International FAIR Digital Objects Implementation Summit 2024
Extended Abstracts
https://doi.org/10.52825/ocp.v5i.1190

Published: 18 Mar. 2025

FDO as an Interoperability Framework for the Biodiversity Digital Twin Project

Julian Lopez Gordillo[1,*] and Sharif Islam[1]

[1]Naturalis Biodiversity Center, Leiden, Netherlands

*Correspondence: Julian Lopez Gordillo, julian.lopezgordillo@naturalis.nl

Abstract. This conference abstract introduces our progress with using the FAIR Digital Object (FDO) paradigm, implemented through RO-Crate, as an interoperability framework in the Biodiversity Digital Twin (BioDT) project. After providing some background on the project and the importance of FAIR in it, the approach followed concerning FDO and RO-Crate is explained, together with our current efforts in structuring the different digital object types in BioDT, as well as the development of metadata profiles and our attempts at aligning with community practices.

Keywords: Digital Objects, FDO, FAIR, RO-Crate, Metadata, Digital Twins, Biodiversity, Machine-Actionability

1. Background

Originally, the FAIR principles (Findable, Accessible, Interoperable, Reusable) were proposed by Wilkinson et al. [1] to tackle issues concerning data, but they have already been adapted to other digital objects such as software, workflows, or machine learning models. Nonetheless, the core idea remains the same: to ensure that such elements can be consistently found, accessed (when appropriate) and reused by the scientific community in a variety of ways. This includes topics such as persistent identification, compliance with open standards or machine-actionability, to name a few. In that sense, it is useful to add a level of abstraction and treat all the aforementioned elements as *digital objects* in order to establish (meta)data formats and protocols that apply to all of them. That is the core idea behind the FAIR Digital Objects (FDOs) paradigm from De Smedt et al. [2], a concept we are using as an interoperability framework in the Biodiversity Digital Twin (BioDT) project funded by the European Union [3].

BioDT aims to push the current boundaries of predictive understanding of biodiversity dynamics by developing digital twins providing advanced modelling, simulation and prediction capabilities. A digital twin (as defined by the Digital Twin Consortium) is a virtual representation of real-world entities and processes, synchronised with real-world data at a specified frequency and fidelity. In turn, digital twinning refers to the process of designing and developing digital twins, and integrating them into their wider operational environment. To that end, the project deals with heterogeneous use cases, each involving different data, models, and workflows. These use cases range from well-established and domain-specific, such as BEEHAVE [4], to citizen-science monitoring mobile apps that are in active development and rapidly evolving [5]. Consequently, the models that are at the core of the use cases are diverse in their nature and objectives: apart from species distribution models, some of them take a different modelling approach (i.e. agent-based, machine learning, or dynamic data driven application systems). In that sense, rather than a single biodiversity digital twin, the project is developing several scoped digital twins that are fundamentally independent among each other. Within the project, they are referred to as "prototype digital twins" (pDTs).

This heterogeneity is also reflected in the volume and type of data sources the pDTs need, as well as the underlying software and workflows they rely on. From the outset of the project, there was not necessarily much overlap on their data requirements or technical implementation. Nonetheless, they all benefit from the developments happening within the BioDT project, which aims to establish common foundations for this wide array of use cases. The challenge lies in organising and integrating these components to support the digital twinning effort. For example, a particular digital twin is based on a certain set of data to compute its prediction (for which it depends on certain software and workflows). Inherent to the concept of a digital twin is its ability to remain synchronised, which might involve updating certain components (i.e. updating some of its data sources or software packages). Those changes and dependencies must be captured to provide reliable provenance about how the digital twin's outputs were produced. In this context, the FAIR principles and implementation thus play an important role for future sustainability. Recognising the importance of adhering to FAIR data principles, the BioDT project has dedicated a work package to improve data, workflows, and models through FAIR Principles.

2. The FDO layer and RO-Crate

The FAIR principles are mostly aspirational and offer little guidance on the specific implementation choices one must make to achieve them. Each of the pDTs started at different points in their FAIR implementation journey, possibly having taken different paths to address the same FAIR principle. To accommodate this diversity, we need a lower-level framework that allows us to structure the different elements of the pDTs at the technical level such that we can work towards more coherent and coordinated implementations for improving their machine-actionability and other aspects regarding FAIR.

