
The CREATIVE Project – Customizing a Generic Repository for Domain Scientists and Creating a Data Steward Network

Sibylle K. Hassler ^{1,2}, Peter Braesicke ³, Mirko Mälicke ², Jörg Meyer ⁴,
Carlos Zuleta Salmon ¹

¹ Institute of Meteorology and Climate Research – Atmospheric Trace Gases and Remote Sensing (IMKASF), Karlsruhe Institute of Technology (KIT);

² Institute for Water and Environment – Hydrology, Karlsruhe Institute of Technology (KIT);

³ German Weather Service (DWD);

⁴ Scientific Computing Center, Karlsruhe Institute of Technology (KIT)

The project CREATIVE aims at facilitating and supporting the publication of data from the scientific communities represented within the Climate and Environment Centre (CEC) at KIT in the generic institutional repository RADAR4KIT. We developed a customised template for domain-specific metadata and an online tool for metadata entries. Additionally, we facilitate establishing a network of data stewards at the CEC as a sustainable support structure for the data publication process.

Keywords: metadata, template, repository, data steward, data publication

1 Introduction

Data publication at the Karlsruhe Institute of Technology (KIT) is supported by the institutional, generic repository RADAR4KIT. While this process is well-established and supported by the research data management (RDM) team at the KIT library and Scientific Computing Center (SCC), the repository cannot fully support any domain-specific metadata description of the data by default. However, such domain-specific metadata options are desirable to better describe environmental data and support its findability in aggregating services that can use such descriptors. With the CREATIVE project we

Published in: Vincent Heuveline, Philipp Kling, Florian Heuschkel, Sophie G. Habinger, and Cora F. Krömer (Hrsg.): E-Science-Tage 2025. Research Data Management: Challenges in a Changing World. Heidelberg: heiBOOKS, 2025. DOI: <https://doi.org/10.11588/heibooks.1652.c23943> (CC BY-SA 4.0).

aim to find a solution to this gap so that data from the KIT researchers working in the earth and environment domain can be published in a well-described and FAIR (findable, accessible, interoperable and reusable) way. In the following sections we will outline the different entities that play a role in the CREATIVE project and portray our project goals, approach, status and outlook.

2 Entities and interactions within the CREATIVE project

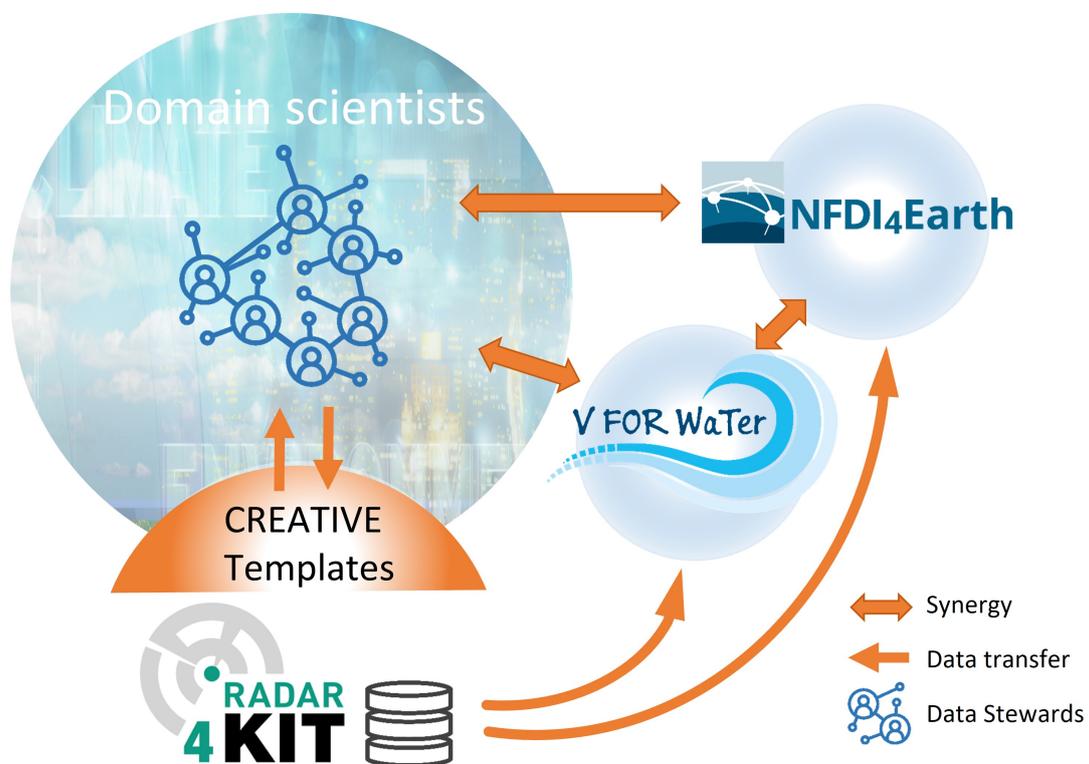


Figure 1: Overview of the CREATIVE project entities and interactions.

