
Uniting the Differences With the DALIA Interchange Format (DIF) – a Minimal Metadata Specification for the DALIA OER Platform

Jonathan D. Geiger ¹, Petra C. Steiner ², Abdelmoneim Amer Desouki ²,
Henrika M. Hüppe ³, Jan-Michael Haugwitz ⁴

¹ Academy of Sciences and Literature Mainz;

² University and State Library Darmstadt, Technical University of Darmstadt;

³ Institute of Inorganic Chemistry, RWTH Aachen University;

⁴ IT Center, RWTH Aachen University

The Data Literacy Alliance (DALIA) project aims to make open educational resources (OER) on data literacy and research data management skills accessible and interoperable. A knowledge graph has been developed to store the information about the materials, using the DALIA Interchange Format (DIF) to ensure transparency and interoperability. This paper focuses on the application of the DIF as a metadata specification for teaching and learning materials of the National Research Data Infrastructure (NFDI). It describes the format and attributes of the DIF and seeks to foster ongoing collaboration with the professional community by inviting all interested parties to provide feedback and make contact.

Keywords: Open Educational Resources, Metadata Specification, Data Literacy, Research Data Management, Search Platform

1 Introduction

Teaching data literacy is a key issue for society and encompasses a broad field: it begins with creating awareness and extends to practical applications, such as programming skills.

Publiziert in: Vincent Heuveline, Philipp Kling, Florian Heuschkel, Sophie G. Habinger und 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.c23945> (CC BY-SA 4.0).

A wide range of data literacy materials and tools is already available, with some being commercial and others available as open educational resources (OER). An essential step in the development of data literacy is therefore the indexing, linking, visualizing, and making accessible of existing materials. The scientific infrastructure project DALIA takes up these tasks within the context of the National Research Data Infrastructure (NFDI), federal RDM initiatives, data competence centers, and beyond.

Since most scientific disciplines and fields in Germany deal with the issue of adequate training of data literacy for their students and researchers, it is reasonable to exploit synergies, share experiences, technically link existing materials, and identify gaps in both general and subject-specific offerings. The Section Training & Education (EduTrain)¹ of the NFDI is committed to overseeing this task; the DALIA project is responsible for the technical implementation, anchoring, and establishment within the NFDI.

A knowledge graph based on the MoDALIA Ontology is used to store information about the educational materials. It is based on semantic web technologies RDF, SPARQL, and OWL, allowing it to be linked to the Linked Open Data (LOD) cloud. To facilitate the ingestion process and serve as a means of communication with the community, the DALIA Interchange Format (DIF) was developed. It is compatible with the ontology.

This short paper provides an overview of the background of the DALIA project and focuses on the DIF as a metadata specification for teaching and learning materials. It also aims to foster ongoing communication with the professional community by inviting all interested parties to provide feedback and make contact.

Section 2 offers an overview of the DALIA project and platform. Section 3 describes the essentials of the DIF in greater detail. Section 4 will discuss future developments and enhancements planned for the near future.

2 DALIA: Project, Platform, and Knowledge Graph

The DALIA project started at the end of 2022. It combines a broad cross-disciplinary competence spectrum of teaching, research and infrastructure, supported by the Technical University of Darmstadt (Chair of Fluid System Technology, and University and State Library Darmstadt), the RWTH Aachen University (Chair of Bioinorganic Chemistry and IT Center), the Academy of Sciences and Literature Mainz (Digital Academy), the University Medical Center Göttingen (Department of Medical Informatics), and the German National Library of Science and Technology Hannover (TIB). Collaborating together, the project partners and the NFDI e.V. and its sections EduTrain, Ethical, Legal & Social As-

¹ <https://www.nfdi.de/section-edutrain/>; *Visited on March 24, 2025.*

pects², and (Meta)data, Terminologies, Provenance³ aim to consistently provide teaching and learning materials. The DALIA project promotes the learning and teaching of data literacy by providing a curated collection of OER. This empowers everyone to collect, analyze, critically evaluate, and use data in a FAIR way. Hence, the DALIA platform aims to be a single point of entry for all different areas of science, career levels, and levels of research data management expertise – from those beginning with RDM to data stewards – providing access to a collection of OER focused on data literacy (Figure 1).

