

Das DALIA Interchange Format – eine Metadaten-Spezifikation für die OER-Plattform DALIA

Geiger, Jonathan

jonathan.geiger@adwmainz.de
Akademie der Wissenschaften und der Literatur | Mainz,
Deutschland
ORCID: 0000-0002-0452-7075

Desouki, Abdelmoneim

abdelmoneim.desouki@tu-darmstadt.de
Technische Universität Darmstadt, Deutschland
ORCID: 0000-0003-2083-1277

Lange, Frank

f.lange@itc.rwth-aachen.de
RWTH Aachen, Deutschland
ORCID: 0000-0002-9346-6031

Steiner, Petra

petra.steiner@tu-darmstadt.de
Technische Universität Darmstadt, Deutschland
ORCID: 0000-0001-8997-2620

Hüppe, Henrika

henrika.hueppe@ac.rwth-aachen.de
RWTH Aachen, Deutschland
ORCID: 0000-0002-0570-7924

Die Vermittlung von Datenkompetenz (Data Literacy) ist seit mindestens einem Jahrzehnt ein für die Gesellschaft zentrales Thema (vgl. Schüller, Busch und Hindinger, 2019; Ridsdale, Rothwell und Smit, 2015). Datenkompetenz ist ein weites Feld: Sie beginnt bei der Schaffung von Awareness und reicht bis zur konkreten Anwendbarkeit, beispielsweise in Form von Programmierfähigkeiten. Mannigfaltiges Material zur Vermittlung von Datenkompetenz liegt bereits vor – beispielsweise Videomitschnitte von Workshops, Guidelines, Podcasts oder Jupyter Notebooks –, teilweise kommerziell, teilweise als offene Bildungsressource (Open Educational Resource, kurz OER). Ein wichtiger Schritt in der planvollen Entwicklung von Data Literacy liegt daher in der Erschließung, Verlinkung, Sichtbar- und Zugänglichmachung des bereits vorhandenen Materials.

Das wissenschaftliche Infrastrukturprojekt DALIA (“Data Literacy Alliance”) greift diese Überlegungen sys-

tematisch und vor dem Hintergrund der sich im Aufbau befindlichen Nationalen Forschungsdateninfrastruktur (NFDI) auf.

Da sich die meisten wissenschaftlichen Disziplinen und Felder in Deutschland mit der Frage nach adäquater Ausbildung von Datenkompetenz für ihre Studierenden und Wissenschaftler*innen beschäftigen, liegt es nahe, Synergieeffekte zu nutzen, Erfahrungen zu teilen, das vorhandene Material technisch zu verknüpfen und in diesem Zusammenhang auch Lücken im allgemeinen und fachspezifischen Angebot aufzeigen zu können. Organisatorisch hat sich dieser Aufgabe im Rahmen der NFDI die Sektion Training & Education (EduTrain) verschrieben; die technische Umsetzung, Verankerung und Etablierung in der NFDI obliegt dem Projekt DALIA.

Das Projekt DALIA begann Ende 2022. Es verbindet ein disziplinübergreifendes Kompetenzspektrum aus Lehre, Forschung und Infrastruktur, getragen von der Technischen Universität Darmstadt (Institut für Fluidsystemtechnik und Universitäts- und Landesbibliothek), der RWTH Aachen (Lehrstuhl für Bioanorganische Chemie und IT Center), der Akademie der Wissenschaften und der Literatur Mainz (Digitale Akademie), der Universitätsmedizin Göttingen (Institut für Medizinische Informatik) und der Technischen Informationsbibliothek Hannover. Durch die enge Kooperation der Projektpartner mit dem NFDI e.V. und seinen Sektionen EduTrain (<https://www.nfdi.de/section-edutrain/>), Ethical, Legal & Social Aspects (<https://www.nfdi.de/section-elsa/>) und (Meta)daten, Terminologien, Provenienz (<https://www.nfdi.de/section-metadata/>) wird eine dauerhafte Bereitstellung der Lehr- und Lernmaterialien angestrebt. Durch Workshops, Personas und User Stories soll die Usability der Plattform für die Zielcommunities gewährleistet werden.

