International FAIR Digital Objects Implementation Summit 2024

Extended Abstracts

https://doi.org/10.52825/ocp.v5i.1202

Published: 18 Mar. 2025

Enhancing FAIR Data Practices in the Norwegian Research Data Archive

Towards Research Objects and Improved Interoperability

Anne Fouilloux[1,*] and Adil Hasan[2]

[1] Simula Research Laboratory, Norway

[2] Sigma2, Norway

*Correspondence: Anne Fouilloux, annef@simula.no

Abstract. The increasing volume and complexity of research data necessitate robust data management practices to ensure data is Findable, Accessible, Interoperable, and Reusable (FAIR). The Norwegian Research Data Archive (NRDA) is at the forefront of efforts to create a comprehensive platform for researchers to share and archive their data. This paper discusses NRDA's ongoing initiatives to enhance its infrastructure in alignment with FAIR principles, emphasizing the integration of Research Objects (ROs) and RO-Crate technologies. These improvements aim to facilitate better data discoverability, accessibility, and interoperability, thereby fostering a more integrated and sustainable data ecosystem. The paper also highlights NRDA's collaborative efforts with other platforms via the use of Research Objects to support data sharing and reuse across repositories. By focusing on standardized metadata, persistent identifiers, and interoperability, NRDA is advancing Open Science practices, ultimately contributing to a more transparent, efficient, and collaborative research environment. The challenges and future directions of these initiatives are also explored, providing insights into the ongoing efforts to create a more open and interconnected scientific landscape.

Keywords: FAIR, Research Object, Data Archive

1. Introduction

The rapid expansion of research data across diverse scientific disciplines has reinforced the need for effective data management, sharing, and accessibility. As part of the global movement towards Open Science, the integration of FAIR (Findable, Accessible, Interoperable, and Reusable) principles [1] into research data infrastructures has become a central focus for repositories worldwide. In Norway, the Norwegian Research Data Archive (NRDA) plays a key role in supporting researchers by providing a research data platform for storing, sharing and archiving (long-term storage) large volumes of data across various research fields.

This paper outlines NRDA's ongoing efforts to enhance its infrastructure in alignment with FAIR principles, with a particular focus on the integration of Research Objects (ROs) and RO-Crate technologies. By improving the discoverability, accessibility, and interoperability of its datasets, NRDA aims to foster a more integrated and sustainable data ecosystem that benefits researchers, institutions, and the broader scientific community. The paper also explores the importance of collaboration with repositories, particularly with DataVerseNO, and how the

seamless integration of Research Objects between these platforms could better support data sharing and enhance the re-usability of research data across repositories.

As NRDA moves forward with its transformation, the successful adoption of these technologies will pave the way for greater data reuse, collaboration, and knowledge sharing, ultimately contributing to the advancement of Open Science in Norway and beyond.

1.1 The Norwegian Research Data Archive (NRDA) Overview

The Norwegian Research Data Archive (NRDA) is a vital service dedicated to supporting Norwegian researchers in managing, storing, and sharing their research data. As a free and widely accessible platform, NRDA currently hosts approximately 1.6 PiB of data across a variety of scientific domains, making it a cornerstone resource for the Norwegian research community.

NRDA provides essential infrastructure for long-term data storage, ensuring that research data is securely archived while remaining readily accessible for future use. Beyond its primary role in long-term data storage, NRDA facilitates collaboration and data sharing among researchers, institutions, and international data repositories.

Committed to adhering to international standards of Open Science and data sharing, NRDA equips users with tools and services to enhance data discoverability, accessibility, and reusability. Supporting a diverse range of research outputs—from large-scale datasets to smaller project-specific datasets—NRDA actively contributes to implementing FAIR (Findable, Accessible, Interoperable, and Reusable) data principles.

Through continuous upgrades and innovations, NRDA seeks to align with evolving scientific needs, addressing the growing demand for seamless data integration, interoperability [2], and cross-platform collaboration.

2. Enhancing FAIR Data Practices in the NRDA

As part of its commitment to advancing Open Science, the NRDA is taking significant steps to align its data management practices with the FAIR principles—Findable, Accessible, Interoperable, and Reusable. These principles are critical for fostering a transparent, efficient, and collaborative research environment, enabling data to be effectively shared, discovered, and reused across disciplines and borders.