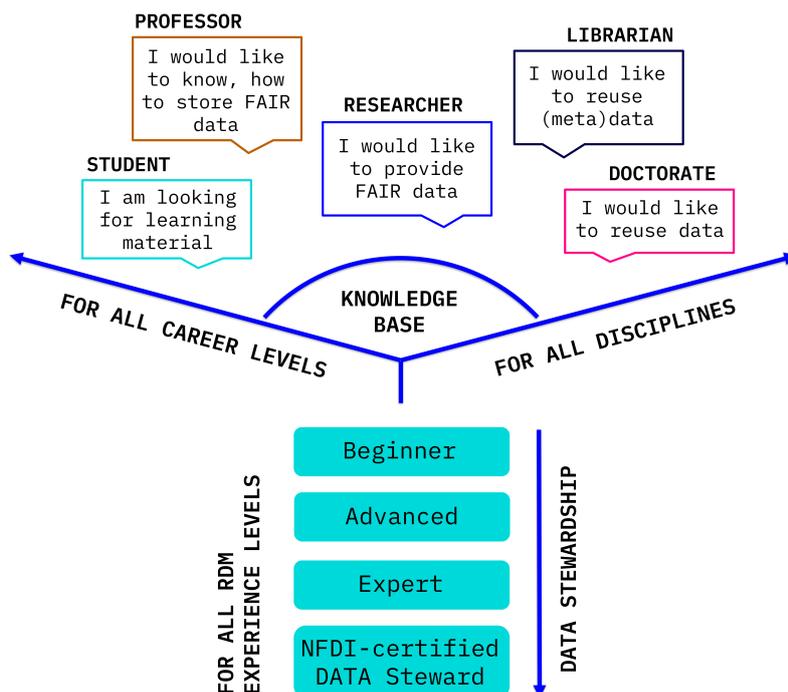


Figure 1: Visualization of the DALIA concept: Teaching and training materials for all career levels, research data management experience levels, and disciplines in one knowledge base.

To link and integrate the diverse needs of all career and competence levels, and academic disciplines, DALIA is developing a knowledge graph that establishes interoperability between existing resources and ensures findability, quality assurance, and adherence to the FAIR Principles (Garcia et al. 2020).

To ensure the interoperability of teaching and learning materials for the DALIA knowledge graph, a schematization of their metadata is required. This is done by the DIF, which provides a framework to make the metadata of teaching and learning materials transparent, comparable, and smoothly integrable into the DALIA platform. It describes and explains the data fields for the online publication of educational resources. Version 1.3 of the DIF was published on Zenodo in June 2024 (Geiger et al. 2024).

² <https://www.nfdi.de/section-elsa/>; Visited on March 24, 2025.

³ <https://www.nfdi.de/section-metadata/>; Visited on March 24, 2025.

3 The DALIA Interchange Format (DIF)

The DIF was developed in close collaboration with the scientific community. The development process included several feedback rounds, leading to a clear structure, and transparent and user-oriented definitions. This process also allowed consideration of existing data standards as well as the requirements of the professional community. The selection of the description elements is based on the Dublin Core Application Profile (DCAP, see Coyle 2022), a set of elements in tabular form that can be used for the application and validation of metadata. Like the DCTAP, the DIF can be converted to RDF format. The description elements are as follows:

- The attribute name, which is a descriptive label
- The priority level *mandatory*, *recommended*, or *optional*
- A description of the attribute, including a definition and if required instructive information
- The values to be used, consisting of the constraint type and constraint
- Examples, for elucidating the definition
- The cardinality, for indicating if the attribute is repeatable or not
- The data type, which provides information on the format of the attribute
- The propertyID: a standardized identifier
- Optionally a comment, for example, more instructions on exceptions