FAIR Digital Objects as a framework can provide a foundation to organise multidisciplinary heterogeneous data by improving machine-actionability, interoperability and reusability. Through the FAIR data principles and the FDO framework, it is possible to achieve contextual and machine-actionable metadata such that enough information for agents (both human and machine) is provided, while also supporting operational requests on each digital object [6], [7]. In other words, we can build a foundation for the Research Infrastructures (RIs) to work with a unified data layer that accommodates computational workflows, interlinking and reusability for the BioDT Digital Twin technical platform. Examples of an infrastructural approach exist in the context of biomedical application [8] and academic publishing [9].

This vision is further developed through the FDO layer, a conceptual view of how to integrate FDO records within BioDT and the RIs taking part in the project (see Figure 1). The FDO layer serves two main purposes:

1. Provide a reference for the design of digital objects within BioDT such that they are coherent and can interact with each other as part of the larger system.
2. Harmonise the description of digital objects to external resources provided by different RIs and the other data and service providers.

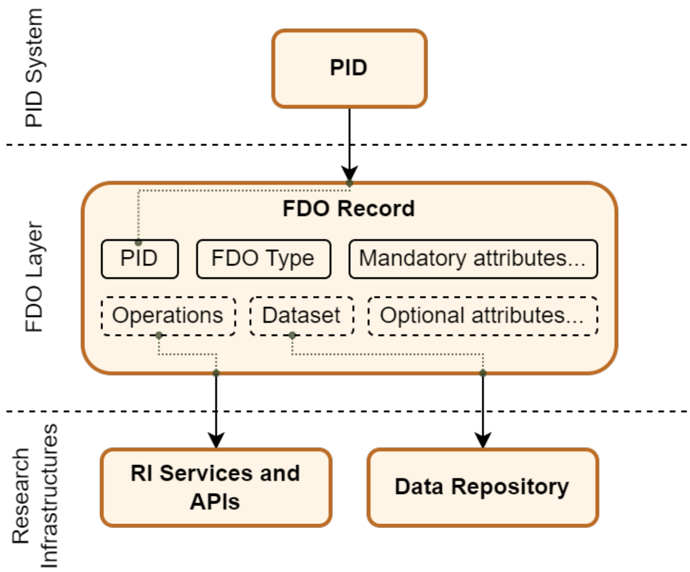

Figure 1. *The FDO layer as an interoperability framework for BioDT. Through machine-actionable metadata in the form of FDO records, it would serve as a middle layer between the RIs and the Persistent Identifier (PID) system.*

Because the FDO layer was envisioned to support different possible implementations, it is inherently conceptual. To put these principles into practice, we need an additional, adaptable implementation that could easily be integrated into the existing use cases. RO-Crate [10], based on Schema.org annotations in JSON-LD, is a mature specification for packaging research data with their metadata. Thanks to its flexibility, it can be used in conjunction with other tools —most notably, FAIR Signposting [11]— to achieve a lightweight yet powerful approach to FDOs that is based on well-established technologies. While we are using FDOs to give a coherent structure to our the technical developments concerning FAIR, the distributed nature of BioDT calls for a bottom-up approach. That is where RO-Crate excels; its simplicity and ease of use made the development of metadata descriptions more approachable for the pDTs.

RO-Crate might not be the most straightforward implementation of the FDO framework (i.e. when compared against the digital object concept with DOIP V2.0 [12]). Following the RO-Crate guidelines, despite helping in achieving FAIR, doesn't guarantee alignment with the FDO framework per se. However, much work has been done in bridging the FDO specifications with the Linked Data world (to which RO-Crate belongs), as shown by Soiland-Reyes et al. [13], and showcasing how RO-Crate can indeed be a valid approach to FDOs, as already mentioned [11] [10]. Similarly, the adequacy of RO-Crate for truly complex distributed resources has been challenged by some. A thorough examination of FDO and Linked Data is given by Soiland-Reyes et al. [14], including the shortcomings and areas of improvement of both sides for the purpose of achieving FAIR. From that research, we would like to highlight some of the concluding thoughts:

> *”We find that both FDO and Linked Data approaches can benefit significantly from each-other and should be aligned further.*
> *[...]*
> *By implementing the goals of FAIR Digital Objects with the mature technology stack developed for Linked Data, EOSC research infrastructures and researchers in general can create and use FAIR machine-actionable research outputs for decades to come.”*

