

Webby FDOs in HUN-REN ARP: Hungary's Multidisciplinary Research Repository

Balázs E. Pataki*, Dr. László Kovács

HUN-REN SZTAKI, DSD
Institute for Computer Science and Control, Hungarian Research Network,
Department of Distributed Systems
Hungary

pataki@sztaki.hu, laszlo.kovacs@sztaki.hu

*Correspondence: Balázs E. Pataki, pataki@sztaki.hu

What is HUN-REN ARP?

The **HUN-REN ARP Adatrepozitórium Platform**¹ (Data Repository Platform) – is the Hungarian Research Network's² (HUN-REN) national initiative to establish a federated infrastructure for the sustainable management of research data across its entire network. [1]

HUN-REN currently includes 7 research centres, 8 independent research institutes, and 101 supported university research groups conducting both basic and applied research throughout Hungary. HUN-REN's disciplinary spectrum is exceptionally broad, covering major fields from the natural sciences (such as physics, chemistry, and materials science) to the life sciences (including biology, genetics, and neuroscience). It is also active in the technical and information sciences, advancing research in mathematics, engineering, artificial intelligence, and quantum technologies.

With this broad disciplinary spectrum come two major challenges:

First, the large volume and diversity of research data that scientists may wish to deposit in the repository; and second, the extensive variety of metadata required to accurately describe these data at both the dataset and the individual file levels.

Moreover, it is a core requirement that all research data be published in the most accessible manner possible — not merely as a philosophical commitment to the FAIR data principles, but through full, operational FAIR implementation. That means the data must not only be **Findable, Accessible, Interoperable, and Reusable**, but also intelligible and actionable by both humans and machines.

In addition to providing a central data repository service, ARP also acts as an aggregator of other Hungarian data repositories, offering a federated metadata-collection service and a unified search interface for the entire research-data production of Hungary.

¹ <https://researchdata.hu/en>

² <https://hun-ren.hu/>

What Do We Need for Being “Genuine” FAIR?

For both humans and machines to understand a dataset and process it further, the data must be richly annotated with metadata. In a heterogeneous research community such as HUN-REN, this means that each discipline – and sometimes even each researcher – may require a different kind of metadata to describe their research data.

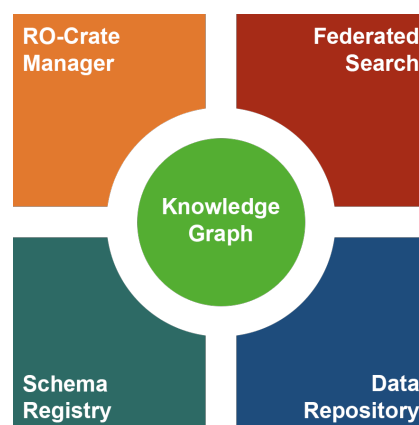
For this reason, we must support the authoring, sharing, and use of metadata schemas throughout the system, so that each component – and human or machine alike – speaks the “same language” and understands what is described in a dataset. Once a dataset is sufficiently annotated, it must be findable via search or a known URL, and the actual data bitstream must then be downloadable.

To achieve all this, we adopt the “**Webby FDO**” [2][3] approach.

The “webby” flavour of FAIR Digital Objects (FDOs) uses standard Web technologies to make objects accessible and processable by both humans and machines:

- persistent HTTP URLs
- metadata embedded or linked using **Signposting**,
- and using a machine-actionable packaging format like **RO-Crate**

Components of ARP Implementing Webby FDOs



Within the ARP system we build upon robust, open-source software solutions and extend them to support true Webby FDOs. [4] At the core is the **Dataverse** [5] platform, serving as our data repository engine: researchers deposit datasets (files, documentation, etc.) and obtain persistent identifiers (PIDs) as part of standard Dataverse capabilities.

On top of that we have built and integrated the **AROMA (ARP RO-Crate Manager)** module so that every data deposit can optionally become a fully packaged RO-Crate object and researchers can apply fine-grained, file-level metadata (not just dataset-level) using our visual RO-Crate editor.

Meanwhile the **CEDAR-based** [6] Metadata Schema Registry allows users (data stewards and/or researchers) to author, version, share and publish metadata schemas (templates, “metadata blocks”, profiles) that can be used both in Dataverse and in the metadata embedded into RO-Crate packages.

The result is that all metadata across packages and files is consistent, extensible across disciplines, and available to both machines and humans.

Finally, ARP includes a **Knowledge Graph-based integrated backbone** and a **Federated Search service** — these enable cross-repository indexing and discovery of both dataset-level and file-level metadata throughout Hungary’s research data repository network.

Webby FDOs in Action

A truly machine-actionable infrastructure like ARP opens many possibilities for more exact and convenient research-data management. While users still interact via intuitive visual interfaces, persistent identifiers, rich metadata profiles, Signposting-based metadata discovery, and fully machine-actionable RO-Crate packaging enable applications to take on tedious or repetitive tasks on behalf of the user.

One such ongoing experiment within the platform is “VibeARP”: leveraging these machine-actionable features, VibeARP allows users to build their dataset metadata via an integrated AI agent, which can access web search results and browse the web with an actual browser to identify relevant “ingredients” for a dataset and help describe it in sufficient detail.

- [1] L. Kovács, “Az ELKH Adatrepozitórium Platform projekt – A kutatási adattárolás és megosztás országos federált rendszere,” *Magyar Tudomány*, vol. 184, no. 7, pp. 839–850, Dec. 2023, doi: <https://doi.org/10.1556/2065.184.2023.7.3>
- [2] S. Soiland-Reyes, “FAIR Research Objects and Computational Workflows – A Linked Data Approach,” Ph.D. dissertation, University of Amsterdam, Amsterdam, Netherlands, Jan. 2025, ISBN 978-94-93391-78-9, doi: <https://doi.org/10.5281/zenodo.8113625>
- [3] S. Soiland-Reyes, P. Sefton, S. Leo, L. J. Castro, C. Weiland, and H. Van de Sompel, “Practical webby FDOs With RO-Crate and FAIR Signposting: Experiences and Lessons Learned,” in *Proc. International FAIR Digital Objects Implementation Summit 2024*, ser. Open Conference Proceedings, vol. 5. Hannover, Germany: TIB Open Publishing, Mar. 2025, doi: <https://doi.org/10.52825/ocp.v5i.1273>
- [4] B. Pataki “Dataverse, CEDAR and RO-Crate: The building blocks of ARP, the federated research data repository of Hungary”, *Research Object Crate (RO-Crate)*, Oct. 2025, <https://www.researchobject.org/ro-crate/blog/2025-10-13/dataverse-cedar-and-ro-crate-the-building-blocks-of-arp>
- [5] G. King, “An Introduction to the Dataverse Network as an Infrastructure for Data Sharing,” *Sociological Methods & Research*, vol. 36, no. 2, pp. 173–199, Nov. 2007, doi: <https://doi.org/10.1177/0049124107306660>
- [6] R. S. Gonçalves, M. J. O’Connor, M. Martínez-Romero, A. L. Egyedi, D. Willrett, J. Graybeal, and M. A. Musen, “The CEDAR Workbench: An Ontology-Assisted Environment for Authoring Metadata that Describe Scientific Experiments,” in *Proc. Int. Semantic Web Conf. (ISWC 2017)*, Vienna, Austria: Springer, 2017, pp. 103–110, doi: https://doi.org/10.1007/978-3-319-68204-4_10