



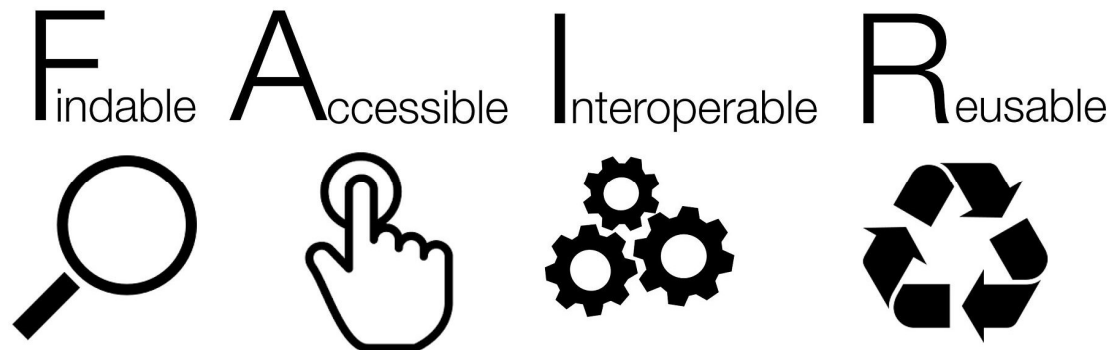
Az RO-Crate alapú kutatási objektum csomagolás keretrendszere az ELKH ARP platformban

Tóth Zoltán
SZTAKI – DSD
toth.zoltan@sztaki.hu

Előadásvázlat

- FAIR irányelvek;
- FAIR Digital Objects (FDO);
- A RO-Crate ismertetése;
- RO-Crate az ELKH ARP platformban.

FAIR irányelvek

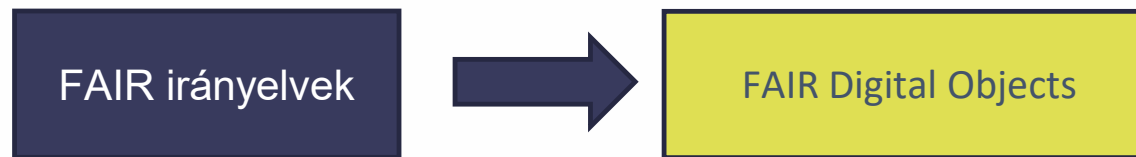


[*SangyaPundir – CC BY SA 4.0](#)

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>

Bár az open science szempontjából (is) nagyon fontos irányelvekről van szó, mert az irányelvek követése megkönnyítheti az adatok általános hozzáférését, kereshetőségét és használatát, az eredeti cél az volt, hogy ne csak ember, hanem gép is megtalálhassa és esetlegesen automatikusan feldolgozhassa a kutatási adatokat (nem pedig az, hogy mindenki hozzáférjen minden kutatási adathoz).

FAIR Digital Objects (FDO)



European Commission, Directorate-General for Research and Innovation, *Turning FAIR into reality : final report and action plan from the European Commission expert group on FAIR data*, Publications Office, 2018, <https://data.europa.eu/doi/10.2777/1524>

Digitális Objektum definíció, ami adott környezetben megvalósíthatja a FAIR irányelveket:

„Az adatok, szoftver és más erőforrások reprezentációja.” ... „Társítva vannak hozzá perzisztens azonosítók, metaadatok és kontextuális dokumentáció, ami lehetővé teszi a felderíthetőséget, idézést és újrahasznosítást.”

Ennek egy lehetséges implementációja az RO-Crate.

