
The DARIAH-DE Data Federation Architecture as a highway to FAIR data

Melina Jander*¹ and Lukas Weimer*¹

¹Göttingen State and University Library – Germany

Abstract of the answer to the call for proposals

Introduction

The ever-evolving digitisation of research across all disciplines yearns for an equally evolving offer of corresponding services. Diverse and large collections of research data become useful for researchers especially when they are FAIR. Comprehensive, user-friendly, and accurate search interfaces are important elements within this context. The DARIAH-DE Data Federation Architecture (DFA) responds to these requirements with its services. As a well-established integral part of DARIAH-DE, the DFA is now also intertwined with the CLARIN search spaces; most notably the Federated Content Search has become an endpoint for the DARIAH-DE Collection Registry. This integration work (Buddenbohm 2020) is pursued by the BMBF-funded project CLARIAH-DE and aims at research institutions and individual researchers alike.

The DARIAH-DE Data Federation Architecture (DFA)

The FAIR principles have become the major guidelines within the realm of digital research. While they are usually applied to research outcomes-most prominently research data like digitised collections or research findings-the path which leads to those outcomes is significant, too. On this path, researchers are often confronted with the question of how to make their data FAIR without having to invest a lot of money, time, and technical expertise?

Fig. 1. Schematic structure of the DARIAH-DE DFA.

An answer to this question is given by the DARIAH-DE DFA (Fig. 1) through its modular design. It is most adequately described as a combination of services and tools connected through multiple interfaces. Research data and collection descriptions can be accessed by various interest groups, e.g., libraries, research facilities, cultural institutions, data centers, or individual researchers. The graphical user interface and language options invite researchers with diverse backgrounds to work with the DFA, and offer further benefits: The data is indexed with a persistent identifier, user support is guaranteed through a helpdesk (CLARIAH-DE 2021), and all services are free of charge. The DFA is composed of five core services and tools which can be combined as well as individually utilized. Each tool has

*Speaker

its own interface, represented in Fig. 1 through the arrows connecting the tools with the researcher and vice versa.

Upload and describe data: The Publikator enables uploading data and describing them with metadata following the Dublin Core metadata standard; the GUI allows for an intuitive and time-saving work-flow (Cremer 2018).

Publish data: With its upload into the Repository (DARIAH-DE 2020), the data is indexed with a Persistent Identifier (DataCite DOI and Epic Handle) and sustainably stored.

Describe collection: Data collections can be registered and individually described in the Collection Registry.

Model data and create Mappings: In the Data Modeling Environment, the data is distinctively modeled or an existing model is applied; furthermore, those models can be associated among each other.

Search and find data: The indexed and modeled data can be searched through the Generic Search.

Thus, all steps that ensure FAIR provision of data are taken into account.

Our talk aims at being a touchpoint to the DFA and an invitation to (international) data providers and users to integrate their collections into the DFA and benefit from its services.

Bibliography

Buddenbohm, Stefan. "CLARIAH-DE - Aligning two Research Infrastructures: Experiences and Challenges." DARIAH Annual Event 2020.

Cremer, Fabian. "DARIAH-DE Repository: Notizen zum Nutzen jenseits der Nutzung." DHd-Blog, August 2018. <https://dhd-blog.org/?p=10368>.

CLARIAH-DE. "Support". Accessed April 7, 2021. <https://www.clariah.de/support>.

DARIAH-DE. DARIAH-DE Repository Documentation: Release 2020-06-25 (2020). <https://repository.de.dariah.eu>

Keywords: DARIAHDE, CLARIAHDE, FAIR, publishing, modeling, metadata, research data, infrastructure