

Advancing Persistent Identifier-based Access in RADAR: Machine-Actionable Data, RO-Crate Integration, and AI-Support for Metadata Enhancement

Authors:

Jan Schweikert*¹, Felix Bach¹, Kerstin Soltau¹, Stefan Hofmann¹, Sergio Pollo Vázquez¹

* info@radar-service.eu

¹ FIZ Karlsruhe – Leibniz Institute for Information Infrastructure, Eggenstein-Leopoldshafen, Germany

Full Abstract (< 500 Words):

RADAR is a cross-disciplinary, FAIR-aligned repository service operated by FIZ Karlsruhe – Leibniz Institute for Information Infrastructure. It provides long-term, format-independent archiving and publication of digital research data across scientific disciplines. While datasets in RADAR are already identified by persistent identifiers and accessible via the WebUI and complete TAR downloads, the internal contents of archived datasets have long remained only partially machine-actionable. This has limited fine-grained referencing, automated reuse, and interoperability in PID-based data ecosystems.

To address this, RADAR has extended its metadata and access framework to support more granular and machine-readable identification of digital objects within published datasets. Streaming-based access to individual files and folders within TAR archives (BagIt) has been available since December 2025, enabling users and machines to retrieve specific data objects without downloading entire datasets. Building on RADAR's DOI-based identification scheme, fragment identifiers enable persistent, fine-grained referencing of files and folders within datasets.

In parallel, RADAR has introduced support for the RO-Crate metadata standard. For each published dataset, an accompanying *ro-crate-metadata.json* is automatically generated, describing the internal dataset structure (files, directories, size, format, timestamps) alongside dataset-level metadata such as title, license, and creators. Combined with DOI-based fragment identifiers, this establishes a machine-actionable layer in which dataset components become precisely identifiable, citable, and directly accessible via PID resolution.

Building on this machine-actionable access layer, RADAR integrates the Terminology Services for NFDI (TS4NFDI) to strengthen the semantic interoperability of PID-identified objects. In an initial implementation, the metadata field “keywords” is linked to standardised terminologies and semantic artefacts provided via TS4NFDI. This supports consistent annotation, improves interpretability across disciplines, and facilitates integration into knowledge graphs and other semantic infrastructures. The approach establishes a foundation for extending terminology-based annotation to other RADAR metadata fields, e. g. subject area or rights information.

Complementing these developments, RADAR explores AI-supported services for metadata enrichment and FAIRness assessment, including automated keyword extraction from descriptive metadata and text-based dataset contents. These approaches are currently in an exploratory phase but demonstrate strong potential to enhance metadata quality and support users in improving the FAIRness of their datasets.

Together, these developments demonstrate the value of persistent identifiers for all stakeholders: enabling granular citation for researchers, efficient data access for machines and applications, and improved interoperability for infrastructures. RADAR thereby strengthens its role within NFDI, EOSC, and the evolving FAIR Digital Object ecosystem, contributing to keeping knowledge connected through PID-based, machine-actionable research data.

Short Abstract (< 150 Words):

RADAR is a FAIR-aligned, cross-disciplinary repository service operated by FIZ Karlsruhe, providing long-term archiving and publication of research data. While datasets are identified by persistent identifiers and accessible via TAR downloads, their internal structure has long remained only partially machine-actionable, limiting fine-grained reuse.

RADAR addresses this by enabling streaming access to files within TAR archives and introducing DOI-based fragment identifiers for persistent, granular referencing. In parallel, RO-Crate metadata (*ro-crate-metadata.json*) is generated for each dataset, describing its internal structure and enabling machine-actionable access.

Further, RADAR integrates TS4NFDI to link keywords to standardised terminologies, improving semantic interoperability and enabling integration into knowledge graphs. AI-supported metadata enrichment and FAIRness assessment are currently explored.

These developments demonstrate how persistent identifiers enable granular access, interoperability, and reuse, supporting more connected and machine-actionable research data ecosystems.