Research Object Crate (RO-Crate) - eredete

Research Object alapelvek:

Azonosíthatóság

Aggregáció

Annotáció

DataCrate:



- Csomagolástechnológia (Bagit + metaadatok JSON-LD formátumban)
- Objektumok RDF leírásai + Schema.org notációk

RO-Crate

RO-Crate leírás

Technikailag:

Adat entitások

- Könyvtárszerkezet (kezdve egy lokális root elemmel)
- Fájlok a könyvtárszerkezetben
- Távoli URI-kkal beazonosítható objektumok

Kontextuális entitások

- Olyan entitások, amik a digitális világon kívül is léteznek (emberek, helyek)
- Elsősorban Metaadat formájában létező leírások (pl. Geokoordináták)

JSON-LD leírás

- Összekapcsolja az adat és kontextuális entitásokat valamilyen publikus séma szerint tipizálva azokat



RO-Crate felépítése

<RO-Crate gyökér könyvtár>/

- | ro-crate-metadata.json # RO-Crate Metadata Fájl – kötelező elem
- | ro-crate-preview.html # RO-Crate Website honlap – megengedett elem
- | ro-crate-preview_files/ # Megengedett elem(ek)
 - | [other RO-Crate honlap fájlok]
- | [payload files and directories] # 0 vagy több

RO-Crate JSON-LD leírás

- Kötelező elem;
- RO-Crate könyvtárstruktúrájának gyökerében található a ro-crate-metadata.json fájlban;
- Leírja a teljes objektumot;
- Leír(hat)ja az egyes adat entitásokat (fájlokat, könyvtárakat és URI-kal beazonosítható entitásokat) úgy, hogy tulajdonságként kontextuális entitásokat rendel hozzájuk;
- Alapvetően a schema.org-ot használja sémaként, de bármi más publikus ontológia használata is támogatott és megengedett;
- A kontextuális entitásoknak minimum a típusa és a címkéje szerepel a leírásban;
- Technológiai szempontból egy flat json-leírás.

RO-Crate JSON-LD példa

```
{ "@context": "https://w3id.org/ro/crate/1.1/context",
  "@graph": [

    {
      "@type": "CreativeWork",
      "@id": "ro-crate-metadata.json",
      "conformsTo": {"@id": "https://w3id.org/ro/crate/1.1"},
      "about": {"@id": "./"}
    },
    {
      "@id": "./",
      "identifier": "https://doi.org/10.4225/59/59672c09f4a4b",
      "@type": "Dataset",
      "datePublished": "2017",
      "name": "Data files associated with the manuscript:Effects of facilitated family case conferencing for ...",
      "description": "Palliative care planning for nursing home residents with advanced dementia ...",
      "license": {"@id": "https://creativecommons.org/licenses/by-nc-sa/3.0/au/"}
    },
    {
      "@id": "https://creativecommons.org/licenses/by-nc-sa/3.0/au/",
      "@type": "CreativeWork",
      "description": "This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Australia License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/au/ or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.",
      "identifier": "https://creativecommons.org/licenses/by-nc-sa/3.0/au/",
      "name": "Attribution-NonCommercial-ShareAlike 3.0 Australia (CC BY-NC-SA 3.0 AU)"
    }
  ]
}
```

RO-Crate importálása az ELKH ARP adatrepozitóriumba

The top part of the screenshot shows the 'Add Data' dropdown menu with options: 'New Dataverse', 'New Dataset', and 'New Ro-Crate'. Below it, the 'Choose a file to upload' interface is shown with a file named 'test-dataset-ro-crate.zip' selected. The 'Upload' button is visible.

The bottom part of the screenshot shows the 'Citation Metadata' form. The 'Title' field contains 'NWS test RO-Crate dataset'. The 'Subtitle' field contains 'NWS dataset with files for testing RO-Crate'. The 'Author' field is filled with 'Admin, Dataverse' and 'Dataverse.org'. The 'Identifier Type' is set to 'Select...'. A note at the top left of the form states: '*Asterisks indicate required fields'.