2.1 Domain scientists at the Climate and Environment Centre of KIT

The KIT Climate and Environment Centre (CEC)¹ is an organisational structure within KIT which encompasses institutes that conduct basic and applied research and generate knowledge on climate and environmental change. This includes concepts, observations and

¹ <https://www.klima-umwelt.kit.edu/english/index.php>; Visited on March 28, 2025.

models as well as nature-based and technological solutions for safeguarding the natural basis of life and shaping the living environment. The CEC offers a broad range of expertise in the natural, technical, and social sciences with about 700 scientists from 30 institutes.

2.2 Institutional repository RADAR4KIT

The repository RADAR4KIT² is the generic institutional repository of KIT. It is based on the RADAR³ service which is offered by FIZ Karlsruhe⁴, with data storage realised on the IT infrastructure of KIT, located at the SCC. The repository allows publication and archiving of data packages from scientific studies and projects, including a coherent description with metadata. After a curation step the data receives a DOI and is openly accessible via the RADAR4KIT portal. The metadata description of RADAR4KIT is generic in order to cater to all research disciplines at KIT. Before uploading a dataset, the metadata needs to be entered by hand or can be uploaded as an XML file. To ensure the quality and consistency of these metadata records the XML files are validated against XML Schema Definition (XSD) files that define the structure and rules of the templates. This validation process checks for structural integrity, data type consistency, and compliance with controlled vocabularies.

2.3 Virtual research environment V-FOR-WaTer

V-FOR-WaTer⁵ is a virtual research environment (VRE) that has been developed in close collaboration between hydrologists (KIT Institute for Water and Environment) and computer scientists (KIT SCC). It caters to the needs of hydrologists and environmental scientists by providing quick and easy access to hydrological data and tools. The main purpose of V-FOR-WaTer is to support quick pre-processing of data from diverse data sources and thus accelerating scientific workflows. Shared tools and a storable workspace support reproducible data analysis.

2.4 NFDI4Earth

The NFDI4Earth⁶ (National Research Data Infrastructure for Earth System Science, ESS) is a consortium of around 70 institutions within Germany, including universities, research

2 <https://radar.kit.edu/radar/en/home>; *Visited on March 28, 2025.*

3 <https://www.fiz-karlsruhe.de/en/produkte-und-dienstleistungen/radar>; *Visited on March 28, 2025.*

4 <https://www.fiz-karlsruhe.de/en>; *Visited on March 28, 2025.*

5 <https://www.vforwater.de/english/>; *Visited on March 28, 2025.*

6 <https://nfdi4earth.geo.tu-dresden.de/>; *Visited on March 28, 2025.*

organisations, infrastructure providers, governmental institutions as well as scientific associations and networks. It aims to provide FAIR, coherent, and open access to all relevant ESS data, to innovative research data management and data science methods. NFDI4Earth was constructed in a community-driven process and offers a range of services. One of them is the OneStop4All⁷, the web-based key entry point for all NFDI4Earth contents and resources. It was developed to fit the needs of a wide range of users and rests on a backbone called the KnowledgeHub where the metadata of ESS relevant resources are collected and referenced.

3 CREATIVE goals and methods

3.1 Goals of the CREATIVE project

The project CREATIVE aims at facilitating and supporting data publications in the generic repositiorium RADAR4KIT for the scientific communities within the CEC at KIT. This is achieved by creating custom templates and entry forms for domain-specific metadata that satisfy the requirements of the CEC researchers. At the same time, suitable interfaces ensure that the domain-specific (meta-)data in RADAR4KIT are accessible from the national research data infrastructure NFDI4Earth and the virtual research environment V-FOR-WaTer. With further activities, e.g. organising specific RDM workshops, a network of data stewards in the individual institutes of CEC is established, strengthening sustainable and advanced RDM at KIT.