To enhance the FAIRness of its data, NRDA is focusing on several key areas:

1. **Data Discoverability and Access:** NRDA is dedicated to ensuring datasets are easily discoverable and accessible, both within its repository and through integration with other national and international platforms. By migrating to the CKAN platform, NRDA aims to provide more efficient data management, improve the visibility of research outputs through enhanced metadata and advanced search functionalities, and integrate with services like domain-specific repositories and scientific applications.
2. **Metadata Standardization:** In alignment with Open Science standards, NRDA is adopting standardized metadata frameworks to ensure datasets are well-described and reusable by both humans and automated systems. These efforts include the implementation of Persistent Identifiers (PIDs), such as DOIs, to enable consistent citation, tracking, and referencing of both research data and associated metadata vocabularies.

3. **Interoperability through Research Objects:** NRDA is enhancing interoperability by adopting Research Objects (ROs), enabling the bundling of all research artifacts—such as datasets, metadata, and related resources—designed to facilitate seamless sharing and reuse of research outputs. Integrating ROs into its infrastructure creates a unified framework that bridges gaps between repositories, enabling easier sharing, reuse, and referencing of datasets.
4. **Collaboration and Integration:** NRDA is actively collaborating with other platforms, including those using CKAN and others, such as Dataverse, which is widely used in Norway. These collaborations foster a cohesive data ecosystem, seamless data flow between platforms, enhancing cross-referencing and interlinking of datasets, providing researchers with broader access to datasets while ensuring efficient, reliable, and Open Science-compliant data sharing.

2.1 The role of persistent identifiers

Persistent Identifiers (PIDs) play a crucial role in ensuring the long-term accessibility, citation, and tracking of datasets within the NRDA. By assigning PIDs to all datasets, NRDA enhances the discoverability and traceability of research data, which is essential for increasing the visibility and impact of scientific work.

2.1.1 Persistent Identifiers (PIDs): Enabling Data Citations and Tracking

NRDA uses the DataCite DOI service to assign unique Digital Object Identifiers (DOIs) to datasets within its repository. This system ensures that each dataset receives a persistent, machine-readable identifier [3] that can be easily cited in academic publications and referenced across various research platforms. The adoption of DOIs aligns NRDA's data management practices with international Open Science standards, simplifying the tracking of dataset use and impact. This contributes to more robust and reliable citation practices within the academic community.

2.1.2 Draft DOIs for Early Integration in Research Workflows

In addition to assigning DOIs for published datasets, NRDA offers researchers the option to reserve draft DOIs. These draft DOIs can be assigned to datasets that are still in preparation, enabling researchers to reference the dataset in publications before it is finalized and made publicly available. This feature is particularly valuable for ensuring that datasets are seamlessly linked to related publications, allowing the dataset to be published and made accessible simultaneously with the associated paper. This practice supports greater alignment between datasets and publications, enhancing their integration into the scientific record.

2.2 Research Objects: A Key Enabler of FAIRness

A significant part of NRDA's strategy for advancing FAIR data practices is the integration of Research Objects (ROs) [4] into the archive's infrastructure. ROs are structured bundles of data, metadata, and related research artifacts that facilitate data discovery, citation, and reuse. By adopting ROs, NRDA can ensure that all components of a research project—whether datasets, code, or supplementary materials—are packaged together in a standardized, machine-readable format.

Research Objects offer several key benefits in the context of FAIR data:

- **Data Packaging and Standardization**: ROs provide a common framework for bundling research data, making it easier to manage and share related datasets and metadata. This ensures that datasets are described in a consistent way, improving their findability and usability.

- **Improved Citation and Traceability**: The use of ROs enables better citation practices by ensuring that all components of a research study are easily traceable and can be attributed to their original authors. This enhances transparency and accountability in research.
- **Seamless Data Integration**: ROs enable datasets to be more easily integrated across different platforms and repositories. This is particularly important in a global research ecosystem, where datasets need to be shared and interlinked to foster collaboration and knowledge exchange.

2.3 Ensuring Interoperability and Collaboration Between NRDA and DataVerseNO

In the Norwegian context, the role of the NRDA extends beyond data storage and management; it also fosters collaboration with other platforms to improve data discoverability and reuse. One such collaboration is with DataVerseNO, a widely used platform in Norway for archiving smaller datasets. The seamless integration between NRDA and DataVerseNO is essential for ensuring that data flows efficiently across systems, enhancing both the accessibility and interoperability of research data.

By ensuring that datasets can easily transition between NRDA and DataVerseNO, researchers are empowered to utilize both platforms based on the size and scope of their datasets. DataVerseNO's focus on smaller datasets complements NRDA's capacity to host large volumes of data, creating a complementary ecosystem for research data storage and sharing. This integration is facilitated by common standards and practices that enable smooth data exchange and metadata synchronization.