The screenshot shows the dataset page for 'NWS test RO-Crate dataset'. The page is in 'Draft' and 'Unpublished' status. The description is 'Test dataset from RO-Crate' and the subject is 'Engineering'. The license is 'CC0 1.0'. The page includes a search bar, a 'Change View' section with 'Table' and 'Tree' options, and a list of files. The file list shows two items: '1200px-Pushkin_population_history.svg.png' and '51bf1af2eab8ea4405000000.webp'. Each file has a download icon and a menu icon.

RO-Crate szerkesztése az ELKH ARP adatrepozitóriumban

The screenshot displays the Dataverse AROMA - ARP RO-Crate Manager interface. At the top, there is a navigation bar with the Dataverse logo, search, user guide, support, and user profile (Dataverse Admin). Below this, there are tabs for 'Files', 'Metadata', 'Terms', 'Versions', and 'AROMA'. The main header reads 'ARP AROMA - ARP RO-Crate Manager'. The interface is divided into several sections:

- Dataset:** A tree view on the left shows the dataset structure: 'TTK dataset - 4 class Motor-Imagery EEG' with subfolders 'raw' and 'standardized'. Under 'standardized', there are files named 'S001R01_raw.fif' through 'S001R07_raw.fif'.
- Contextual Entities:** A list of entities including 'author', 'CreativeWork', 'Dataset', 'datasetContact', 'dsDescription', and 'File'.
- Metadata Editor:** A central panel with a search bar and a 'Filter' button. It contains a toolbar with 'Root Dataset', '</> Add', 'Edit Context', 'Preview', and 'Browse Entities'. The main area shows metadata fields: 'Title' (with a value 'TTK dataset - 4 class Motor-I' and a green checkmark), 'Subtitle', 'Alternative Title', 'Alternative URL', and 'Other Id'. Each field has a '+ Text' or '+ URL' or '+ otherid' button to add or edit content.

Copyright © 2023

Powered by  v. 5.13

RO-Crate szerkesztése az AROMA RO-Crate szerkesztővel

The screenshot displays the AROMA - ARP RO-Crate Manager interface. The top navigation bar includes the ARP logo and the text "AROMA - ARP RO-Crate Manager". Below this, a search bar and a "Filter" button are visible. The main content area is divided into two columns. The left column shows a tree view of the dataset structure, including "TTK dataset - 4 class Motor-Imagery EEG" with subfolders for "raw" and "standardized", and a list of files like "S001R01_raw.fif". The right column shows the "About" tab of the dataset editor, with fields for "Title", "Subtitle", "Alternative Title", "Alternative URL", "Other Id", and "Author". Each field has a corresponding "+ [Field Name]" button. The "Title" field is currently filled with "TTK dataset - 4 class Motor-Imagery EEG". Below the "Author" field, there are three rows of author information, each with a unique ID and a red "X" icon.

Dataset

- TTK dataset - 4 class Motor-Imagery EEG
 - raw
 - S001R01_raw.fif
 - S001R02_raw.fif
 - S001R03_raw.fif
 - S001R04_raw.fif
 - S001R05_raw.fif
 - S001R06_raw.fif
 - S001R07_raw.fif
 - S001R08_raw.fif
 - S001R09_raw.fif
 - S001R10_raw.fif
 - S002R01_raw.fif
 - S002R02_raw.fif
 - S002R03_raw.fif
 - S002R04_raw.fif
 - standardized

Contextual Entities

- > author
- > CreativeWork
- > Dataset
- > datasetContact
- > dsDescription
- > File

About

Title
The main title of the Dataset
TTK dataset - 4 class Motor-Imagery EEG

Subtitle
A secondary title that amplifies or states certain limitations on the main title
+ Text

Alternative Title
Either 1) a title commonly used to refer to the Dataset or 2) an abbreviation of the main title
+ Text

Alternative URL
Another URL where one can view or access the data in the Dataset, e.g. a project or personal webpage
+ URL

Other Id
Another unique identifier for the Dataset (e.g. producer's or another repository's identifier)
+ otherid

