RADAR: Enhancing FAIR Research Data Management with AI Support

Kerstin Soltau[1], Stefan Hofmann[1], and Felix Bach[1]

[1]FIZ Karlsruhe - Leibniz Institute for Information Infrastructure, Eggenstein-Leopoldshafen, Germany
*Corresponding author: info@radar-service.eu.

Motivation

RADAR [1], developed and operated by FIZ Karlsruhe, is an established interdisciplinary repository for the archiving, publication, and long-term preservation of research data. Since its launch in 2017, RADAR has continuously evolved to meet the growing demands of Open Science and to promote the implementation of the FAIR data principles [2].

The repository supports persistent identifiers (e.g, DOI, ORCID, ROR, GND), comprehensive metadata, and semantic enrichment (e.g., Schema.org and FAIR Signposting). It also provides access to discipline-specific terminologies via TS4NFDI (Terminology Services 4 NFDI) [3] and integrates seamlessly with platforms such as GitHub and GitLab. Flexible deployment options (RADAR Cloud and RADAR Local) and community-specific services – such as RADAR4Chem, RADAR4Culture, and RADAR4Memory – ensure wide applicability and alignment with the needs of diverse research domains.

To further strengthen FAIR data implementation, the RADAR team is currently investigating, in consultation with the RADAR User Advisory Council, the use of artificial intelligence (AI) to assist researchers in improving metadata quality and FAIR compliance. Two complementary developments are currently being pursued:

- ***FAIRness Assessment***: to motivate data providers to enhance datasets with comprehensive and semantically rich metadata.
- ***Metadata Enhancement***: to support data providers in enriching datasets / files with significant and semantic metadata derived from metadata, file contents, and external resources.

Both initiatives aim to reduce manual effort while improving metadata completeness, interoperability, and overall data stewardship.

Methods, Tools and Evaluation

In our exploratory work, we examined two complementary use cases where AI can add tangible value to repository services: FAIRness assessment and metadata enhancement. For each, we tested two methodological approaches. Method 1 employed generic large language models (LLMs) such as ChatGPT and Mistral, using prompt engineering to generate FAIRness evaluations and enrich metadata. Method 2 focused on open-source, task-specific tools — the ***Fair-Way Service*** for FAIRness assessment and ***KeyBERT/KeyLLM*** [4] for keyword extraction and metadata enhancement.

The Fair-Way Service returns FAIRness scores with detailed feedback according to the FAIRsFAIR Data Object Assessment Metrics v0.5 [5], though not yet all checks are implemented. Based on KeyBERT, we set up a ***RADAR Keyword Service*** in which keywords are extracted from metadata (and, at a later stage, also from external resources and file content). These extracted keywords are then mapped to semantic terms via the TS4NFDI Terminology Service, newly integrated into RADAR.

Our evaluation showed that generic LLMs provide flexibility but face limitations in reproducibility, interpretability, privacy, and cost. In contrast, the open-source tools proved more transparent, stable, and easier to integrate into existing repository workflows.

Consequently, we selected Method 2 for both

use cases and adopted a ***self-hosted AI infrastructure*** at FIZ Karlsruhe (ai.radar-service.eu). This setup ensures data protection, cost efficiency, and reproducibility, operating with 20 GB GPU capacity and supporting both small and large language models as well as domain-specific AI models.

Outlook

By integrating trustworthy, domain-aware AI services into repository workflows, RADAR aims to further strengthen its position as a FAIR-by-design repository that supports sustainable and reproducible Open Science practices.

In the next phase, we plan to test these AI services with domain-specific data and further develop AI-supported functionalities. Planned work includes refining the FAIRness metrics coverage, expanding the RADAR Keyword Service to more metadata elements, and evaluating specialized small language models (SLMs) for community repositories such as RADAR4Chem and RADAR4Culture.

The poster presents the methods applied so far, compares their practical advantages and disadvantages, and illustrates the resulting technical implementation and current status of work in progress.

References

[1] (2025) re3data.org: Radar. [Online]. Available: https://doi.org/10.17616/R3ZX96

[2] M. Wilkinson, M. Dumontier, I. Aalbersberg *et al.*, "The fair guiding principles for scientific data management and stewardship," *Sci Data 3*, 160018 (2016). doi: 10.1038/sdata.2016.18 .

[3] R. Baum, S. Fillies, N. Karam *et al.* (2025) Technical architecture of terminology services 4 nfdi (ts4nfdi). From: Zenodo Publication, 2025. [Online]. Available: https://doi.org/10.5281/zenodo.15182339

[4] M. Grootendorst. (2025) Maartengr/keybert. From: Zenodo Publication, 2025. [Online]. Available: https://doi.org/10.5281/zenodo.4461264

[5] A. Devaraju, R. Huber, M. Mokrane *et al.* (2025) Fairsfair data object assessment metrics. From: Zenodo Publication, 2025. [Online]. Available: https://doi.org/10.5281/zenodo.3775793