A key aspect of this collaboration is the adoption of Research Objects (ROs) and their integration with RO-Crate. Both NRDA and DataVerseNO are working towards implementing these standards, which will allow datasets and their associated metadata to be packaged and shared in a standardized, machine-readable format. This not only improves the discoverability of datasets but also facilitates the cross-referencing of data from different repositories, enabling researchers to gain deeper insights by combining multiple datasets from different sources.

This collaboration ensures that Norwegian researchers benefit from a unified, interoperable data ecosystem where datasets are easily accessible, discoverable, and reusable across both platforms. The integration between NRDA and DataVerseNO is a crucial step towards achieving broader goals of Open Science, where seamless data sharing and collaboration are key to advancing research.

3. User Impact and Challenges

The ongoing efforts to enhance the NRDA through the integration of FAIR principles, Research Objects (ROs), and RO-Crate technology hold significant potential to impact researchers, institutions, and the broader scientific community. However, these advancements come with both benefits and challenges, particularly when considering user engagement, adoption, and the scalability of the system.

3.1 User Impact

1. **Improved Data Discoverability and Accessibility:** Researchers will benefit from increased data visibility through standardized metadata and integration with external platforms like DataVerseNO. This improvement will make it easier for users to locate datasets, enhancing collaboration across disciplines and institutions. With better metadata and support for Persistent Identifiers (PIDs), datasets are easier to cite, track,

and reference in publications, ensuring greater attribution and recognition for data creators.

2. **Enhanced Reusability and Interoperability:** By adopting Research Objects and RO-Crate standards, NRDA is creating an environment that fosters seamless data sharing and reuse. Researchers can now bundle data, code, and related materials in machine-readable formats, making it easier to repurpose data for new analyses. This ensures that datasets can be more effectively reused, improving the overall value and longevity of research outputs.
3. **Streamlined Collaboration:** The integration with DataVerseNO and other platforms facilitates cross-repository data exchange, improving collaboration and enabling researchers to access a broader range of datasets without navigating complex systems. This is particularly valuable for interdisciplinary research where data from multiple sources need to be combined.
4. **Encouragement of Open Science Practices:** NRDA's alignment with FAIR principles promotes open data sharing, fostering a more transparent and reproducible research environment. Users are encouraged to adopt practices that make data available for others to build upon, thus accelerating scientific discovery and innovation.

3.2 Challenges

1. **Adoption Barriers:** Although the benefits of FAIR practices are clear, convincing researchers to adopt new data management workflows can be challenging. Many researchers may be unfamiliar with or resistant to new metadata standards, the concept of Research Objects, or the integration of persistent identifiers in their workflows. Overcoming this resistance requires effective training, outreach, and demonstration of the tangible benefits of these practices.
2. **Technical Complexity:** The integration of RO-Crate into existing systems and platforms such as CKAN introduces a layer of technical complexity. While the prototype development has shown promising results, scaling this solution to handle the large volumes of data stored in NRDA, and ensuring that it remains compatible with other systems and repositories, presents ongoing technical challenges. Furthermore, ensuring the continued stability and scalability of the infrastructure as the data ecosystem evolves is crucial.
3. **Interoperability Issues:** Despite the adoption of common standards like RO-Crate, ensuring true interoperability between diverse platforms and repositories can be difficult. Different systems may implement metadata standards in subtly different ways, leading to challenges in seamless data sharing and integration. Addressing these issues requires continuous collaboration and refinement of standards, as well as regular updates to ensure that all systems remain compatible.
4. **Resource and Training Requirements:** The transition to a more interoperable, FAIR-aligned system necessitates significant resources, both in terms of technical infrastructure and user training. Researchers, particularly those who may not be accustomed to working with complex metadata or the concept of structured research objects, will need support in understanding how to adopt these new practices. Providing comprehensive documentation, training workshops, and user support is critical to the success of this transformation.
5. **Data Quality and Consistency:** To fully realize the benefits of the FAIR principles, datasets need to be consistently well-documented, with clear metadata that enhances usability. Ensuring the consistent quality and completeness of metadata across all datasets in NRDA will require ongoing monitoring and user engagement. Researchers will need to be incentivized to adopt and maintain high standards for metadata quality, which may require periodic audits and feedback mechanisms.

4. Prototype Development: Integrating RO-Crate and CKAN

As part of its ongoing efforts to enhance the FAIRness and interoperability of its data management practices, the NRDA has undertaken the development of a prototype integrating RO-Crate with its newly implemented CKAN platform. CKAN, an open-source data management system, serves as the backbone for NRDA's infrastructure, allowing for efficient storage, retrieval, and management of research datasets. The integration of RO-Crate, a standardized format for packaging research data and metadata, aims to further improve the discoverability and accessibility of datasets stored within NRDA.

