

Making the Maturity of Data and Metadata Visible with Datacite DOIs

Amandine Kaiser [1], Daniel Heydebreck [1], Anette Ganske [2], Angelina Kraft [2]

[1] German Climate Computing Centre (DKRZ), [2] Technische Informationsbibliothek (TIB)



PRESENTED AT:



INTRODUCTION

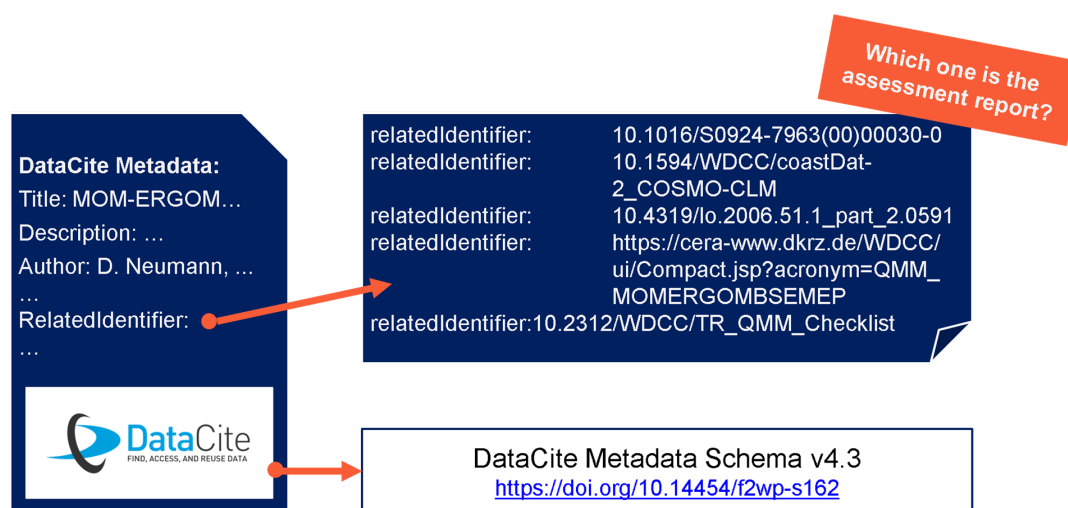
Data publication numbers (incl. DataCite DOI's) are increasing. However, the reusability of data is often difficult, e.g. due to incomplete metadata, missing discipline-specific metadata or insufficient quality information. Therefore, a place where maturity information (e.g. FAIRness and quality) can be stored is needed to easily identify records that are suitable for reuse.

Within the AtMoDat project (<https://www.atmodat.de/>) we have developed a concept that proposes the introduction of a Maturity Indicator as a new property in the DataCite Metadata Schema. This indicator is generic and independent of any scientific discipline and data stewardship tool. Hence, it can be used in a variety of research fields. The goal of the indicator is to enable the scientific community to present high quality data sets, in our case from the field of atmospheric model data.

Based on the discussions with many interested colleagues, we are sure that data users from different research areas will benefit from an explicit feature that displays information about data maturity. It would be gratifying if the property "Maturity Indicator" would be included in the DataCite metadata schema in the near future and we would appreciate your feedback on this proposal.

CURRENT SITUATION

The problem for us: The [DataCite Metadata Schema 4.3](https://doi.org/10.1002/essoar.1004947.1), does not provide an explicit property to link/store assessment reports on data maturity (e.g. FAIRness or quality of data/metadata). Such information can only be added as RelatedIdentifier, which makes it hard for data users to find the information, if and how e.g. data sets have been validated.



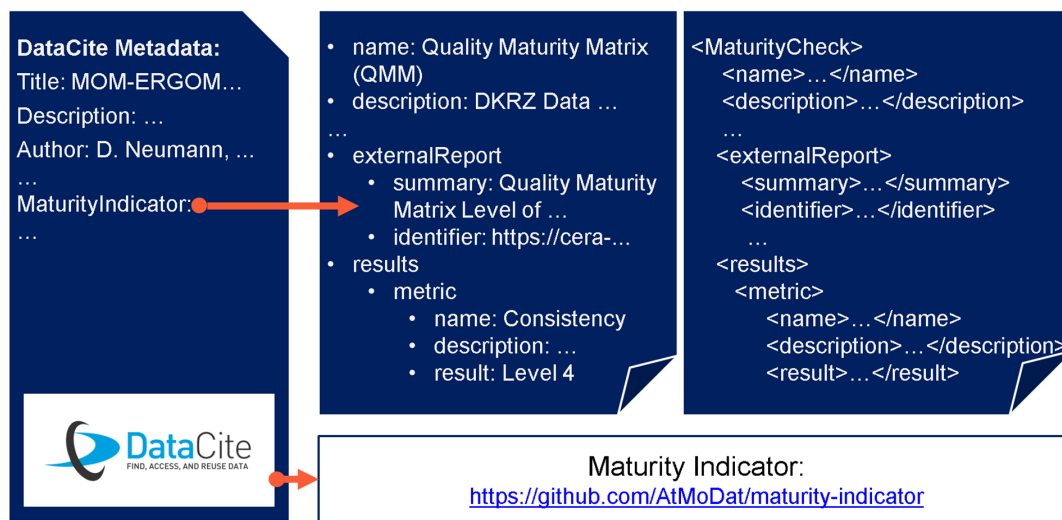
We need a place to store maturity information (e.g. FAIRness and Quality) to enable easy identification of datasets suitable for reuse.

