3rd FAIR Digital Objects Conference
TIBconferencesession
TIB-OP_will_set_DOI

Published: 1970-01-01

Implementation of Fair Digital Objects in the CLARIN Infrastructure

Willem Elbers[1]*, Twan Goosen[1], and Dieter van Uytvanck[1]

[1]CLARIN ERIC, Utrecht, The Netherlands
*Correspondence: Willem Elbers, willem@clarin.eu

Abstract. There is a clear need to improve interoperability in the wider research landscape and within CLARIN specifically. Fair Digital Objects (FDOs) can strengthen interoperability within CLARIN in various ways. Three concrete examples are adding FDO support to the Digital Object Gateway (DOG) service and the Virtual Collection Registry (VCR) and explain how FDOs improve the citation of datasets.

Keywords: FAIR Signposting, RO-Crate, Digital Object Gateway, Virtual Collection Registry, Dataset Citation

1. Background

The Strategic Innovation and Research Agenda (SRIA) of the European Open Science Cloud (EOSC) [1] highlights the establishment of the Web of FAIR data and a Minimum Viable EOSC (MVE) by 2027 as priority. Together with the introduction of an EOSC Federation of Nodes in 2023 this clearly emphasizes the need to improve interoperability.

Within CLARIN[1], a digital infrastructure offering data, tools and services to support research based on language resources, there has been a longstanding acceptance of, discussion on [2] and involvement with Fair Digital Objects (FDOs). One of the community specific approaches proposed is to embed signposting information into the CLARIN specific Component Metadata [3]. The Australian LDaCa [4] sister infrastructure is already making use of RO-crates. In this paper we will explore more options to position FDOs within CLARIN's infrastructure by means of widely supported standards to improve interoperability. Concretely we will discuss how FDOs can contribute to make the Digital Object Gateway (DOG) and Virtual Collection Registry (VCR) more interoperable, and how they may play a role in addressing common dataset citation challenges.

2. Digital Object Gateway

The Digital Object Gateway (DOG) is an abstraction layer for accessing resources and services referenced by a Persistent Identifier (PID) or URL. It can be used both by end-users directly, and by other services within the infrastructure. A user of the DOG does

[1] https://www.clarin.eu

not need to know the specifics of how to access one particular data repository. Instead, by invoking it with the identifier of an object (which typically resolves to the landing page by default), the DOG can perform a range of standard operations to explore and process the object. This is a natural fit for the concept of FDOs where the DOG acts as a consumer. Interoperability across the infrastructure could be improved by implementing functionality to recognize FAIR signposting [5] and RO-crate [6] objects, and to retrieve, process and relay the relevant metadata.

3. Virtual Collection Registry

A virtual collection is a coherent set of links to digital objects that can be easily created, accessed, (both manually and automated), and cited. The links can originate from different archives. CLARIN provides a registry (VCR) where scholars can manage and publish their virtual collections. The collection metadata is currently only offered as HTML, CMDI and DOI metadata which works in our domain but is fairly CLARIN specific. Serving the collection metadata via FAIR signposting in addition to the existing approaches will improve interoperability of the virtual collections, especially outside of the CLARIN domain.

4. Data citation

Citation of datasets has always been a challenge [CITATION?]. The most common practice within CLARIN is to refer, by PID, to a dataset's landing page at the repository where it is stored. However this is not always followed and not implemented consistently across repositories. Both RO-Crate and the FAIR signposting specifications have adopted the "cite-as" relation based on RFC8574 [7] which would give a clearly defined URI (ideally a DOI) to cite the dataset in a standardised, interoperable way.

5. Conclusion

The concrete steps to implement FDO support, both serving and consuming, contribute to the FDO landscape within the CLARIN infrastructure and more broadly. Via the DOG, FDO support for FAIR signposting and RO-Crates is exposed within the infrastructure to all connected services. By offering our virtual collections via FAIR signposting, the collections are also made accessible outside our domain. Additionally, the mentioned specifications offer a common approach to unambiguous datasets citation.

Data availability statement

No data were created or analyzed in this article.

Author contributions

Twan Goosen (Writing – original draft), Willem Elbers (Writing – original draft), Dieter van Uytvanck (Writing – review & editing)

Competing interests

The authors declare that they have no competing interests.

References

[1] E. Association, *Strategic research and innovation agenda (sria) of the european open science cloud (eosc)*, 2024. [Online]. Available: https://eosc.eu/wp-content/uploads/2024/12/20241031_SRIA_1.3_final_Annex.pdf.

[2] D. Broeder, W. Elbers, M. Gawor, C. Concordia, N. Larrousse, and D. Van Uytvanck, *Towards fair data access.* 2022. DOI: 10.3897/rio.8.e94386. [Online]. Available: https://doi.org/10.3897/rio.8.e94386.

[3] T. T. Denis Arnold Bernhard Fisseni, *Signposts for clarin.* 2021. DOI: 10.3384/ecp1803. [Online]. Available: https://doi.org/10.3384/ecp1803.

[4] R. M. Michael Haugh Simon Musgrave, "The language data commons of australia: Towards a nationally distributed research infrastructure," in *CLARIN Annual Conference Proceedings*, Oct. 2025, pp. 21–25. [Online]. Available: https://cris.brighton.ac.uk/ws/portalfiles/portal/60194662/CLARIN2025_ConferenceProceedings.pdf.

[5] M. Wilkinson et al., *Report on fair signposting and its uptake by the community*, Jan. 2024. DOI: 10.5281/zenodo.10490289. [Online]. Available: https://doi.org/10.5281/zenodo.10490289.

[6] S. Soiland-Reyes et al., "Packaging research artefacts with ro-crate," *Data Science*, vol. 5, no. 2, pp. 97–138, 2022. DOI: 10.3233/DS-210053. [Online]. Available: https://doi.org/10.3233/DS-210053.

[7] H. V. de Sompel, *Cite-as: A link relation to convey a preferred uri for referencing*, RFC 8574, Apr. 2019. DOI: 10.17487/RFC8574. [Online]. Available: https://www.rfc-editor.org/rfc/rfc8574.html.