RO-Crate is an emerging standard designed to package datasets and their associated metadata in a machine-readable and human-understandable format, facilitating the exchange and reuse of research data across platforms. By adopting this format, NRDA is moving towards a more structured approach for managing research objects, which can include datasets, software, publications, and other research-related resources. This allows for better traceability, citation, and reuse of research outputs, aligning with the principles of Open Science.

The development of the prototype began with NRDA participating in the **FAIR-Impact initiative's Support offer #2, focused on enabling FAIR signposting and RO-Crate for content and metadata discovery**. This initiative provided valuable guidance and resources to integrate RO-Crate and signposting within the CKAN environment, ensuring that metadata associated with NRDA's datasets is standardized and enriched. This integration aims to improve data visibility across different platforms, making it easier for researchers to find and access datasets.

To validate the integration, a test instance was deployed using a Docker container, a lightweight and portable environment ideal for testing and deployment. The successful testing of the prototype demonstrated that RO-Crate could be effectively incorporated into CKAN, offering enhanced metadata management, better signposting for datasets, and improved data discoverability. The integration ensures that datasets stored in NRDA are not only accessible within the archive but also interoperable with other data repositories that support the RO-Crate standard.

This prototype development represents a significant milestone in NRDA's efforts to create an interoperable ecosystem for research data, linking datasets from various platforms and repositories, both national and international. The integration of RO-Crate into CKAN sets the stage for broader adoption of FAIR data practices, facilitating the seamless exchange of research data and ensuring that it can be easily integrated into other platforms and systems.

4.1 Implementation Details

Although the development and testing of the RO-Crate prototype have been outlined, additional detail on how the implementation will proceed operationally is essential to clarify the pathway for adoption and ensure successful deployment.

4.1.1 Implementation Phases

The implementation of the RO-Crate integration into NRDA will take place in clearly defined phases to ensure a structured and manageable transition from prototype to full-scale operational system. The first phase will involve the finalization of the RO-Crate prototype, including testing with a select group of early adopters. This will be followed by broader testing across a wider range of datasets to assess scalability and interoperability with other repositories like DataVerseNO.

Subsequent phases will focus on full integration into the NRDA ecosystem, expanding compatibility with additional platforms and services. Each phase will include a feedback loop

where user input will guide refinements and adjustments to both the technical system and user workflows.

4.1.2 Timelines

The timeline for the implementation will be divided into quarterly milestones, with the initial phase of prototype finalization and testing expected to take 3 months. Afterward, integration with NRDA will begin, followed by scaling efforts to support larger datasets and expand to more external repositories. The goal is to achieve full operational status within 12–18 months, with ongoing improvements based on user feedback.

4.1.3 Support Structures

To ensure smooth adoption of the RO-Crate system, a comprehensive support infrastructure will be put in place. This will include:

- **User Training:** Training materials, workshops, and webinars will be developed to guide researchers through the adoption of the new metadata standards and integration of Research Objects into their workflows.
- **Technical Support:** We will be using the NRDA support email to assist users with technical issues, data upload challenges, and integration questions.
- **Documentation:** Comprehensive documentation will be provided, covering not only the technical aspects of RO-Crate but also practical guides for researchers on how to create and manage research objects effectively.

The support structures will be continuously evaluated and adjusted based on user feedback to ensure that researchers have the resources they need to succeed with the new system.

5. Future Directions: Operational Deployment and Broader Implications

The integration of RO-Crate into the Norwegian Research Data Archive (NRDA) represents a pivotal advancement toward enhanced interoperability and adherence to FAIR (Findable, Accessible, Interoperable, Reusable) principles. This integration lays the groundwork for improving metadata management, discoverability, and data citation within NRDA, fostering transparency, reproducibility, and collaboration in the research ecosystem.

Operational deployment of RO-Crate will standardize data packaging, enabling seamless linking across platforms and facilitating long-term data preservation. By embedding RO-Crate standards, NRDA will empower researchers to more easily cite and track datasets, contributing to the creation of a more transparent and traceable scientific record.

5.1 Broader Implications

NRDA’s adoption of RO-Crate offers a scalable model for implementing FAIR practices in research data archives. By demonstrating the feasibility of integrating standardized formats, NRDA is paving the way for broader adoption across the international research community. This initiative also strengthens interoperability and aligns with global Open Science goals, promoting seamless data sharing across disciplines and borders.

