

RADAR: Building a FAIR and Community Tailored Research Data Repository

Felix Bach¹, Kerstin Soltau¹, Sandra Göller¹, Christian Bonatto Minella¹ and Stefan Hofmann¹

¹ e-Research Institute, FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Germany

Abstract. The research data repository [RADAR](#) is designed to support the secure management, archiving, publication and dissemination of digital research data from completed scientific studies and projects. Developed as a collaborative [project funded](#) by the [Deutsche Forschungsgemeinschaft](#) (DFG, German Research Foundation) (2013-2016), the system is operated by [FIZ Karlsruhe](#) - Leibniz Institute for Information Infrastructure - and currently serves as a generic cloud service for about 20 universities and non-university research institutions. Since its launch, RADAR has witnessed significant changes in the landscape of research data repositories and the evolving needs of researchers, research communities and institutions. In our presentation within the “Enabling RDM” Track, we will show how RADAR is responding to these dynamic changes. In order to create a sufficiently large user base for the sustainable operation of the system, we have moved RADAR away from its previous single focus on a discipline-agnostic cloud service and towards a demand-driven functional optimisation. In 2021, we introduced an additional operating model for institutions ([RADAR Local](#)), where we operate a separate RADAR instance locally at the institution site exclusively using the institutional IT-infrastructure. In 2022 we opened up RADAR to new target groups with community-specific service offerings, in particular in the context of the [National Research Data Infrastructure](#) (NFDI). Beside the expansion of the functional scope, our ongoing development work focuses also on strengthening the system's support for the FAIR principles [1] and the concepts of [FAIR Digital Objects](#) (FDO) [2] and [Schema.org](#). Our presentation will outline recent RADAR developments and achievements as well as future plans thus providing solutions and synergy potential for the scientific community and for other service providers.

Keywords: NFDI, Chemistry, Culture, Infrastructure

1. Enhancements of the RADAR metadata schema

Aiming to enrich the descriptive metadata and strengthen the implementation of the FAIR principles, two important standard data - the [Research Organisation Registry](#) (ROR) and the “[Gemeinsame Normdatei](#)” (GND, Integrated Authority File) - were implemented into the generic [RADAR metadata schema](#) in December 2023. Besides, further controlled vocabularies (e. g. for licence information and related resources) were introduced. These enhancements promote in particular findability and interoperability of published datasets. For the same reason, the integration of the [TIB Terminology Service](#), which helps encoding the [DFG Classification of Subject Areas Ontology](#) (DFGFO) into an ontology, is currently under consideration as a useful standardisation measure for a future update of the RADAR schema. Moreover, we intend to explore different FAIR assessment tools to optimise and finetune our future developments.

2. Subject-specific metadata schemas

In 2022, we launched our community specific offerings [RADAR4Chem](#) and [RADAR4Culture](#) to provide researchers in the fields of chemistry ([NFDI4Chem](#)) and cultural studies ([NFDI4Culture](#)) with free data publication services. Similar sustainable research data publishing offerings in other scientific communities are underway.

It is especially in these subject-specific contexts that the established generic RADAR metadata schema needs to be opened up. Currently, this requirement is already supported by an upload option for subject-specific metadata files (XML) that are validated against configurable schemas given as XSD files.

However, we aim to enable those metadata annotations via more user friendly input masks which is currently ongoing work. For this purpose, we are extending the RADAR backend to support JSON based metadata (i.e. JSON-LD) in addition to XML, which will give users greater flexibility in creating and/or importing their own discipline-specific metadata in manifold schemas into RADAR. Nevertheless, a great challenge for all subject-based repositories in this context is to agree upon standards in the rapidly changing landscape and to make the existence of subject-specific metadata visible in evidence systems for publications such as DataCite that are focused on generic metadata.

3. Embedded metadata on dataset landing page

In addition to submitting descriptive metadata of published datasets to DataCite, we offer metadata harvesting via our [OAI provider](#) for all RADAR data publications, including discipline-specific metadata, according to the standardised [Open Archives Initiative Protocol for Metadata Harvesting](#) (OAI-PMH). This supports the dissemination, visibility and discoverability of RADAR research data and enables research communities to integrate RADAR content into specialist portals or knowledge graphs. However, findability, interoperability and in particular machine-readability are further maximised by applying established standards such as Schema.org. Moreover, new alternative approaches such as data visiting, Fair Digital Objects (FDO), and [FAIR signposting](#) are gaining significance as promising ways to increase machine-actionability, e.g. to allow distributed machine learning.

Data visiting targets landing pages directly, allowing machines to dynamically access both metadata and data. FAIR signposting takes a low-threshold approach that allows machines to interact with scholarly portals in a unified way, bypassing proprietary APIs and following typed links (e. g. DOI, author PIDs, web addresses of landing pages and descriptive metadata) in the [HTTP link headers](#). Late 2022, we enriched the source code of all dataset landing pages in RADAR, improving their machine readability and -actionability.

In the following software releases in early 2023, we optimised our embedded metadata into the landing pages for datasets using Schema.org annotation in JSON-LD and in Turtle format relevant for facilitating search, integration, and analysis of metadata of research datasets. In the future, we also plan to evaluate the [RO-Crate](#) approach (a community effort to practically achieve FAIR packaging of research objects) with their structured metadata to optimise our current preservation container solution based on [BagIt-standard](#).

4. Summary

In summary, the RADAR data repository has evolved to meet the changing needs of researchers and institutions, with a focus on demand-driven functional optimisation, community-specific service provision and increased support for concepts such as the FAIR principles, Schema.org and FDO. Subject-specific metadata annotation and the exploration of alternative approaches in contrast to traditional metadata harvesting have been achieved, nevertheless developments

continue. Future work includes the integration of further community-specific terminologies and ontologies as well as the exploration of FAIR assessment tools to improve research data published in RADAR.

Overall, RADAR aims to provide a complete solution for the management, archiving and publishing of research data for the scientific community and supporting best practises for other service providers.

Data availability statement

-

Underlying and related material

-

Author contributions

-

Competing interests

Authors declare no competing interests.

Funding

-

Acknowledgement

We would like to extend our gratitude to NFDI4Chem (DFG) – 441958208 and NFDI4Culture (DFG) – 441958017 for their financial support.

References

1. Wilkinson M., Dumontier M., Aalbersberg I., The FAIR Guiding Principles for scientific data management and stewardship (2016). *Sci Data* 3: 160018. <https://doi.org/10.1038/sdata.2016.18>
2. Schultes, E., Wittenburg, P. FAIR Principles and Digital Objects: Accelerating Convergence on a Data Infrastructure. In: Manolopoulos, Y and Stupnikov, S (eds.), *Data Analytics and Management in Data Intensive Domains. DAMDID/RCDL 2018*. (2018) *Communications in Computer and Information Science*, 1003. Cham: Springer. DOI: <https://doi.org/10.1007/978-3-030-23584-0>