The PropertyIDs that specify the attributes are particularly important: Common standards such as the DCMI Metadata Terms (DCMI Usage Board 2020) and Schema.org are used here. As an application profile, the DIF comprises elements compiled from other namespaces. However, when standardized definitions of the required attributes are unavailable, custom IDs are defined in the underlying MoDALIA ontology (mo)⁴ to meet formal requirements such as range and domain. The priority levels were carefully selected. In practice, often some information about the resources is missing. However, this should not impede the process of ingestion to the platform. To address this, only four DIF attributes are mandatory (*Authors*, *License*, *Link*, *Title*). Eleven are strongly recommended (*Community*, *Description*, *Discipline*, *FileFormat*, *Keywords*, *Language*, *LearningResourceType*, *MediaType*, *ProficiencyLevel*, *PublicationDate*, *TargetGroup*), and three are optional (*RelatedWork*, *Size*, *Version*). Table 1 provides an example of the metadata for Biernacka et al. (2024).

⁴ <https://purl.org/ontology/modalia>; Visited on March 24, 2025.

Table 1: Example for the metadata of a resource according to the DIF. *Authors*, *Description*, and *LearningResourceType* have been abbreviated for conciseness.

Attribute Name	Example
Authors	Biernacka, Katarzyna :https://orcid.org/0000-0003-4864-8036 * Castillo Upiachihua, Claudia Beatriz :https://orcid.org/0009-0009-2186-3212 * ...
License	CC-BY-4.0
Link	https://doi.org/10.5281/zenodo.13927614
Title	Train-the-Trainer Concept on Research Data Management
Community	DINI/nestor UAG Schulungen/Fortbildungen (SR)
Description	This version of the 'Train-the-Trainer Concept on Research Data Management' is the English translation of the fifth updated version of the German 'Train-the-Trainer-Konzept zum Forschungsdatenmanagement' ...
Discipline	https://w3id.org/kim/hochschulfaechersystematik/n0
FileFormat	.pdf * .zip * .txt * .docx * .rtb * .pptx * .xlsx
Keywords	research data * research data management * RDM * train-the-trainer * OER * training * Forschungsdaten * Forschungsdatenmanagement * Didactics * ...
Language	en
Learning Resource Type	Book * https://w3id.org/kim/hcrt/course * https://w3id.org/kim/hcrt/lesson_plan * https://w3id.org/kim/hcrt/worksheet * ...
MediaType	text * presentation
ProficiencyLevel	novice * advanced beginner * competent
PublicationDate	2024-10-31
Target Group	data steward * teacher (higher education)
Related Work	isNewerVersionOf: https://doi.org/10.5281/zenodo.10122153 * isNewerVersionOf: https://doi.org/10.5281/zenodo.4071471
Size	195.2
Version	5.1

Further restrictions apply to the cardinalities. For instance, each OER should possess exactly one license and one title. The latter is for the sake of simplicity, as the DIF is a minimal standard. The attributes are explained in the description, and the technical constraints are made explicit. Controlled vocabularies are specified by picklists. Examples round off the DIF. Geiger et al. (2025) provide an example of the same OER showing the cardinalities. The DIF is partly compatible with the General Metadata Profile for Educational Resources (Pohl et al. 2024) and the controlled vocabulary used therein. It also considers the LRMI (Barker and Sutton 2022), and the RDA ETHRD-IG minimum metadata set recommendations (Hoebelheinrich et al. 2022). Additionally, a crosswalk to Schema.org attributes on learning resources is provided. To facilitate the contribu-

tion of metadata for OER to the DALIA platform, we developed a .csv template. This template was published alongside the DIF version 1.3 on Zenodo (Geiger et al. 2024). It includes a column for each attribute, allowing contributors to submit metadata for one or multiple OERs in a single document. This document is interoperable with the backend implementation of the DALIA platform.

4 Further Developments

Detailed feedback and inquiries from the community have shown the necessity and importance of such a metadata specification. Particularly during the preparation of metadata collections following the DIF structure, numerous questions arose, underscoring the potential for further improvements.

