New Developments in RADAR: Safeguarding and Publishing Research Data for Long-term Usability

Kerstin Soltau [1,*], Stefan Hofmann [1], Sandra Göller [1], Felix Bach [1]

[1] FIZ Karlsruhe – Leibniz-Institut für Informationsinfrastruktur
[*] Corresponding author: info@radar-service.eu

RADAR[1], launched in 2017 as a cloud service, is a generic research data repository designed to securely archive, publish, and optionally review research data. Developed in a cooperation project funded by the DFG, it is now operated by FIZ Karlsruhe – Leibniz Institute for Information Infrastructure[2]. RADAR stores data according to the BagIt standard in a geo-redundant manner, maintaining three copies across two academic data centers, located at KIT[3] and TU Dresden[4] using tape archives as storage medium. With a minimum retention period of 25 years for published data and flexible archival options compliant with the DFG Code of Conduct, RADAR serves a variety of scientific disciplines. Agile software development has driven the development of the system, resulting in numerous improvements in terms of metadata annotation and FAIRness of metadata (e.g. OAI provider, FAIR signposting, Schema.org integration, a knowledge graph, SPARQL endpoints etc.) as well as functional enhancements, especially in terms of uploading research data (e.g. WebDAV protocol support and Git integration). In addition, community-tailored publication services developed in the framework of the NFDI[5] as well as alternative operating variants (RADAR Local and Hybrid) expand the range of services offered by RADAR and meet the needs of additional target groups. This paper provides an overview of RADAR's technical architecture, service enhancements, and its significant role in advancing research data management.

---

1 https://www.radar-service.eu/; *Visited on May 16, 2025*

2 https://www.fiz-karlsruhe.de; *Visited on May 16, 2025.*

3 https://www.kit.edu/; *Visited on May 16, 2025.*

4 https://tu-dresden.de/; *Visited on May 16, 2025.*

5 https://www.nfdi.de/; *Visited on May 16, 2025.*

1 Introduction

The exponential growth of research data over recent decades has necessitated robust, secure, and long-term data management solutions. Research institutions worldwide are increasingly challenged by the need to archive and share datasets while ensuring compliance with policies related to data retention, security, and accessibility. In this context, RADAR was conceived as a cross-disciplinary research data repository to address these challenges. Initially funded by the German Research Foundation (DFG)[6] during a 3-year project phase and launched as a cloud service for academic institutions in 2017, RADAR represents a significant step towards establishing an integrated infrastructure that supports the findability, accessibility, interoperability, and reusability of research data.

The primary objective of RADAR is to provide a secure, scalable, and user-friendly platform for archiving and publishing research data independent of scientific discipline. By leveraging cloud technologies and reliable archival techniques, RADAR offers a comprehensive solution tailored to the needs of research institutions, scientific communities, individual researchers, regional networks, and funding organizations. This paper presents an overview of the system architecture and data integrity, metadata management and the implementation of the FAIR principles, recent functional enhancements and user-focused features, and community-specific services and RADAR operating variants.

2 System Architecture and Data Integrity

RADAR was designed as a cloud-based service to address the challenges inherent in managing long-tail research data. The infrastructure is built on a modular architecture that ensures both scalability and resilience. At the heart of this system is a secure archival process that employs geo-redundancy – each dataset is stored in three copies across two different academic data centers at SCC[7] and ZIH[8]. This approach guarantees data integrity and availability even in the event of hardware failures and also mitigates risks associated with natural disasters or localized cyberattacks. The primary storage medium for RADAR are tape archives. Though often slower in terms of data retrieval compared to disk-based systems, it offers significant cost benefits and stability for archival purposes in the long term. The system is engineered such that the retention period for

---

6 https://www.dfg.de; *Visited on May 16, 2025.*

7 https://www.scc.kit.edu; *Visited on May 16, 2025.*

8 https://tu-dresden.de/zih; *Visited on May 16, 2025.*

published data is set to a minimum of 25 years, while archived data can be stored for flexible periods (e.g. 10 years in accordance with the DFG Code of Conduct[9]).

Data integrity is a key aspect of RADAR's design. Datasets are provided with a checksum (MD5 hash value) before being transferred to a tape archive. This checksum is validated after each copying process. The data centers also validate the checksums internally after migration to new tapes.

3 Metadata Management and FAIR Principles

A dataset annotated with meaningful and interoperable metadata increases its visibility, discoverability, reusability and reproducibility. Thus effective data management goes hand in hand with comprehensive metadata annotation. RADAR has developed its own generic metadata schema which is listed as FAIRsharing Standard (FAIRsharing Team 2023). It is based on the DataCite Metadata Kernel and comprises 10 mandatory and 13 optional elements. Over time, however, it became obvious that this generic approach is not sufficient for all use cases and that sometimes researchers need to describe their data according to their discipline- or project-specific needs. To meet these needs, in 2021 RADAR introduced a flexible subject-specific annotation option for custom defined XSD schemas that are either RADAR-compliant or generate the mandatory RADAR metadata by an additionally provided XSLT transformation.

The FAIR principles are integral to modern research data management. So far, RADAR has already implemented several approaches to maximise the FAIRness of datasets:

- **Operation of a dedicated OAI provider**, via which the metadata of all data publications are available for harvesting by third parties in various formats.
- **Integration of standard data in RADAR metadata**, e.g. for individuals (ORCID), organisations (ROR), funding organisations (Crossref Funder Registry), and keywords (GND, the Integrated Authority File).
- **Schema.org Integration**: Adoption of this widely recognized metadata standard, facilitating broader machine-readability, discoverability, and interoperability by internet search engines, like Google Dataset Search.
- **FAIR Signposting**: Enrichment of the source code of landing pages with typed links, ensuring machine readability and actionability in a uniform way by making data resources easy to find and reference.
- **RADAR Knowledge Graph and SPARQL Endpoint**: The development of a RADAR knowledge graph enables the semantic linking of data elements, while the SPARQL endpoint[10] provides researchers with powerful querying capabilities.

