

Establishing a generic Research Data Repository: The RADAR Service

Angelina Kraft
Technische Informationsbibliothek
(TIB) German National Library of
Science and Technology
Welfengarten 1 B
D-30167 Hannover, Germany
+49 (0)511 762 14238
angelina.kraft@tib.eu

Matthias Razum
FIZ Karlsruhe – Leibniz Institute
for Information Infrastructure
Hermann-von-Helmholtz-Platz 1
D-76344 Eggenstein-Leopoldshafen,
Germany
+49 (0)7247 808 457
matthias.razum@fiz-
karlsruhe.de

Jan Potthoff
Karlsruhe Institute of Technology
(KIT)
Hermann-von-Helmholtz-Platz 1
D-76344 Eggenstein-Leopoldshafen,
Germany
+49 (0)721 608 25 666
jan.potthoff@kit.edu

ABSTRACT

Science and its data management are in transition. And while the research data environment has become heterogeneous and the data dynamic, funding agencies and policy makers push towards findable, accessible, interoperable and reuseable (= FAIR) research data [1]. A popular issue of the management of data originating from (collaborating) research infrastructures is their dynamic nature in terms of growth, access rights and quality. On a global scale, systems for access and preservation are in place for the big data domains (e.g. environmental sciences, space, climate). However, the stewardship for disciplines of the so-called long tail of science remains uncertain. This poster gives the impression of an interdisciplinary infrastructure facilitating **research data archival and publication**.

The **RADAR - Research Data Repository** - project strives to make a decisive contribution in the field of long tail research data: On one hand it enables clients to **upload, edit, structure and describe (collaborative) data in an organizational workspace**. In such a workspace, administrators and curators can manage access and editorial rights before the data enters the preservation and optional publication level. Data consumers on the other hand may **search, access, download and get usage statistics** on the data via the RADAR portal. For data consumers, findability of research data is of utmost importance. Therefore the metadata of published datasets can be harvested via a local **RADAR API** or the DataCite Metadata Store.

Being the proverbial “**transmission belt**” between **data producers and data consumers**, RADAR specifically targets researchers, scientific institutions, libraries and publishers. In the data lifecycle, RADAR services are placed in the “Persistent Domain” of the conceptual data management model described in the “domains of responsibility”[2]. These domains of responsibility are used to show duties and responsibilities of the actors involved in research data management. Simultaneously, the domains outline the contexts of shared knowledge about data and metadata information, with the goal of a broad reuse of preserved and published research data.

RADAR applies different preservation and access strategies for open vs. closed data:

- For open datasets, RADAR provides a Digital Object Identifier (DOI) to enable researchers to clearly reference data. The service offers the publication service of research data together with format-independent data preservation for at least 25 years. Each published dataset can be enriched with discipline-specific metadata and an optional embargo period can be specified.
- For closed datasets, RADAR uses handles as identifiers and offers format-independent data preservation between 5 and 15

years, which can also be prolonged. By default, preserved data are only available to the respective data curators, which may selectively grant other researches access to preserved data.

With these two services, RADAR aims to meet demands from a broad range of research disciplines: To provide a secure, citable data storage and citability for researchers which need to retain restricted access to data on one hand, and an e-infrastructure which allows for research data to be stored, found, managed, annotated, cited, curated and published in a digital platform available 24/7 on the other.

E-research projects often require comprehensive collaborative features. These include data storage, access rights management and version control. RADAR possesses a **modular software architecture** based on the e-research infrastructure eSciDoc Next Generation. The data storage is managed by a repository software consisting of two parts: A back end addresses general tasks such as storage access, **bitstream preservation** and regular reports on data integrity, whereas the front end implements RADAR-specific workflows. Front end workflows include various data services: Metadata management, access control, data ingest processes, as well as the licensing for re-use and publishing of research data with DOI. Archival Information Packages (AIP) and Dissemination Information Packages (DIP) are provided in a **BagIt-structure**[3] in ZIP container format. As part of the import/export strategy, an API for RADAR will be provided. The API allows the import/export of data as well as metadata.

The **RADAR API enables users to integrate the archival backend into their own systems and processes**. Another option is to install the RADAR software locally. The customer may choose to only deploy the management and User Interface layer, while archiving the data in the hosted RADAR service via the API, or to run everything locally. Additionally, there is the option to run the complete software stack locally and use the hosted RADAR service as a replica storage solution.

RADAR is developed as a cooperative project of five research institutes from the fields of natural and information sciences. The technical infrastructure for RADAR is provided by the FIZ Karlsruhe – Leibniz Institute for Information Infrastructure and the Steinbuch Centre for Computing (SCC), Karlsruhe Institute of Technology (KIT). The sustainable management and publication of research data with DOI-assignment is provided by the German National Library of Science and Technology (TIB). The Ludwig-Maximilians-Universität Munich (LMU), Faculty for Chemistry and Pharmacy, and the Leibniz Institute of Plant Biochemistry (IPB) provide the scientific knowledge and specifications and ensure that RADAR services can be implemented to become part of the scientific workflow of academic institutions and universities.

KEYWORDS

Research Data Infrastructure; Data Management; Repository

ACKNOWLEDGMENTS

The repository is developed as part of a three-year project funded by the German Research Foundation from 2013 to 2016 (<http://www.radar-projekt.org>) and is placed within the program ‘Scientific Library Services and Information Systems (LIS)’ on restructuring the national information services in Germany.

REFERENCES

[1] Wilkinson, M. D. 2016 The FAIR Guiding Principles for scientific data management and stewardship. *Scientific*

Data. 3, 160018 (2016). DOI=
<http://dx.doi.org/10.1038/sdata.2016.18>

- [2] Treloar, A., Harboe-Ree, C. 2008 Data management and the curation continuum. How the Monash experience is informing repository relationships. In: *14th Victorian Association for Library Automation*, 2008, Conference and Exhibition, Melbourne, Australia. URL=
<http://arrow.monash.edu.au/hdl/1959.1/43940> (accessed on 04.04.2016)
- [3] Kunze et al. (2016) The BagIt File Packaging Format (V0.97). URL= <https://tools.ietf.org/html/draft-kunze-bagit-13> (accessed on 05.04.2016)