Indeed, we think that using RO-Crate (and its rich ecosystem) as a vehicle, we will be able to build up from the common but necessary groundwork of developing metadata descriptions and tackling some of the basic aspects of achieving FAIR into developing a community practice that increasingly conforms to the requirements laid out by the FDO specifications by Anders et al. [15]. Naturally, we acknowledge the critical importance for FAIR and FDO of some aspects that we are not fully addressing yet, most notably Persistent Identifiers (PIDs) and PID records. However, this is a conscious decision which is influenced by several factors from the context of the project, like its distributed nature. From the researchers' perspective, dealing with RO-Crate metadata files and URLs is more approachable than working with PID records from the Handle system. Focusing on the bottom-up collaboration with the use cases results in a better integration of the concepts regarding FAIR and FDO that we envision for the project, even if some aspects are not addressed from the start as a result. While implementing the FDO record with a Persistent Identifier (PID) system like Handle is one way to achieve it, it is important to remember that it is an implementation choice. The core concept of the FDO specification and its approach of abstracting metadata remain consistent regardless of the chosen PID system.

3. Progress on implementation

To that end, some of the earlier work within the project consisted in conceptualising the frameworks and approaches discussed here and informing our project collaborators about such ideas, their significance, and our overall vision. Once our goals for the FDO layer were established, the next steps fell in place accordingly. The first prominent task was to start developing metadata profiles that could be used to gather the digital object types and metadata attributes that were needed to describe digital objects within the use cases in BioDT.

However, these goals were set out just before the definition of the prototype digital twins within BioDT took place, which changed the internal distribution of the project from traditional work packages into pDTs, centered around the use cases but involving collaborators from all areas (technical infrastructure, FAIR implementation, etc...). This change in the working paradigm of the project led to the development of metadata descriptions for the pDTs' digital objects co-occurring with the development of the metadata profiles they would eventually conform to, namely the FDO profiles. As discussed earlier, a positive outcome of this change was a more back-and-forth discussion on topics like alignment with community standards.

3.1 FDO Profiles and Kernel Attributes

These FDO profiles, as well as some other parts of the FDO work in BioDT, are in active development, growing dynamically according to the needs of the project. So far, progress has been made in the definition of metadata attributes that should be present in all BioDT digital objects, regardless of their type, as well as some type-specific attributes. The former align with the Kernel Attributes as defined by the FDO specifications by Broeder et al. [16], while the latter would fall into the Community Attributes. Yet, following our current thinking about what information would be necessary for the different digital object types to seamlessly work together, we foresee that we need stronger restrictions on the cardinality of the metadata attributes than those stated on [16].

As stated previously, we are not dealing with PID records yet (although it remains a key objective of the FDO layer) but with metadata descriptions that exist as RO-Crate metadata files and could be referenced in different ways. Nevertheless, we have started adopting some recent developments in the RO-Crate specification (i.e. RO-Crate draft version 1.2, [17]) that will facilitate the definition of ”Profile Crates” to collect different resources linked to a certain profile. Those include the specification, schema, validation, examples, and others, as defined by *The Profiles Vocabulary* [18] from W3C. Such resources will help greatly in generalising the metadata efforts happening around RO-Crate into PID records and the wider FDO framework, a task that will likely be tackled in the later part of the project, after a wider consensus has been reached about certain project-wide choices, such as common solutions for repositories and data storage.

3.2 FDO Types and Community Attributes

The fundamental digital object types within BioDT were initially considered to be Dataset, Model and Workflow, but others might be added as required (i.e. Mapping Set, relevant for a one of our work package’s tasks). The selection of their type-specific metadata attributes has been centred around the needs of the pDTs which are more advanced in their usage or have them as a distinctive feature. We have aimed at aligning with existing community standards and protocols in their respective domains, although we are exploring several avenues simultaneously even though we might not be able to fully commit to all of them for the duration of the project. Current efforts for each digital object type include the following:

- **Model:** developing a machine-actionable version of the ODMAP (Overview, Data, Model, Assessment and Prediction) protocol for reporting species distribution models [19]; aligning with HugginFace’s Model cards for sharing machine learning models [20]; adopting FAIR-IMPACT’s Research Software Metadata Guidelines [21].
- **Dataset:** reusing granular terms from Darwin Core [22] and DCAT-3 [23]; interoperability with the research infrastructures of BioDT (such as the GBIF Data Model [24], the DiSSCo Open Digital Specimen [25], or the LTER Digital Asset Register); alignment with Biodiversity Data Cubes [26] (possibly via the Reliance RO-Crate profile [27]) and other data formats.
- **Workflow:** conformance with Bioschemas’ ComputationalWorkflow profile [28]; use of WorkflowHub from Goble et al. [29] as the standard working repository for workflows; capturing provenance from HPC workflows (possibly through the more granular Workflow Run RO-Crate [30]); integration with the OPeNDAP protocol [31].
- **Mapping Set:** adoption of the Simple Standard for Sharing Ontology Mappings (SSSOM) [32]; registration in EOSC’s Metadata Schema and Crowsswalk Registry (MSCR) [33]; integration with mapping.bio (based on Cordra [34]).

4. Conclusion and future directions

All in all, there is still much to be done in terms of increasing the breadth and depth of our current FDO implementation within BioDT. Current FDO profiles need to be developed further, new ones might need to be established, and metadata descriptions conforming to them need to be created for all of the pDTs in a more systematic manner. Our current bottom-up approach, driven by RO-Crate implementation, needs to mature enough to meet the FDO layer ideas defined earlier in the project. This should become easier as BioDT's technical platform takes shape and the pDTs start adopting more definitive solutions.

As a notable advancement on this front, we have started to shape out how to implement the FDO layer as a service that can interact with BioDT's technical architecture. This could involve defining certain operations on digital objects coming from the research infrastructures such that they can be served following the specifications established within the project. We envision transforming the FDO layer into a service or framework that seamlessly integrates with BioDT's technical architecture, aligning with the different data and services expertise provided by the different research infrastructures.

This strategic move aims to enhance accessibility, interaction, and overall efficiency within, solidifying BioDT's position at the forefront of innovative biodiversity research and digital twinning endeavours.

Author contributions

Authors contributions to this article according to the Contributor Roles Taxonomy (CreDIT) [35]:

- **Julian Lopez Gordillo:** Conceptualization, Software, Writing – original draft, Writing – review and editing
- **Sharif Islam:** Conceptualization, Supervision, Writing – review and editing
- **Wouter Addink:** Writing – review and editing
- **Soulaine Theocharides:** Writing – review and editing

Data availability statement

This article discusses openly available resources, published on the web, and does not use any data for the conclusions stated here. When appropriate, URLs and DOIs are included in the Reference section.

Underlying and related material

No supplementary materials were used for this article. All the resources referenced through the article are available online. Materials developed within the scope of the project and for aspects related to FAIR and FDO are being made public in the following repository on GitHub: https://github.com/BioDT/biodt-fair.

Competing interests

The authors declare that they have no competing interests.

Funding

This study has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement No 101057437 (BioDT project, https://doi.org/10.3030/101057437). Views and opinions expressed are those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

Funded by the European Union

Acknowledgements

We would like to thank our project partners from BioDT for their contributions to the topics discussed here.

References

[1] M. D. Wilkinson et al., "The FAIR guiding principles for scientific data management and stewardship," *Scientific Data*, vol. 3, no. 1, p. 160 018, Mar. 15, 2016, ISSN: 2052-4463. DOI: 10.1038/sdata.2016.18. Accessed: Jan. 4, 2023. [Online]. Available: https://www.nature.com/articles/sdata201618.

[2] K. De Smedt, D. Koureas, and P. Wittenburg, "FAIR digital objects for science: From data pieces to actionable knowledge units," *Publications*, vol. 8, no. 2, p. 21, Apr. 11, 2020, ISSN: 2304-6775. DOI: 10.3390/publications8020021. Accessed: Jan. 19, 2023. [Online]. Available: https://www.mdpi.com/2304-6775/8/2/21.

[3] "Biodiversity digital twin for advanced modelling, simulation and prediction capabilities," Accessed: Mar. 4, 2024. [Online]. Available: https://cordis.europa.eu/project/id/101057437.