3.2 Methods of the CREATIVE project

In a series of interviews and workshops with CEC researchers we collected and defined the metadata requirements for a comprehensive description of the relevant datasets, we co-developed the templates, collected feedback and adapted the templates accordingly. In close collaboration with the RADAR4KIT and RADAR team we established how the templates can be integrated and verified in the repository and how a future harvesting of the data would work. The template was developed on the existing metadata schema of V-FOR-WaTer (based on ISO19115 and DataCite) to ensure the processing possibility within the VRE. Further, communicating with the data model team of NFDI4Earth will also enable findability of the data in the OneStop4All as soon as the RADAR4KIT repository is harvested for data originating from the CEC researchers. As a direct integration of the template as an entry form in RADAR4KIT was not (yet) possible, we implemented an independent web form to enter the metadata and export it in different formats, e.g. XML which can be used while uploading the metadata in the RADRA4KIT repository.

⁷ <https://onestop4all.nfdi4earth.de/>; *Visited on March 28, 2025.*

A procreative exchange among CEC researchers and data stewards at the workshops supported the cultural change towards more FAIR data publications and open data. While dedicated data stewards are lacking at many CEC institutes, the exchange formats brought together researchers interested in data publications and data managers from the Helmholtz Research Field Earth and Environment DataHub. The interaction between these different parties can be seen as a starting point for a network among CEC data stewards and creates a climate of mutual help and support.

4 Transferability

While the CREATIVE metadata template was targeted to support CEC researchers in publishing their data on the local repository RADAR4KIT, the approach itself is easily transferable to other disciplines and other repositories and is meant to serve as a blueprint of the possibility to customise generic repositories. This kind of approach has also been used successfully in bioinformatics (Brandt et al. 2023).

The code behind the metadata editor web form is available open source⁸ and can be reused to cater to different constellations of metadata requirements and repositories.

5 Status and outlook

The CREATIVE project resulted in the collaborative design of domain-specific metadata templates, their incorporation into RADAR4KIT, and the development of a standalone web form application. This app enables researchers to upload datasets easily and generate XML files that comply with the required metadata structure, which are validated using the developed XSD validation schema for CREATIVE within RADAR4KIT. Once validated in RADAR4KIT, they are made discoverable with proper metadata standards, including DOI assignment. The metadata from the template was based on V-FOR-WaTer, ensuring compatibility, however, the direct connection of the two systems still needs to be established. The metadata is also compatible to the NFDI4Earth schema so it can be harvested automatically as soon as the connection to RADAR4KIT is established.

Further development steps are in progress and include, for example, the integration of existing ontologies, and the automatic completion of metadata fields from well-structured and widely used file formats at CEC such as netCDF. While the full implementation of interfaces to NFDI4Earth and V-FOR-WaTer is still ongoing, the current state is already a clear improvement for sharing and publishing datasets generated by CEC researchers.

⁸ https://github.com/KIT-HYD/creative_frontend; *Visited on March 28, 2025.*

The CEC data steward network was initialised and generates interest. Further steps in this regard will be visiting the CEC institutes and presenting the tool in hands-on sessions so that it can be thoroughly tested and promoted. In case the metadata requirements change or further institutes that were not involved in previous workshops join in, the template can also be updated. Directly supporting the researchers in using the tools will also strengthen the institutional effort strengthening data publications at KIT.

Acknowledgements

We acknowledge the support of the RADAR4KIT and RADAR teams in the development of the template, the support by the NFDI4Earth in the project and the fruitful exchange and discussions with the IMK data management team (supported by the Helmholtz Research Field Earth and Environment DataHub) and the other V-FOR-WaTer team members during the project.

Authorship Contributions

SH developed the project concept, was mainly responsible for the project proposal, initiated contacts between the project staff, CEC researchers and data managers, fostered connections to the NFDI4Earth and V-FOR-WaTer and supported the technical developments. She presented the poster at the E-Science-Tage and mainly wrote the article. PB facilitated RDM at IMK as co-chair of the Helmholtz RF E&E DataHub and as co-spokesperson of the NFDI4Earth. MM developed and implemented the web form, implemented the link between the web form and V-FOR-WaTer. JM contributed to the project concept, served as contact for V-FOR-WaTer. CZ co-developed the subject-specific metadata templates for RADAR4KIT, fostered collaboration between researchers and data stewards across the CEC institutes. PB, MM, JM and CZ all contributed to the writing of the article.

Conflict of Interest

There are no conflicts of interest.

Bibliography

Brandt, Olaf, Holger Gauza, Jan Kaltenbach, Maximilian E. Müller, Gabriel Schneider, and Claus Zinn. 2023. “Ein Werkzeug zur XSD-basierten Metadatenannotation”. In *E-Science-Tage 2023: Empower Your Research – Preserve Your Data*, edited by Vincent Heuveline, Nina Bisheh, and Philipp Kling, 276–281. heiBOOKS. <https://doi.org/10.11588/heibooks.1288.c18088>.