

ROkit

FAIR from the start

Introduction

ROkit (from Research Object) is intended to serve as a modular, expandable software framework for research data management interlinked with national and international infrastructure that can be customised to create domain-specific applications. ROkit responds to the heterogeneity of data and methods in modern research not only in built cultural heritage and is meant for applications that support researchers as early as possible in the data life cycle with FAIR data standards and practices (Wilkinson 2016).

Domain specific repositories and metadata

Established institutional and generic repositories do not yet offer sufficient domain-specific metadata, search and exploration tools and cannot already support scientists in the organization and documentation of data within the research phase. Established vocabularies such as Dublin Core, DCAT, DataCite Metadata Schema, Lightweight Information Describing Object (LIDO) offer administrative, but sparse descriptive metadata.

Domain specific repositories and specialized tools with a historical perspective on the built environment are currently being developed in Germany and Europe. Baureka.online is developing a repository that links metadata on projects, data and objects from historical building research. The DFG-3D-Viewer which can be integrated into various repositories focuses on a minimal metadata set for 3D models of historical reconstructions. ARIADNE RI integrates archaeological data from various European data providers in an open access data catalog with a modular metadata approach. The Europeana platform collects digital objects from various European cultural institutions but provides little object-specific metadata. The IDOVIR project provides a web-tool for the documentation of digital reconstructions according to the source-argument-reconstruction scheme. Within the consortia of the German "Nationale Forschungsdateninfrastruktur (NFDI)" multiple domain specific FAIR data solutions are under development, for example Semantic Kompakkt, that allows 3D mesh models to be annotated. The additional information is stored in a Wikibase instance as linked open data. The EU FP7 project "Durable Architectural Knowledge (DURAARK)" has developed metadata schemas for the digital long-term archiving of building data: Its *buildm* schema contains descriptive metadata for physical and digital objects. Its *ifcm* and *e57m* schemas supplement metadata that can be extracted from BIM models and point clouds, including georeferencing in WKT format in accordance with the Open Geospatial Consortium (OGC) specification (Lindlar and Tamke 2014).

A flexible framework

ROkit has been developed as part of the Specialized Information Service (FID) BAUdigital at the University and State Library Darmstadt (ULB) as a web-based service for the ingest, harvesting and

semantic enrichment of research data related the built environment. Currently, it serves this purpose within the NFDI4Ing (Schmitt 2020).¹

ROkit is based on a microservice infrastructure (Kubernetes cluster) that combines open-source components to create flexible workflows for ingesting, harvesting, curating and disseminating research data and metadata. The core of this infrastructure is a triple store for metadata and an S3 object store for storing files. Typical contents of a research data set can be processed individually, and metadata can be extracted from an extendable list of file formats. The platform allows the publication of research datasets in CKAN and DSPACE repositories and long-term archiving in Rosetta and Archivematica systems. In addition to delivering datasets to repositories, the platform allows the presentation of research data with advanced search and exploration tools. Currently research data in IFC, 3dm, OBJ and e57 files are parsed for metadata and converted into web-enabled formats and displayed using the integrated Model Editor (glTF) and POtree (point cloud) instances. These are supplemented by conversion and display components for various 2D formats commonly used in building research. All components can be accessed via an API² or via the OAI-PMH (import/export) and SPARQL (search) interfaces.

A special feature of ROkit is the flexible handling of metadata based on modular metadata profiles (Grönewald 2022). In addition to the aforementioned *buildm* schema, ROkit currently supports three more: one minimal set of administrative metadata for submission to repositories and long-term archiving complying with the DCAT standard, one for digitized architectural models with additional information for size, scale and collection, and one for harvested retro-digitized copies of a plan repository. In principle, ROkit allows to render input forms or masks for faceted search for any metadata set defined in form of a SHACL shape³ and validate and process any metadata in this form. This allows customized modular solutions for heterogeneous research data without the loss of an overarching search or the rigidity and complexity of a large monolithic data model. In this way, research data can be made machine actionable and reusable for a multitude of domain specific methods.

ROkit is deeply integrated into the NFDI4Ing infrastructure environment. The NFDI4Ing Terminology service provides ontologies and norm data, the Metadata Profile Service stores SHACL shapes that researchers can create with a web-based applications for their individual needs, the prototypical CSV-RDF-Mapper⁴, develop in the NFDI4Ing seedfund program with participants from the Cluster of Excellence IntCDC, allows the semantic uplifting and enrichment of tabulated data with SHACL shapes. All metadata processed by ROkit will be fed into the upcoming NFDI Federated Knowledge

¹ <https://ingest.nfdi4ing.de/>

² <https://ingest.nfdi4ing.de/api/>

³ <https://ulb-darmstadt.github.io/shacl-form/-intro>

⁴ <https://winroger.github.io/csv-rdf-mapper/>

Graph. German researchers can use ROKit with an account of their home university using the DFN AAI⁵, the upcoming IAM4NFDI identity provider supports a wide range of international institutions.

Future applications

The technological basis of ROKit is not meant for the provision of a single platform. It is designed as an extendable basis for multiple projects. In the near future it will support the IDOVIR project by harvesting source material and workflows for long term preservation and dissemination, it will provide an export function for the DFG 3D-viewer and will help to generate FAIR Digital Objects (FDO) within the proposed extension of the NFDI4Ing infrastructure.

The future development of ROKit focuses on data collection and organization within the research phase, providing FAIR workflows from the start with benefits for researchers in large workgroups and with complex data processing tasks and data sets. More components are planned to support developers modelling complex research processes and data structures. ROKit is available under an open-source license.

References

- Grönewald, M. et al. (2022). 'Mit AIMS zu einem Metadatenmanagement 4.0: FAIRe Forschungsdaten benötigen interoperable Metadaten', *E-Science-Tage 2021: Share Your Research Data*, Heuveline, V. and Bisheh, N., heiBOOKS, Heidelberg, pp. 91--104.
- Lindlar, M. and Tamke, M. (2014). 'A Domain driven Approach to Digital Curation and Preservation of 3D Architectural Data: Stakeholder Identification and Alignment in the DURAARK Project'. *Archiving Conference: Archiving 2014 Final Program and Proceedings*, The Society for Imaging Science and Technology, pp. 204–209
- Schmitt, R. H. (2020) „NFDI4Ing - the National Research Data Infrastructure for Engineering Sciences“. Zenodo. doi: 10.5281/zenodo.4015201.
- Wilkinson, M. D. et al. (2016). 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific data* 3:160018. DOI: <https://doi.org/10.1038/sdata.2016.18> .

⁵ Alternatively, login via ORCID is available for international researchers.