[4] M. A. Becher, V. Grimm, P. Thorbek, J. Horn, P. J. Kennedy, and J. L. Osborne, "Beehave: A systems model of honeybee colony dynamics and foraging to explore multifactorial causes of colony failure," *Journal of Applied Ecology*, vol. 51, no. 2, E. Morgan, Ed., pp. 470–482, Apr. 2014, ISSN: 0021-8901, 1365-2664. DOI: 10.1111/1365-2664.12222. Accessed: Dec. 14, 2023. [Online]. Available: https://besjournals.onlinelibrary.wiley.com/doi/10.1111/1365-2664.12222.

[5] P. Lauha et al., "Domain-specific neural networks improve automated bird sound recognition already with small amount of local data," *Methods in Ecology and Evolution*, vol. 13, no. 12, pp. 2799–2810, 2022, ISSN: 2041-210X. DOI: 10.1111/2041-210X.14003. Accessed: Oct. 20, 2023. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/2041-210X.14003.

[6] E. Schultes et al., "FAIR digital twins for data-intensive research," *Frontiers in Big Data*, vol. 5, p. 883 341, May 11, 2022, ISSN: 2624-909X. DOI: 10.3389/fdata.2022.883341. Accessed: Mar. 1, 2023. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fdata.2022.883341/full.

[7] U. Schwardmann and T. Kálmán, "Two examples on how FDO types can support machine and human readability," *Research Ideas and Outcomes*, vol. 8, e96014, Oct. 12, 2022, ISSN: 2367-7163. DOI: 10.3897/rio.8.e96014. Accessed: Dec. 29, 2023. [Online]. Available: https://riojournal.com/article/96014/.

[8] M. Conte, A. Flynn, P. Barrison, P. Boisvert, Z. Landis-Lewis, and C. Friedman, “Digital objects to make computable biomedical knowledge FAIR: An infrastructural approach to knowledge representation, dissemination and implementation,” *Research Ideas and Outcomes*, vol. 9, e109307, Dec. 18, 2023, ISSN: 2367-7163. DOI: 10.3897/rio.9.e109307. Accessed: Dec. 29, 2023. [Online]. Available: https://riojournal.com/article/109307/.

[9] E. Schultes, “FAIR digital objects for academic publishers,” *Information Services & Use*, vol. 44, no. 1, M. Duine, Ed., pp. 15–21, Dec. 22, 2023, ISSN: 01675265, 18758789. DOI: 10.3233/ISU-230227. Accessed: Dec. 29, 2023. [Online]. Available: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/ISU-230227.

[10] S. Soiland-Reyes et al., “Packaging research artefacts with RO-crate,” *Data Science*, vol. 5, no. 2, S. Peroni, Ed., pp. 97–138, Jul. 20, 2022, ISSN: 24518492, 24518484. DOI: 10.3233/DS-210053. Accessed: Feb. 1, 2023. [Online]. Available: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/DS-210053.

[11] S. Soiland-Reyes et al., “Creating lightweight FAIR digital objects with RO-crate and FAIR signposting,” Oct. 26, 2022. DOI: 10.5281/ZENODO.7231714. Accessed: Feb. 1, 2023. [Online]. Available: https://zenodo.org/record/7231714.

[12] “Digital object interface protocol specification, version 2.0,” 2018. Accessed: Nov. 12, 2018. [Online]. Available: https://hdl.handle.net/0.DOIP/DOIPV2.0.

[13] S. Soiland-Reyes, L. J. Castro, D. Garijo, M. Portier, C. Goble, and P. Groth, “Updating linked data practices for FAIR digital object principles,” Oct. 27, 2022. DOI: 10.5281/ZENODO.7256427. Accessed: Feb. 1, 2023. [Online]. Available: https://zenodo.org/record/7256427.

[14] S. Soiland-Reyes, C. Goble, and P. Groth, “Evaluating FAIR digital object and linked data as distributed object systems,” *PeerJ Computer Science*, vol. 10, e1781, Apr. 2024, ISSN: 2376-5992. DOI: 10.7717/peerj-cs.1781. Accessed: Nov. 25, 2024.

[15] I. Anders et al., “FDO forum FDO requirement specifications,” 2023. DOI: 10.5281/zenodo.7782262.