Currently, we are conducting a detailed analysis of the feedback and questions received. This process has led to the preparation of a revised version of the DIF (version 1.4). On the one hand, there was further feedback from research data management projects with their respective needs and perspectives. While these are largely addressed by the DIF, they also introduce new requirements in certain areas. On the other hand, the practical application of the DIF by the projects has revealed ambiguities, which can be resolved through more precise definitions, adapted examples, and, in particular, more detailed and extended picklists, for example in the area of scientific disciplines or domain-specific keywords. Compatibility with the DIF version 1.3 is important here. Similarly, future versions will also be backward compatible. This iterative development approach supports us to provide solutions that fit the community best and achieve widespread acceptance. In the near future, metadata for OER will no longer need to be submitted via .csv file in a manual process. Instead, DALIA will provide a curation form to registered contributors. This will enhance the useability through features like autocomplete and API-based data import from platforms such as Zenodo. Consequently, metadata can be added in accordance with the DIF and the MoDALIA Ontology with less effort and reduced error potential, thereby improving the quality of the metadata. As registered users will be able to edit metadata through a reviewed process, DALIA will support the RDM and data literacy communities with functionalities for enhancing both the quantity and quality of metadata, advancing the goal of FAIR OER.

Acknowledgements

This project with the funding code 16DWWQP07A was funded by the Federal Ministry of Education and Research (BMBF) and the funding measure from the EU's Capacity Building and Resilience Facility.

Authorship Contributions

All authors were writing, reviewing, and editing.

Conflict of Interest

There are no conflicts of interest.

Bibliography

- Barker, Phil, and Stuart Sutton, editors. 2022. “LRMI Terms (RDF) 2022-06-14”. Visited on March 27, 2025. https://www.dublincore.org/specifications/lrmi/lrmi_terms/2022-06-14/.
- Biernacka, Katarzyna, Claudia Beatriz Castillo Upiachihua, Ron Dockhorn, Claudia Engelhardt, Laura Grunwald-Eckhardt, Kerstin Helbig, Juliane Jacob, et al. 2024. “Train-the-Trainer Concept on Research Data Management”. <https://doi.org/10.5281/zenodo.13927614>.
- Coyle, Karen, editor. 2022. “Elements for DC Tabular Application Profiles”. With contributions from Tom Baker, Phil Barker, John Huck, and Nishad Thalath. Visited on March 27, 2025. <https://www.dublincore.org/specifications/dctap/elements/>.
- DCMI Usage Board. 2020. “DCMI Metadata Terms”. Visited on March 27, 2025. <https://www.dublincore.org/specifications/dublin-core/dcmi-terms/>.
- Garcia, Leyla, Bérénice Batut, Melissa L. Burke, Mateusz Kuzak, Fotis Psomopoulos, Ricardo Arcila, Teresa K. Attwood, et al. 2020. “Ten simple rules for making training materials FAIR”. Edited by Scott Markel. *PLOS Computational Biology* 16 (5): e1007854. <https://doi.org/10.1371/journal.pcbi.1007854>.
- Geiger, Jonathan D., Petra Steiner, Abdelmoneim Amer Desouki, Henrika Maria Hüppe, Frank Lange, Sonja Herres-Pawlis, Alessandra Kuntz, Ulrich Sax, and Jochen Ortmeier. 2025. *Uniting the Differences with The DALIA Interchange Format (DIF) – a Minimal Metadata Specification for the DALIA OER Platform*. Zenodo. Poster at E-Science-Tage, Heidelberg, Germany. <https://doi.org/10.5281/zenodo.15020137>.
- Geiger, Jonathan D., Petra Steiner, Abdelmoneim Amer Desouki, and Frank Lange. 2024. *DALIA Interchange Format (Version 1.3)*. Zenodo. <https://doi.org/10.5281/ZENODO.11521029>.

Hoebelheinrich, Nancy J., Katarzyna Biernacka, Michelle Brazas, Leyla Jael Castro, Nicola Fiore, Margareta Hellström, Emma Lazzeri, et al. 2022. *Recommendations for a minimal metadata set to aid harmonised discovery of learning resources*. Zenodo. <https://doi.org/10.15497/RDA00073>.

Pohl, Adrian, Axel Klinger, Boris Hartmann, Carl Schuurbiers, Manuel Kummerländer, Manuel Oellers, Mikey Stengel, et al. 2024. "Allgemeines Metadatenprofil für Bildungsressourcen (AMB)". Entwurf vom 21. November 2024. Visited on March 27, 2025. <https://dini-ag-kim.github.io/amb/draft/>.