Author
The entity, e.g. a person or organization, that created the Dataset
+ author

author: #131507ae-89a9-4959-b5f3-fff770a5b601

author: #654e10db-f9a6-478a-95e6-e2f11f017811

author: #69a58547-0343-478b-b78e-fff68acc053

RO-Crate példa felhasználás

1 to 8 of 8 Results

Tüskevár + Almásfüzitő **Draft** **Unpublished**
Apr 9, 2023
HF, 2023, "Tüskevár + Almásfüzitő", <https://doi.org/10.5072/FK2/3YXMBT>, Root, DRAFT VERSION, UNF:6:+CC01Ewg2lm3fwr7pJ85A== [fileUNF]
Tüskevár és Almásfüzitő feltárási helyszín

Tüskevár, fazekaskemence. Felmérési rajz. **Draft** **Unpublished**
Apr 9, 2023
HF, 2023, "Tüskevár, fazekaskemence. Felmérési rajz.", <https://doi.org/10.5072/FK2/Q1CYI5>, Root, UNF:6:+CC01Ewg2lm3fwr7pJ85A== [fileUNF]
Épített fazekaskemence hosszmetraje (222-1). Kiegészítő adatok: méretadat.

HF, 2023, "Tüskevár, fazekaskemence. Felmérési rajz.", <https://doi.org/10.5072/FK2/Q1CYI5>, Root, UNF:6:+CC01Ewg2lm3fwr7pJ85A== [fileUNF]
Épített fazekaskemence hosszmetraje (222-1). Kiegészítő adatok: méretadat.

Almásfüzitő-Vasút melletti hát, település. Ásatási helyszínrajz. **Draft** **Unpublished**
Apr 9, 2023
HF, 2023, "Almásfüzitő-Vasút melletti hát, település. Ásatási helyszínrajz.", <https://doi.org/10.5072/FK2/3YXMBT>, Root, DRAFT VERSION
1, 3, 4, 5. árok felülnézeti és részben metszterajza, foltok dokumentálása. Kiegészítő adatok: mélységadat, felir



Dataset page for "Tüskevár + Almásfüzitő" on DataVerse. The page shows the dataset title, draft status, and a list of files including metadata and images. The description, subject, and license information are also visible.

Tüskevár + Almásfüzitő
Draft Unpublished

HF, 2023, "Tüskevár + Almásfüzitő", <https://doi.org/10.5072/FK2/3YXMBT>, Root, DRAFT VERSION, UNF:6:+CC01Ewg2lm3fwr7pJ85A== [fileUNF]
Cite Dataset - Learn about Data Citation Standards.

Description Tüskevár és Almásfüzitő feltárási helyszín
Subject Arts and Humanities
License/Data Use Agreement Custom Dataset Terms

Files Metadata Terms Versions AROMA

Search this dataset... [Q]

Filter by File Type: All Access: All

1 to 4 of 4 Files [Edit Files] [Download]

File Name	Format	Size	Deposited	MD5	Download
127_RIA_metadata.tab	Tabular Data	64 B	Apr 9, 2023	1 Variables, 1 Observations UNF:6:+CC01Ewg2lm3fwr7pJ85A==	[Download]
44_RIA_Metadata.csv	Comma Separated Values	327 B	Apr 9, 2023	MD5: 5c6...fa0	[Download]
RIA_RK_127.jpg	JPEG Image	211.1 KB	Apr 9, 2023	MD5: 195...5f	[Download]
RIA_RK_44.jpg	JPEG Image	164.9 KB	Apr 9, 2023	MD5: e23...d0b	[Download]

KÖSZÖNÖM A FIGYELMET!

SZTAKI - <https://www.sztaki.hu/>

ELKH ARP honlap - <https://science-research-data.hu/>

ELKH ARP repozitórium - <https://science-data.hu/>

Tóth Zoltán
SZTAKI – DSD
toth.zoltan@sztaki.hu