1st Conference on Research Data Infrastructure

Poster presentations II

https://doi.org/10.52825/CoRDI.v1i.373

Published: 07 Sept. 2023

DALIA FAIR Open Educational Federation

Aggregation, Harmonisation, Curation, and Quality Assurance with DALIA

Canan Hastik[1][https://orcid.org/0000-0003-1729-4642], Gábor Kismihók [2][https://orcid.org/0000-0003-3758-5455], Frank Lange[3][https://orcid.org/0000-0002-9346-6031], and Petra C. Steiner [1][https://orcid.org/0000-0001-8997-2620]

[1] Technical University of Darmstadt, Germany

[2] TIB - Leibniz Information Centre for Science and Technology, Germany

[3] RWTH Aachen University, Germany

Abstract. As a funding measure of the German Federal Ministry of Education and Research and from the EU's Development and Resilience Facility, the project “DALIA - Knowledge-Base for FAIR data usage and supply" not only creates technical bridges between NFDI consortia but also an evolving approach that contributes to transdisciplinary community building in the long term. By harmonising, networking and making visible teaching and learning materials for the development and promotion of data literacy, the project makes an important contribution to education, training and cultural change.

Keywords: Training and Education, Knowledge Graph, Research Data Management, Harmonising RDM

1. Project Introduction and Objective

The DALIA project will develop a central entry point for a federated knowledge base, comprising teaching and learning materials to build and strengthen data literacy and knowledge in related FAIR [5] data use (data science) and supply (FAIR research data management). Thus, based on the DALIA Knowledge Graph, which is currently under development, the DALIA Knowledge-Base will make the heterogeneous, distributed and subject-specific teaching and learning materials of the NFDI e.V. [4], the Research Data Management (RDM) and Open Educational Resource (OER) communities visible, accessible, and networked in a sustainable way. By implementing the DALIA Knowledge-Base as Linked Open Data (LOD), not only the semantic interoperability of the learning content will be established but also the curation of teaching and learning materials. The DALIA Knowledge-Base will ensure optimal findability and best possible access for users from a wide range of disciplines, career and competence levels. It will also foster the (re)usability of the educational materials by means of a monitoring and personal learning recommendation system.

2. Challenges

As a federated infrastructure for FAIR data use and supply, DALIA addresses content providers and curators for educational purposes with learning content repositories such as NFDI4Ing’s Education Repository (https://git.rwth-aachen.de/nfdi4ing/education), NFDI4Chem’s Knowledge Base (https://knowledgebase.nfdi4chem.de/knowledge_base/),

DARIAH-Campus (https://campus.dariah.eu/), curated group repositories on Zenodo, e.g. Research data management (RDM) open training materials [2] or future data competence centers using their own community-specific taxonomies, for instance the Taxonomy of Digital Research Activities in the Humanities (TaDiRAH, https://tadirah.info). Similarly, to the European Open Science Cloud (EOSC, https://eosc-portal.eu/), DALIA wants to generate added value for users by networking these heterogeneous and distributed resources.

Challenges regarding the use of standards for material indexing, or the use of vocabularies for curation and search, are systematically tackled and harmonised across disciplines. In this context, characteristic questions are: How do we use vocabulary to harmonise resources for discovery and recommendation? What is the role of controlled vocabularies? Are keywords related to a discipline or resource? How is curation best organized? What is the common level of control?

The user-oriented findability, curation and quality assessment of teaching and learning materials in DALIA are the main cornerstones of the federated infrastructure. DALIA reuses existing open technologies, pipelines (e.g. data aggregator workflow in Europeana [1]), and assessment criteria (e.g. RDM learning goal matrix [3]) from FDM initiatives, which are working closely with NFDI e.V. and as well as communities from different RDM areas. These and the associated FAIR compliance are considered as important prerequisites for a trustworthy DALIA methodology, leading to the provision of a generic, but at the same time discipline-oriented educational service. In DALIA, the outcome of the harmonisation of metadata, including controlled vocabularies, will be represented in the DALIA Knowledge-Graph. A particular focus is to identify the scope and level of detail of controlled vocabularies and other metadata standards to be used to describe teaching and learning materials. Projects such as TaDiRAH and standards developed in focus group workshops and working groups such as the NFDI sections (Meta-)Data, Terminologies, Provenance, and Training & Education will be taken into account. This will yield networks of concepts which interlink to concepts from other resources. Controlled vocabularies, taxonomies and other ontologies, even if imperfect, are useful due to their semantic relations and expressiveness compared to other kinds of knowledge representation. They can be extended through language models and point to relevant similar concepts from other resources.