Um die unterschiedlichen Bedarfe aller Karriere- und Kompetenzstufen und Disziplinen zu vernetzen und zu integrieren, wird in DALIA ein Wissensgraph entwickelt, der die Interoperabilität zwischen bestehenden Ressourcen herstellt und die Auffindbarkeit, Qualitätssicherung und Einhaltung der FAIR-Principles der OER-Metadaten gewährleistet (vgl. Garcia et al., 2020). Eine Prüfung formaler Art und die Bereitstellung von Feedback-Mechanismen auf der Plattform durch Rating- und Kommentierungs-Optionen gewährleistet die Qualität der OER und ihrer Metadaten.

Zur Gewährleistung der Interoperabilität von Lehr- und Lernmaterialien für den DALIA-Wissensgraphen ist eine Schematisierung ihrer Metadaten erforderlich. Diese Aufgabe übernimmt das DALIA Interchange Format (DIF), welches einen Rahmen bietet, um die Metadaten von Lehr- und Lernmaterialien transparent, vergleichbar und nahtlos in die DALIA-Plattform integrierbar zu machen. Es umfasst die Beschreibung und Erläuterung der Datenfelder für die Online-Veröffentlichung von Bildungsressourcen und wurde im Juni 2024 als V1.3 auf Zenodo publiziert (Geiger, Steiner, Desouki, Lange, 2024).

Das DIF wurde in enger Zusammenarbeit mit der Fachcommunity entwickelt. Dieser Entwicklungsprozess um-

fasste mehrere Feedbackrunden, welche zu transparenten und benutzerorientierten Definitionen der Datenfelder sowie einer klaren Struktur führten. Zudem ermöglichte dieser Prozess die Berücksichtigung bereits existierender Datenstandards und die Bedürfnisse der Fachcommunity. Das DIF ist mit dem Allgemeinen Metadatenprofil für Bildungsressourcen (Pohl et al. 2024) und den dort verwendeten kontrollierten Vokabularen zum Teil kompatibel und berücksichtigt das LRMI (vgl. Barker und Sutton, 2022), den Standard Learning Object Metadata (LOM) (IEEE Computer Society, 2020) und die Empfehlungen zum minimalen Metadatensatz der RDA ETHRD-IG (vgl. Hoebelheinrich et al., 2022). Der Weg einer neuen Spezifikation wurde gewählt, da kein bereits vorhandener Standard die folgenden Anforderungen erfüllen konnte: 1. die spezifischen Anforderungen der Fachcommunities, wie zum Beispiel die Berücksichtigung von OER in der Schule, 2. den technischen Anforderungen des Wissensgraphen, zum Beispiel die Validierung der Metadatenschemata, und 3. dem Erkenntnisinteresse des Projektes: vorgesehen sind verschiedene Analysen des verfügbaren Materials, um Aussagen über den Status quo der Data-Literacy-OER treffen zu können.

Die Auswahl der deskriptiven Elemente orientiert sich am Dublin Core Application Profile (DCTAP) (vgl. Coyle, 2022), eine Menge von Elementen in tabellarischer Form, die für die Anwendung und Validierung von Metadaten verwendet werden kann. Ebenso wie das DCTAP kann das DIF in das RDF-Modell konvertiert werden. Die deskriptiven Elemente umfassen u. a. die Kardinalität, den Datentyp und Angaben zu Constraints. Besonders wichtig sind die Angaben von PropertyIDs, die die Attribute spezifizieren. Dabei werden gängige Standards wie die DCMI Metadata Terms (vgl. DCMI Usage Board, 2020) und Schema.org benutzt. Als Applikationsprofil besteht das DIF aus Elementen, die aus anderen Namensräumen kompiliert werden (vgl. Steiner, Hastik, Fuhrmans, 2024). Falls es jedoch keine standardisierten Definitionen gibt, werden eigene IDs in der zugrundeliegenden MoDALIA ontology (mo) (<https://purl.org/ontology/modalia>) definiert. Ein Beispiel hierfür ist mo:Community für den Begriff der fachlichen Community, wie er bislang nicht als Standard definiert war.

Von den 18 Attributen des DIF V1.3 sind vier obligatorisch ("Authors", "License", "Link", "Title"), elf empfohlen ("Community", "Description", "Discipline", "File-Format", "Keywords", "Language", "Learning Resource Type", "MediaType", "Proficiency Level", "Publication Date", "Target Group") und drei optional ("Related Work", "Size", "Version"). Die Attribute werden erläutert und die technischen Constraints expliziert. Für Auswahl Listen werden kontrollierte Vokabulare angegeben. Beispiele runden das DIF ab.