OUR PROPOSAL

Add an unambiguous property for maturity information to the DataCite Metadata Schema.

Concept: Extension of the DataCite Metadata Schema to include the Maturity Indicator, i.e. by adding e.g.

- a link to an external maturity metric
- a link to the result report of an application of a maturity metric
- a maturity metric summary



Examples of Maturity Metrics

- **F-UJI Automated FAIR Data Assessment Tool** (FAIRsFAIR): FAIRsFAIR has developed F-UJI, a service based on REST, and is piloting a programmatic assessment of the FAIRness of research datasets in five trustworthy data repositories. The F-UJI assessment is based on 16 out of 17 core FAIR object assessment metrics developed within FAIRsFAIR and each corresponding to a part or the whole of a FAIR principle. F-UJI adheres to existing web standards and PID resolution services best practices and utilises external registries and resources such as re3data and Datacite APIs, SPDX License List, RDA Metadata Standards Catalog, and Linked Open Vocabularies (LOV). For information on the practical tests implemented against the metrics, see [Devaraju, Huber, et al., 2020](#). The source code is now available with a free license through [Github](#).
- **Quality Maturity Matrix** (WDCC): The WDCC (World Data Climate Center) focuses on climate data products, specifically those resulting from climate simulations. Based on already existing Maturity Matrix models, WDCC developed a generic Quality Assessment System for Earth System data. A self-assessment is performed using a maturity matrix evaluating the data quality for five maturity levels with respect to the criteria data and metadata consistency, completeness, accessibility and accuracy.
- **Data Stewardship Maturity Matrix** (NOAA): Nine key components are identified based on requirements imposed on digital environmental data and information that are cared for and disseminated by U.S. Federal agencies by U.S. law, i.e., Information Quality Act of 2001, agencies' guidance, expert bodies' recommendations, and users. These components include: preservability, accessibility, usability, production sustainability, data quality assurance, data quality control/monitoring, data quality assessment, transparency/traceability, and data integrity.

EXAMPLES AND FURTHER INFORMATION

XML or JSON examples can be found here:

- xml: [example_MI_WDCC_Maturity.xml](#)
- json: [example_MI_WDCC_Maturity.json](#)
- xml: [example_MI_ARDC_FAIR.xml](#)
- json: [example_MI_ARDC_FAIR.json](#)

Further information can be found here:

- see our [poster about the ATMODAT standard](#)
- see for example a recording of a previous webinar about the Maturity Indicator:
www.atmodat.de/p/dmi_webin_talk_view
- EGU presentation: <https://meetingorganizer.copernicus.org/EGU2020/EGU2020-8463.html> (Note: the Maturity Indicator was previously named Quality Indicator)

ACKNOWLEDGEMENTS AND CONTACT INFORMATION

We thank our colleagues for answering questions and providing support on individual topics of the AtMoDat project.

AtMoDat is funded by the German Federal Ministry for Education and Research (BMBF) within the framework of

“Forschungsvorhaben zur Entwicklung und Erprobung von Kurationskriterien und Qualitätsstandards von Forschungsdaten” (FKZ: 16QK02A).



Federal Ministry
of Education
and Research

Contact us: info@atmodat.de

Visit us: <https://www.atmodat.de>

Visit us on GitHub: <https://github.com/AtMoDat>

AUTHOR INFORMATION

Amandine Kaiser: kaiser@dkrz.de; <https://orcid.org/0000-0002-2756-9964>

Daniel Heydebreck: heydebreck@dkrz.de; <https://orcid.org/0000-0001-8574-9093>

Anette Ganske: anette.ganske@tib.eu; <https://orcid.org/0000-0003-1043-4964>

Angelina Kraft: angelina.kraft@tib.eu; <https://orcid.org/0000-0002-6454-335X>

ABSTRACT

Data maturity describes the degree of the formalisation/standardisation of a data object with respect to FAIRness and quality of the (meta-) data. Therefore, a high (meta-) data maturity increases the reusability of data. Moreover, it is an important topic in data management, which is reflected by a growing number of tools and theories trying to measure it, e.g. the FAIR testing tools assessed by RDA¹ or the NOAA maturity matrix².

If the results of stewardship tasks cannot be shown directly in the metadata, reusers of data cannot easily recognise which data is easy to reuse. For example, the DataCite Metadata Schema does not provide an explicit property to link/store information on data maturity (e.g. FAIRness or quality of data/metadata). The AtMoDat project³ (**A**tmospheric **M**odel **D**ata) aims to improve the reusability of published atmospheric model data by scientists, the public sector, companies and other stakeholders. These data are valuable because they form the basis to understand and predict natural events, including the atmospheric circulation and ultimately the atmospheric and planetary energy budget. As most atmospheric data has been published with DataCite DOIs, it is of high importance that the maturity of the datasets can be easily found in the DOI's Metadata. Published data of other fields of research would also benefit from easily findable maturity information.

Therefore, we developed a Maturity Indicator concept and propose to introduce it as a new property in the DataCite Metadata Schema. This indicator is generic and independent of any scientific discipline and data stewardship tool. Hence, it can be used in a variety of research fields.

1 <https://doi.org/10.15497/RDA00034>

2 Peng et al., 2015: <https://doi.org/10.2481/dsj.14-049>

3 www.atmodat.de