Interlinking learning content through DALIA is organized by content providers, who are responsible for the registration, quality assurance, and maintenance of learning content. Content curation is based on a small obligatory (closed) set of terms such as title, author, license, competency level, learning goal, topic/subject, link of the resource, and additional (open) keyword field to determine the precise context of a given learning resource.

3. Solution and Outreach

Thus, a common framework for the visibility of teaching and learning materials, their quality assessment and (re)usability is created. The aggregated content is curated in an RDF triplestore and is provided as Linked Open Data via a public SPARQL endpoint. In DALIA not only an API for an aggregated federated vocabulary service will be developed, but also a predefined set of common vocabularies for external repositories will be defined and a vocabulary charter with guidelines for DALIA users will be developed. This is intended to convey best practices for different curators, different metadata, and different vocabularies, and to encourage multipliers to assess quality.

Establishing interoperability is not a purely technical challenge. The harmonisation of the DALIA technology stack goes hand in hand with the development of framework conditions and didactic concepts for the quality-assured curation of practically usable teaching and learning materials for self-learning. To ensure the usability of DALIA in the long term, we will not only develop a train-the-trainer concept for learning content curation, but we will also collect user feedback in focus group workshops and via user monitoring continuously. The Data Literacy

Alliance behind DALIA ensures that the learning and teaching materials in the NFDI e.V. are continuously offered via the DALIA Knowledge-Base, and that they are further developed in terms of methodology and technology in line with requirements.

Data availability statement

This submission is not based on data.

Author contributions

CH - conceptualisation, writing original draft, review, and editing

GK – reviewing, and editing

FL – reviewing, and editing

PS - writing, reviewing, and editing

Competing interests

The authors declare that they have no competing interests.

Funding

This project with the federal label 16DWWQP07A is funded by the German Federal Ministry of Education and Research (BMBF) and by the EU's Reconstruction and Resilience Facility.

Acknowledgement

We would like to thank our DALIA project partners and related NFDI consortia for their valuable input.

References

1. Nuno Freire, Enno Meijers, Sjors de Valk, Julien A. Raemy, and Antoine Isaac, “Metadata Aggregation via Linked Data: Results of the Europeana Common Culture Project,” in *Metadata and Semantic Research. 14th International Conference, MTSR 2020, Madrid, Spain, December 2–4, 2020, Revised Selected Papers* (Communications in Computer and Information Science 1355), Emmanouel Garoufallou, María-Antonia Ovalle-Perandones, Eds., Cham, Germany: Springer, 2021, pp. 383–394, doi: https://doi.org/10.1007/978-3-030-71903-6_35.
2. Laura Molloy. “Research data management (RDM) open training materials.” https://zenodo.org/communities/dcc-rdm-training-materials/?page=1&size=20 (26.04.2023).
3. Britta Petersen *et al.*, “Lernzielmatrix zum Themenbereich Forschungsdatenmanagement (FDM) für die Zielgruppen Studierende, PhDs und Data Stewards (Version 2),” 2023, Zenodo. doi: https://doi.org/10.5281/zenodo.8010617 (29.08.2023).
4. York Sure-Vetter, Eva Lübke, Sophie Kraft, Hendrik Seitz-Moskaliuk, “Nationale Forschungsdateninfrastruktur (NFDI) e. V. - Satzungsvorstellung,” 2021, doi: https://doi.org/10.5281/ZENODO.5735196.
5. Mark D. Wilkinson *et al*., “The FAIR Guiding Principles for scientific data management and stewardship,” in *Scientific Data*, vol. 3, Mar. 2016, Art no. 160018, doi: https://doi.org/10.1038/sdata.2016.18.