---

9 https://www.dfg.de/en/basics-topics/basics-and-principles-of-funding/good-scientific-practice/code-of-conduct; *Visited on May 16, 2025.*

10 https://www.radar-service.eu/sparql; *Visited on May 16, 2025.*

In 2023, RADAR participated in a support measure of the EU project FAIR-IMPACT (Soltau et al. 2024). During this FAIRness assessment process, the *F-UJI score*[11] as indicator for the FAIRness of data publications was significantly increased (up to 91 %). Starting in spring 2025, RADAR integrates the *TS4NFDI terminology service*[12] developed as part of the NFDI, which allows to include a large number of different discipline specific terminologies and terminology collections in several metadata fields (keywords, subject area etc.).

4 Functional Enhancements and User-Focused Features

To accommodate the various workflows of researchers and to allow data transfer even for larger data volumes, RADAR continuously evolves its data ingestion methods. Since early 2024, an alternative upload option via the *WebDAV protocol* has been available. This method offers a familiar interface for users who prefer a flexible, network drive-like process that simplifies the upload process also for larger data volumes. End of 2024, RADAR introduced *Git integration* allowing researchers to seamlessly import content from both public and private GitHub repositories into RADAR. Integration with GitLab will be released in the first half of 2025.

Further, RADAR will soon offer more flexible access to research data and expand its download options. Currently, datasets stored on tape are only available for download as packed .tar files. Around mid 2025, users will be able to *download individual files or folders* from these datasets via selective streaming. In future, this enhancement will also enable data previews and facilitate data transfer prior to versioning. At the same time, access to data by external reviewers will be simplified by additionally offering the dataset content as downloadable .zip file.

The RADAR team follows an *agile software development* workflow, continuously releasing improvements based on user feedback, discussions and prioritization with the RADAR User Advisory Council, as well as emerging needs and developments in the research data ecosystem, including the NFDI. *Future improvements* include the versioning of datasets, enhanced search and terminology driven functionalities, optional integration of artificial intelligence for automated metadata annotation, and integration with other research data management tools and authentication infrastructures like IAM4NFDI[13].

---

11 https://www.f-uji.net/; *Visited on May 16, 2025.*
12 https://base4nfdi.de/projects/ts4nfdi; *Visited on May 16, 2025.*
13 https://www.nfdi-aai.de/; *Visited on May 16, 2025.*

5 Community-Specific Service Offerings and Alternative Operating Variants

Over time, FIZ Karlsruhe has not only continuously enhanced RADAR functionality, but also expanded the service to new target groups and introduced alternative operating variants for use by academic institutions.

FIZ Karlsruhe is actively involved in several NFDI consortia. Analyses in NFDI4Chem, NFDI4Culture, and NFDI4Memory demonstrated the need for a generic repository based on RADAR to complement the portfolio of existing specialized repositories. So far, FIZ Karlsruhe has developed three community-specific data publication services, which are free of charge and directly address researchers independent from their institutional affiliation. The RADAR team will continuously enhance these services, for example by integrating discipline-specific terminologies and metadata standards developed by the respective consortia:

- **RADAR4Culture**[14]: for communities in the field of cultural heritage.
- **RADAR4Chem**[15]: for the chemistry community.
- **RADAR4Memory**[16]: for the historically working humanities.

Further, in response to the needs of academic institutions and their diverse IT environments, FIZ Karlsruhe launched alternative operating variants in 2021. Unlike the cloud service, these allow to integrate own IT resources within the RADAR system:

- **RADAR Local**: Enables institutions to host a local RADAR instance on their own IT and under their own domain operated by FIZ Karlsruhe, providing complete control over data management.
- **RADAR Hybrid**: Allows institutions to store research data in an alternative archive storage (either locally or externally), while continuing to use RADAR´s cloud-based software and infrastructure.

As of spring 2025, seven RADAR Local instances are either operated by FIZ Karlsruhe or currently in the implementation phase. The instance at the University of Potsdam is being expanded into a cross-university research data portal for the federal state of Brandenburg, as part of the project FDM-BB[17]. This centrally provided, state-wide data repository based on RADAR could serve as a model for other federal states.

---

14 https://radar4culture.radar-service.eu/; *Visited on May 16, 2025.*

15 https://radar4chem.radar-service.eu/; *Visited on May 16, 2025.*

16 https://radar4memory.radar-service.eu/; *Visited on May 16, 2025.*

17 https://fdm-bb.de/; *Visited on May 16, 2025.*

6 Conclusion

All future developments remain firmly aligned with our core principles: long-term data security and integrity, demand-driven and sustainable software development, and the FAIR principles. To do so, we are in close consultation with the institutions using RADAR and with the NFDI consortia in which we participate, and are committed to the ongoing integration of RADAR into regional, national and the European research data ecosystems.

Acknowledgements

Authorship Contributions

The basic structure of the text was created with the help of ChatGPT.

Conflict of Interest

The authors have no conflicts of interest to declare.

Bibliography

FAIRsharing Team. 2023. *RADAR Metadata Schema.* FAIRsharing record. https://doi.org/10.25504/FAIRsharing.e26f92.

Soltau, Kerstin, Stefan Hofmann, Agnes Jasinska, and Clara Linés. 2024. *FAIRness assessment as a catalyst of chemistry research repository transformation in Germany.* Zenodo. https://doi.org/10.5281/ZENODO.13135009.