférési és/vagy felhasználási feltételek,
- absztrakt,
- szerzők neve

A kutatási adatokra vonatkozó tények és információk összessége, adatok az adat(ok)ról.

Legáltalánosabb (Dublin Core



<https://www.dublincore.org>):

- abstract
- accessRights
- accrualMethod
- accrualPeriodicity
- accrualPolicy
- alternative
- audience
- available
- bibliographicCitation
- conformsTo
- contributor
- *coverage*
- created
- creator
- date
- dateAccepted
- dateCopyrighted
- dateSubmitted
- *description*
- educationLevel
- extent
- *format*
- hasFormat
- hasPart
- hasVersion
- *identifier*
- instructionalMethod
- isFormatOf
- isPartOf
- isReferencedBy
- isReplacedBy
- isRequiredBy
- issued
- isVersionOf
- *language*
- license
- mediator
- medium
- modified
- provenance
- *publisher*
- references
- *relation*
- replaces
- requires
- *rights*
- rightsHolder
- *source*
- spatial
- *subject*
- tableOfContents
- temporal
- *title*
- *type*
- valid

Add Resource:

Step 1: Resource Configuration

Step 2: Metadata Profile

Step 3: Resource Metadata

Step 4: Overview & Confirm

Metadata Profile: ⓘ * Microscopy in a Rotating Magnetic Field 📄 Create

Choose the Metadata Profile carefully, as it cannot be changed once it is assigned to a resource.

Creator * 🔒 👁 +

Student 🔒 👁 +

Date and time of measurement * 📅 🔒 👁 +

Production method * 🔒 👁 +

Sample ID * 🔒 👁 +

SFB 985 Sample Management PID 🔒 👁 +

Measurement Series * 🔒 👁 +

Solvent * 🔒 👁 +

Field strength [mT] * 🔒 👁 +

Rotation rate [gps] * 🔒 👁 +

Frame rate [fps] * 🔒 👁 +

Rotation [°] * 🔒 👁 +

Metaadatséma:
[https://
coscine.rwth-
aachen.de](https://coscine.rwth-aachen.de)

Egy minta metaadatai példa

Sample Field_name	example_input	Comment
ID*	MySample	ID for the object, be it sample, experiment, sim...
External/alias ID		Alternative ID for this object, e.g. if one is using a different name for another project
User*	Niklas Siemer	The user responsible for this digital record.
Date*	01.01.2022	Date of the current digital record; Default is today.
Affiliation	MPIE	
DOIs		To associate publications produced using this object.
Location	Right drawer	Where is this in the real world.
Parent ID	S8	The super-sample from which this came.
Preparation IDs	Polishing_fine	What was done to this sample that differentiates it from its parent, in order of operation. Multiple routines should only be given for perfectly contiguous sample chains.
Comments		

https://www.mpie.de/4699570/metadata_schemes

Szabványos metaadatokra van szükségünk, azt fogják tudni mások is használni, így lesznek a fogalmaink egyértelműek.

Vocabulary: [DCMI Metadata Terms](http://purl.org/dc/terms/)

URI: <http://purl.org/dc/terms/accessRights>

Label: Access Rights

Definition: Information about who access the resource or an indication of its security status.

Comment: Access Rights may include information regarding access or restrictions based on privacy, security, or other policies.

Type of Term: Property

Kontrollált szótár – egyedi nevek

Taxonómia (rendszerteran) – hierarchikus rendszer

Ontológia – Relációk is vannak a különböző fogalmak között

EMMO (European Materials Modelling Ontology <https://emmo-repo.github.io/>),

Általános (mereotopológia, szemiotika), de még fejlesztés alatt, egyelőre kevés példa.

Ráépül már:

- CHAMEO (Characterisation Methodology Domain Ontology),
- BTO (Battery testing ontology)

- Adattípus, adatformátum, könyvtárszerkezet, file elnevezési konvenciók, verziókövetés
- Metaadatok
- Megosztás módja, licenc
- Tárolás módja, biztonsági mentés
- Költségek
- A terv változhat, 6 hónaponta ajánlott is felülvizsgálni
- Online programok a terv készítéshez: <https://argos.openaire.eu>, <https://ds-wizard.org/>
- Licenc kiválasztásához: License Clearance Tool (<https://lct.ni4os.eu/lct/>)

Ajánlott és nem ajánlott fileformátumok

Adattípus	Ajánlott fájlformátum	Elfogadható fájlformátum	Nem ajánlott fájlformátum
Szöveges adat	.pdf(PDF/A), .txt, .tex, .odt, .xml, .csv, .epub, .md	.docx, .pdf, .html, .rtf, .pptx	.doc, .ppt
Táblázatos adat	.csv, .ods, .xml, .tab, .dta	.xlsx, .mdb	.xls, .xlsb
Kép	.jpg, .png, .svg, .tif, .jp2	.psd, .gif, .dng, .nef,	.ait, .psd
Hang	.aiff, .bwf, .flac, .wav,	.aac, .mp3, .mp4, .gif	
Video	.mp4, .mkv, .mj2, .avi	.ogm, .webm	.mov
Térinformatikai adat CAD/vektor/raszter adat	.tif (GeoTiff), .gml, .svg, .dxf, .dwg, .GeoJSON, .x3d	.psd, .jpf, .eps, .mif, .kml,	
Generikus adat	.xml, .json, .ref		
Adatbázis	.tsv, .csv, .db, .sqlite, .sql	.xlsx	.xls
Kutatási adat	.mat, .hdf, .h5, .cdf, .sav, .py	.mdb, .por, .txt	.rdata

Forrás:

<https://urn.fi/urn:nbn:fi-fe2020100578096>

Adatcsomag értékelők, mennyire FAIR



F-UJI is a web service to programmatically assess FAIRness of research data objects at the dataset level based on the FAIRsFAIR Data Object Assessment Metrics [↗](#)

[Click here to assess a dataset](#)



Your first step towards your FAIR data(set)

2021. november 1. SZTAKI, TK, Wigner FK

- Harvard Dataverse alapú adatrepozitórium
- Elsősorban HUN-REN kutatóhálózat intézményei, valamint a hazai felsőoktatási intézmények munkatársai számára.

(Amennyiben az adott intézmény nem rendelkezik saját intézményi tárolóval az ARP adatrepozitóriumában, a munkatárs közvetlenül vegye fel a kapcsolatot a szolgáltatóval.)

- Adatcsomag kvóta 100 GB
- Intézményi kvóta is van ez kb. 10 TB. (19 HUN-REN intézmény)
- 1 × tárolás, 2 × georedundáns tárolás (SZTAKI, Wigner FK), 3 × geored. (+ Wigner mágneses szalag) A georedundáns tárolást kérni kell.
- Lehetséges embargót, vendégkönyvet beállítani
- jelenleg 10 évre van tervezve
- Olyan adatokat kellene letenni, amelyek előreláthatóan már nem fognak változni. (Korrekcióra, új verzióra van lehetőség)
- Ha van elfogadott szakterületi adatrepozitórium, akkor inkább oda kellene elhelyezni az adatokat.
- Napi munkára inkább az intézményi felhőt kellene használni.

- ARP Közös Kereső magyar vonatkozású adatok keresése az ARP mellett: a Debreceni Egyetem Adatrepozitóriumában, TK KDK Adatrepozitóriumában, és Zenodoban.
- ARP Sémátár felhasználói metaadatok, metaadatsémák definiálására nyújt lehetőséget (Cedar alapon).
- ARP AROMA (RO-Crate Manager) az adatcsomagon belüli fileokat is el lehet látni metaadatokkal.