Das Poster beleuchtet die Hintergründe des DALIA-Projekts und stellt das DALIA Interchange Format als Metadaten-Spezifikation für Lehr- und Lernmaterialien in den Mittelpunkt. Die Kommunikation mit der Fachcommunity soll durch das Poster fortgeführt werden, indem allen Interessierten die Einladung zu Feedback und Kontaktaufnahme ausgesprochen wird. Dafür werden auf dem Poster die Hin-

tergründe und Ziele des Projekts DALIA zusammengefasst, das DALIA Interchange Format erläutert, sowie Diskussionen der kritischen Attributspezifikationen dargestellt.

Bibliographie

Barker, Phil und Stuart Sutton (Hg). 2022. "LRMI Terms (RDF) 2022-06-14." https://www.dublincore.org/specifications/lrmi/lrmi_terms/2022-06-14/ . (letzter Zugriff am 02.12.2024)

Coyle, Karen (Hg). 2022. "Elements for DC Tabular Application Profiles." Unter Mitarbeit von Tom Baker, Phil Barker, John Huck und Nishad Thalath. <https://www.dublincore.org/specifications/dctap/elements/> . (letzter Zugriff am 02.12.2024)

DCMI Usage Board . 2020. "DCMI Metadata Terms." <https://www.dublincore.org/specifications/dublincore/dcmi-terms/> . (letzter Zugriff am 02.12.2024)

Garcia, Leyla, Bérénice Batut, Melissa L. Burke, Mateusz Kuzak, Fotis Psomopoulos, Ricardo Arcila, Teresa K. Attwood, Niall Beard, Denise Carvalho-Silva, Alexandros C. Dimopoulos, Victoria Dominguez del Angel, Michel Dumontier, Kim T. Gurwitz, Roland Krause, Peter McQuilton, Loredana Le Pera, Sarah L. Morgan, Päivi Rauste, Allegra Via, Pascal Kahlem, Gabriella Rustici, Celia W. G. van Gelder und Patricia M. Palagi . 2020. "Ten simple rules for making training materials FAIR." PLOS Computational Biology . DOI: 10.1371/journal.pcbi.1007854.

Geiger, Jonathan, Petra Steiner, Abdelmoneim Amer Desouki und Frank Lange. 2024. "DALIA Interchange Format (Version 1.3)." Zenodo. DOI: 10.5281/zenodo.11521029.

Hoebelheinrich, Nancy J., Katarzyna Biernacka, Michelle Brazas, Leyla Jael Castro, Nicola Fiore, Margareta Hellström, Emma Lazzeri, Ellen Leenarts, Paula Maria Martinez Lavanchy, Elizabeth Newbold, Amy Nurnberger, Esther Plomp, Lucia Vaira, Celia W.G. van Gelder und Angus Whyte. 2022. "Recommendations for a minimal metadata set to aid harmonised discovery of learning resources." Zenodo. DOI: 10.15497/RDA00073.

IEEE Computer Society. 2020. "IEEE Standard for Learning Object Metadata IEEE." Std 1484.12.1™-2020. DOI: 10.1109/IEEESTD.2020.9262118.

Pohl, Adrian, Axel Klinger, Boris Hartmann, Carl Schuurbijs, Manuel Kummerländer, Manuel Oellers, Mikey Stengel, Mirjan Hoffmann, Steffen Rörtgen, Stephan Kulla und Tobias Bülte . 2024. "Allgemeines Metadatenprofil für Bildungsressourcen (AMB). Entwurf vom 21. November 2024." <https://dini-ag-kim.github.io/amb/draft/> . (letzter Zugriff am 27.11.2024).

Ridsdale, Chantel, James Rothwell, Mike Smit, Michael Bliemel, Dean Irvine, Daniel Kelley, Stan Matwin, Brad Wuetherick und Hossam Ali-Hassan. 2015. "Strategies and Best Practices for Data Literacy

Education: Knowledge Synthesis Report.” DOI: 10.13140/RG.2.1.1922.5044.

Schüller, Katharina, Paulina Busch und Carina Hindinger . 2019. “Future Skills: Ein Framework für Data Literacy.” Hochschulforum Digitalisierung 47. DOI: 10.5281/zenodo.3349865.

Steiner, Petra, Canan Hastik und Marc Fuhrmans . 2024. “Compiling Controlled Vocabularies of Contributor and User Roles for a Platform of Open Educational Resources.” DHd 2024 Quo Vadis DH (DHd2024), Passau, Deutschland . 2024. Zenodo. DOI: 10.5281/zenodo.10709934.