[16] D. Broeder, P. Wittenburg, I. Anders, and K. Peters-von Gehlen, “FDO – kernel attributes & metadata,” May 10, 2023. DOI: 10.5281/ZENODO.7825692. Accessed: Mar. 4, 2024. [Online]. Available: https://zenodo.org/record/7825692.

[17] “RO-crate 1.2-DRAFT,” Accessed: Mar. 4, 2024. [Online]. Available: https://www.researchobject.org/ro-crate/1.2-DRAFT/.

[18] “The profiles vocabulary,” Accessed: Mar. 4, 2024. [Online]. Available: https://www.w3.org/TR/dx-prof/#introduction.

[19] D. Zurell et al., “A standard protocol for reporting species distribution models,” *Ecography*, vol. 43, no. 9, pp. 1261–1277, 2020, ISSN: 1600-0587. DOI: 10.1111/ecog.04960. Accessed: Oct. 25, 2023. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/ecog.04960.

[20] M. Mitchell et al., “Model cards for model reporting,” in *Proceedings of the Conference on Fairness, Accountability, and Transparency*, Jan. 29, 2019, pp. 220–229. DOI: 10.1145/3287560.3287596. arXiv: 1810.03993[cs]. Accessed: Dec. 14, 2023. [Online]. Available: http://arxiv.org/abs/1810.03993.

[21] M. Gruenpeter et al., *D4.4 - guidelines for recommended metadata standard for research software within EOSC*, Jul. 2023. DOI: 10.5281/zenodo.8199104. [Online]. Available: https://doi.org/10.5281/zenodo.8199104.

[22] J. Wieczorek et al., "Darwin core: An evolving community-developed biodiversity data standard," *PLoS ONE*, vol. 7, no. 1, I. N. Sarkar, Ed., e29715, Jan. 6, 2012, ISSN: 1932-6203. DOI: 10.1371/journal.pone.0029715. Accessed: Mar. 4, 2024. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0029715.

[23] "Data catalog vocabulary (DCAT) - version 3," Accessed: Mar. 4, 2024. [Online]. Available: https://www.w3.org/TR/vocab-dcat-3/#acknowledgments.

[24] "GBIF new data model," Accessed: Mar. 4, 2024. [Online]. Available: https://www.gbif.org/new-data-model.

[25] Distributed System of Scientific Collections, *open Digital Specimen specification*. [Online]. Available: https://github.com/DiSSCo/openDS.

[26] "Data and evidence," Accessed: Mar. 4, 2024. [Online]. Available: https://b-cubed.eu/data-and-evidence.

[27] "RELIANCE RO-crate profile," Accessed: Mar. 4, 2024. [Online]. Available: https://reliance-eosc.github.io/reliance-ro-crate/.

[28] "Bioschemas - 1.0 release," Accessed: Mar. 4, 2024. [Online]. Available: https://bioschemas.org/profiles/ComputationalWorkflow/1.0-RELEASE#nav-description.

[29] C. Goble et al., *Implementing FAIR digital objects in the EOSC-life workflow collaboratory*, Mar. 2021. DOI: 10.5281/zenodo.4605654. [Online]. Available: https://doi.org/10.5281/zenodo.4605654.

[30] S. Leo et al., "Recording provenance of workflow runs with RO-crate," 2023. DOI: 10.48550/ARXIV.2312.07852. Accessed: Mar. 4, 2024. [Online]. Available: https://arxiv.org/abs/2312.07852.

[31] "Opendap," Accessed: Mar. 4, 2024. [Online]. Available: https://www.opendap.org/.

[32] N. Matentzoglu et al., "A simple standard for sharing ontological mappings (SSSOM)," *Database*, vol. 2022, baac035, Jan. 1, 2022, ISSN: 1758-0463. DOI: 10.1093/database/baac035. Accessed: Mar. 4, 2024. [Online]. Available: https://doi.org/10.1093/database/baac035.

[33] "Metadata schema and crosswalk registry (MSCR)," Accessed: Mar. 4, 2024. [Online]. Available: https://faircore4eosc.eu/eosc-core-components/metadata-schema-and-crosswalk-registry-mscr.

[34] "What's cordra?" Accessed: Mar. 4, 2024. [Online]. Available: https://www.cordra.org/cordra.html.

[35] "CRediT – contributor roles taxonomy," Accessed: Mar. 4, 2024. [Online]. Available: https://credit.niso.org/.