5.2 Future Directions

NRDA’s commitment to Open Science will continue to shape its evolution. Key areas for future focus include:

- **FAIR Digital Objects (FDOs)** ([5], [6], [7]) **:** Building on RO-Crate capabilities to integrate FDOs for enhanced metadata and data sharing, which we will implement when relevant for our users and our archive.
- **Emerging Technologies:** Exploring AI/ML for metadata generation and intelligent search, as well as decentralized systems for data provenance and trust.
- **Advanced FAIRification Tools:** Simplifying the creation of FAIR-compliant datasets through automation and enrichment.

Ongoing collaboration with the Open Science community will ensure NRDA remains responsive to emerging trends and user needs. By prioritizing innovation and interoperability, NRDA aims to position Norwegian researchers at the forefront of Open Science, contributing to a sustainable and accessible global research ecosystem.

6. Conclusion

The integration of Research Objects (ROs) and the adoption of RO-Crate within the Norwegian Research Data Archive (NRDA) represents a significant advancement in enhancing the archive's alignment with FAIR (Findable, Accessible, Interoperable, and Reusable) principles. By embracing these standards, NRDA is not only improving its own data management and accessibility but also contributing to the broader Open Science movement. Through this initiative, NRDA is ensuring that research data is more easily discoverable, interoperable across platforms, and reusable, thus helping to foster collaboration both within Norway and on an international scale.

The successful development and testing of the RO-Crate prototype demonstrates NRDA's commitment to continuously evolving its infrastructure to meet the needs of the research community. By focusing on interoperability, data discoverability, and the seamless integration of data across platforms like DataVerseNO, NRDA is helping to create a more integrated and sustainable data ecosystem. This will ultimately enhance the quality and impact of research by enabling more efficient data sharing and reuse, providing researchers with the tools and resources they need to collaborate and innovate.

Looking ahead, the operational deployment of RO-Crate in NRDA will play a crucial role in further strengthening the archive's capacity to support Open Science practices. As NRDA continues to refine its infrastructure, the lessons learned from this initiative will provide valuable insights for other research data repositories seeking to implement FAIR principles and foster more open, accessible, and interoperable research data management.

Through these efforts, NRDA is playing a key role in advancing global Open Science practices and ensuring that research data is managed in a way that benefits not only researchers but also the wider scientific community and society at large.

Data availability statement

All the data and information mentioned in this paper are made available in a Research Object stored in the ROHub Research Object Hub at https://w3id.org/ro-id/7733ebf0-b1be-4522-a909-8089251749c4 .

Author contributions

Adil Hasan is in charge of the NRDA and serves as the project administrator, overseeing the overall coordination and management of the project. Anne Fouilloux contributed to all other aspects of the work, including the development, implementation, and testing of the system enhancements, as well as drafting the article.

Competing interests

The authors declare that they have no competing interests.

Funding

This initiative has been funded through the open call "Support offer #2: Enabling FAIR Signposting and RO-Crate for content/metadata discovery and consumption," which enabled the development and testing of the integration of RO-Crate and signposting within the NRDA infrastructure.

Acknowledgement

We would like to thank the FAIR-Impact project for its support throughout this work. We are especially grateful to Stian Soiland-Reyes for his expert guidance on RO-Crate and to Herbert Van de Sompel for his support on the implementation of signposting. Their guidance and resources were essential in making these advancements possible.

References

[1] 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

[2] Sansone, SA., Rocca-Serra, P., Field, D. *et al.* Toward interoperable bioscience data. *Nat Genet* 44, 121–126 (2012). https://doi.org/10.1038/ng.1054

[3] Juha Hakala. (2010). Persistent Identifiers - An Overview. *TWR Technology Watch Review,* http://www.metadaten-twr.org/2010/10/13/persistent-identifiers-an-overview/

[4] Bechhofer, S., De Roure, D., Gamble, M. *et al.* Research Objects: Towards Exchange and Reuse of Digital Knowledge. *Nat Prec* (2010). https://doi.org/10.1038/npre.2010.4626.1

[5] Koenraad De Smedt, Dimitris Koureas, and Peter Wittenburg. An analysis of scientific practice towards FAIR digital objects. DOI: 10.23728/b2share.e14269d07ce84027a7f79ee06b994ef9.

[6] Koenraad De Smedt, Dimitris Koureas, and Peter Wittenburg. “FAIR Digital Objects for Science: From Data Pieces to Actionable Knowledge Units.” In: Publications 8.2 (2020). DOI: 10.3390/publications8020021.

[7] FAIR Digital Objects Forum.