

Current Developments in the Research Data Repository RADAR

Felix Bach[‡], Kerstin Soltau[‡], Sandra Göller[‡], Christian Bonatto Minella[‡]

[‡] FIZ Karlsruhe - Leibniz Institute for Information Infrastructure, Karlsruhe, Germany

Corresponding author: Felix Bach (Felix.Bach@fiz-karlsruhe.de), Kerstin Soltau (kerstin.soltau@fiz-karlsruhe.de)

Abstract

RADAR is a cross-disciplinary internet-based service for long-term and format-independent archiving and publishing of digital research data from scientific studies and projects. The focus is on data from disciplines that are not yet supported by specific research data management infrastructures. The repository aims to ensure access and long-term availability of deposited datasets according to FAIR criteriaWilkinson et al. 2016 for the benefit of the scientific community. Published datasets are retained for at least 25 years; for archived datasets, the retention period can be flexibly selected up to 15 years. The RADAR Cloud service was developed as a cooperation [project funded](#) by the DFG (2013-2016) and started operations in 2017. It is operated by [FIZ Karlsruhe](#) - Leibniz-Institute for Information Infrastructure.

As a distributed, multilayer application, RADAR is structured into a multitude of services and interfaces. The system architecture is modular and consists of a user interface (frontend), management layer (backend) and storage layer (archive), which communicate with each other via application programming interfaces (API). This open structure and the access to the APIs from outside allows integrating RADAR into existing systems and work processes, e. g. for automated upload of metadata from other applications using the RADAR API. RADAR's storage layer is encapsulated via the Data Center API. This approach guarantees independence from a specific storage technology and makes it possible to integrate alternative archives for the bitstream preservation of the research data.

The data transfer to RADAR takes place in two steps: In the first step, the data is transferred to a temporary work storage. The ingest service accepts individual files and packed archives, optionally unpacks them while retaining the original directory structure and creates a dataset. For each file found, the MIME Type (see [Multipurpose Internet Mail Extensions](#) specification)) is analysed and stored in the technical metadata. When archiving and publishing, a dataset is created in the second step. The structure of this dataset - the AIP (archival information package) in the sense of the [OAIS](#) standard -

corresponds to the [BagIt](#) standard. It contains, in addition to the actual research data in original order, technical and descriptive metadata (if created) for each file or directory as well as a manifest within one single TAR ("tape archive", a [unix archiving format and utility](#)) file as an entity in one place. This TAR file is stored permanently on magnetic tapes redundantly in three copies at different locations in two academic computing centres.

The FAIR Principles are currently being given special importance in the research community. They define measures that ensure the optimal processing of research data, accessibility for both humans and machines, as well as reusability for further research. RADAR also promotes the implementation of the FAIR Principles with different measures and functional features, amongst others:

Descriptive metadata are recorded using the internal [RADAR Metadata Schema](#) (based on [DataCite Metadata Schema 4.0](#)), which supports 10 mandatory and 13 optional metadata fields. Annotations can be made on the dataset level and on the individual files and folders level. A [user licence](#) which rules re-use of the data, must be defined for each dataset. Each published dataset receives a DOI which is registered with [DataCite](#). RADAR metadata uses a combination of controlled lists and free text entries. Author identification is ensured by using an ORCID ID and funder identification by [CrossRef Open Funder Registry](#). More interfacing options, e.g. [ROR](#) and the Integrated Authority File ([GND](#)) are currently implemented. Datasets can be easily linked with other digital resources (e.g. text publications) via a "related identifier". To maximise data dissemination and discoverability, the metadata of published datasets are indexed in various formats (e.g. DataCite and [DublinCore](#)) and offered for public metadata harvesting e.g. via an [OAI-provider](#).

These measures are - to our minds - undoubtedly already significant, but not yet sufficient in the medium to long term. Especially in terms of interoperability, we see development potential for RADAR. The [FAIR Digital Object \(FDO\) Framework](#) seems to offer a promising concept, especially to further promote data interoperability and to close respective gaps in the current infrastructure and repository landscape.

RADAR aims to participate in this community driven approach also in its role within the National Research Data Infrastructure ([NFDI](#)). As part of the NFDI, RADAR already plays a relevant role as a generic infrastructure service in several NFDI consortia (e.g. [NFDI4Culture](#) and [NFDI4Chem](#)). With [RADAR4Chem](#) and [RADAR4Culture](#), FIZ Karlsruhe for example offers researchers from chemistry and the cultural sciences low-threshold data publication services based on RADAR. We successively develop these services further according to the needs of the communities, e.g. by integrating and linking them with subject-specific terminologies, by providing annotation options with subject-specific metadata or by enabling selective reading or previewing options for individual files in existing datasets.

In our presentation, we would like to describe the present and future functionality of RADAR and its current level of FAIRness as possible starting points for further discussion with the FDO community with regard to the implementation of the FDO framework for our service.

Keywords

FAIR data, data publication, NFDI, NFDI4Chem, NFDI4Culture, interoperability

Presenting author

Felix Bach

Presented at

First International Conference on FAIR Digital Objects, presentation, Leiden, Netherlands, 2022

Conflicts of interest

References

- Wilkinson M, Dumontier M, Aalbersberg I, (2016) The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3: 160018. [In English]. <https://doi.org/10.1038/sdata.2016.18>