

DataCite – International Data Citation e.V.
Welfengarten 1 B
30167 Hannover

Prof. Dr. Erhard Hinrichs
Spokesperson Text+
Leibniz-Institut für Deutsche Sprache
R-5, 6-13
68161 Mannheim

Letter of support to Text+, consortium for text- and language-based research data

Dear Prof. Hinrichs,

DataCite is a leading global non for profit organization that provides persistent identifiers (DOIs) for research data and other research outputs. DataCite DOI names, as one integrated component of PID services are rooted in and developed by the global research community, thereby ensuring that the community is at the center of activities to guide development, standards, guidelines and best practice, drive adoption, promote advocacy, and participate in the governance of all research-related issues. By providing reliable community owned infrastructure to register DOI names and metadata, DataCite plays a key role in making data FAIR (Findable, Accessible, Interoperable, Reuseable). To showcase and expose metadata, DataCite provides an integrated search interface, DataCite Search, where you can search, filter and extract all the details from a collection of millions of records. This interface is complemented by our OAI-PMH and API services, providing an actionable end-point to harvest DataCite's metadata collection and metadata provenance information. DataCite Search is used by Google Dataset Search to index all research data with a DataCite DOI name. DataCite metadata are also harvested by organizations such as OpenAIRE and Clarivate's Data Citation Index.

DataCite and partners in the EU financed project FREYA have developed a PID Graph, a network of interconnected PID systems, as a basis for a wide range of services. The PID graph can link PI together via relations in their metadata to enable the discovery of connections two steps away. This allows us to e.g. aggregate the citations for all datasets hosted in a particular repository, funded by a particular funder, or created by a particular researcher.

Generic metadata are widely used in research communities of all disciplines. Their use facilitates easier interoperability between systems and disciplines, enable standard—based quality management and in this way provide reliability for re—use of data. In support of the aforementioned benefits DataCite's metadata schema follows the generic approach. Subject-specific metadata, on the other hand, are tailored to the special features of the data and standards used and provide very specific information that is difficult to integrate in a generic schema. Therefore the integration of more extended metadata, using

established community standards in the humanities, into the metadata provided by PIDs is important for FAIR data. DataCite will aim for a generic integration that works with multiple metadata standards, and update our metadata schemata where needed.

Human-, machine-readable and interoperable metadata standards are a central element for the implementation of quality management in repositories. The integration of PIDs is key for the establishment of such standards and is addressed by several quality management concepts. The EU-funded project FAIRsFAIR with participation of 22 representatives of the European research landscape including DataCite, Karlsruhe Institute of Technology (KIT), DANS, Data Curation Center (DCC), SUB Göttingen, Research Data Alliance (RDA) and several European universities aim to extend the certification CoreTrustSeal model to include the FAIR principles. The existing repository metadata provided by re3data.org and FAIRsharing.org will be a starting point for an enriched metadata record including the repository PID build upon the FAIR principles. An additional project goal is to apply and align the interoperability of FAIR repositories and resources in existing and forthcoming research infrastructures across the EU society and globally. This work will be done by re3data COREF (Community Driven Open Reference for Research Data Repositories) a DFG funded project.

DataCite DOI names make research data and other outputs discoverable and citable. This means it becomes easier for the data to be reused. DataCite wants to help track the reuse of data. Therefore, we make data citations obtained through our members and via our collaborator Crossref available through our API and in our search index. DataCite was also involved in drafting the COUNTER code of practice for research data and works with members to help them count views and downloads of their datasets in a standardized fashion. When these counts are submitted to DataCite, DataCite makes these available through APIs and in our search index, making it easy for you to demonstrate data reuse and foster data sharing.

We strongly believe that all of our above mentioned services, which have been developed or are under development in collaboration with the global research community, will add substantial value to Text+. We would like to express our strong interest in working together with the planned consortium.

Sincerely,

Irina Sens, 24.09.2020

Dr. Irina Sens
DataCite Board Member