Increased visibility of FAIRness and Maturity of Data and Metadata 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)
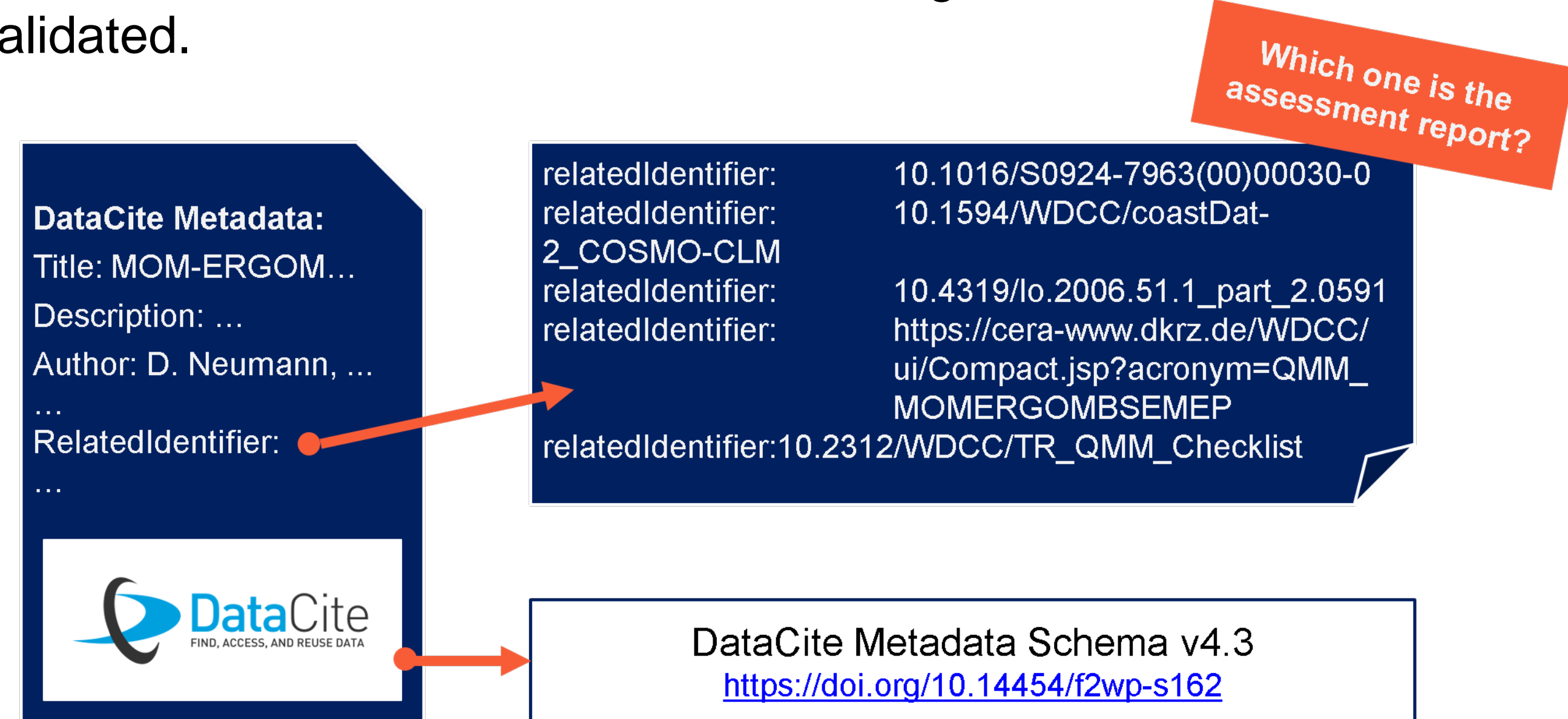
Introduction

Data publication numbers (incl. DataCite DOI's) are increasing, however data reusability is often difficult due to for example incomplete metadata, no discipline-specific metadata or insufficient quality information. Therefore, we need a place to store maturity information (e.g. FAIRness and quality) to enable easy identification of datasets suitable for reuse.

Within the AtMoDat project (https://www.atmodat.de/) we have developed a Maturity Indicator concept which proposes to introduce 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 display high-quality data sets, in our case coming from the field of atmospheric model data.

Due to the discussions with many interested colleagues we are sure that data users from a variety of research fields will profit from an explicit property which displays information about data maturity. We would be delighted if the Maturity Indicator property would be added to the DataCite Metadata Schema in the near future and would welcome

Current Situation

The problem for us: The DataCite Metadata Schema 4.3. 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.

Which one is the assessment report?

DataCite Metadata:
Title: MOM-ERGOM…
Description: …
Author: D. Neumann, …
…
RelatedIdentifier: …
…

relatedIdentifier:	10.1016/S0924-7963(00)00030-0
relatedIdentifier:	10.1594/WDCC/coastDat-2_COSMO-CLM
relatedIdentifier:	10.4319/lo.2006.51.1_part_2.0591
relatedIdentifier:	https://cera-www.dkrz.de/WDCC/ui/Compact.jsp?acronym=QMM_MOMERGOMBSEMEP
relatedIdentifier:	10.2312/WDCC/TR_QMM_Checklist

DataCite Metadata Schema v4.3
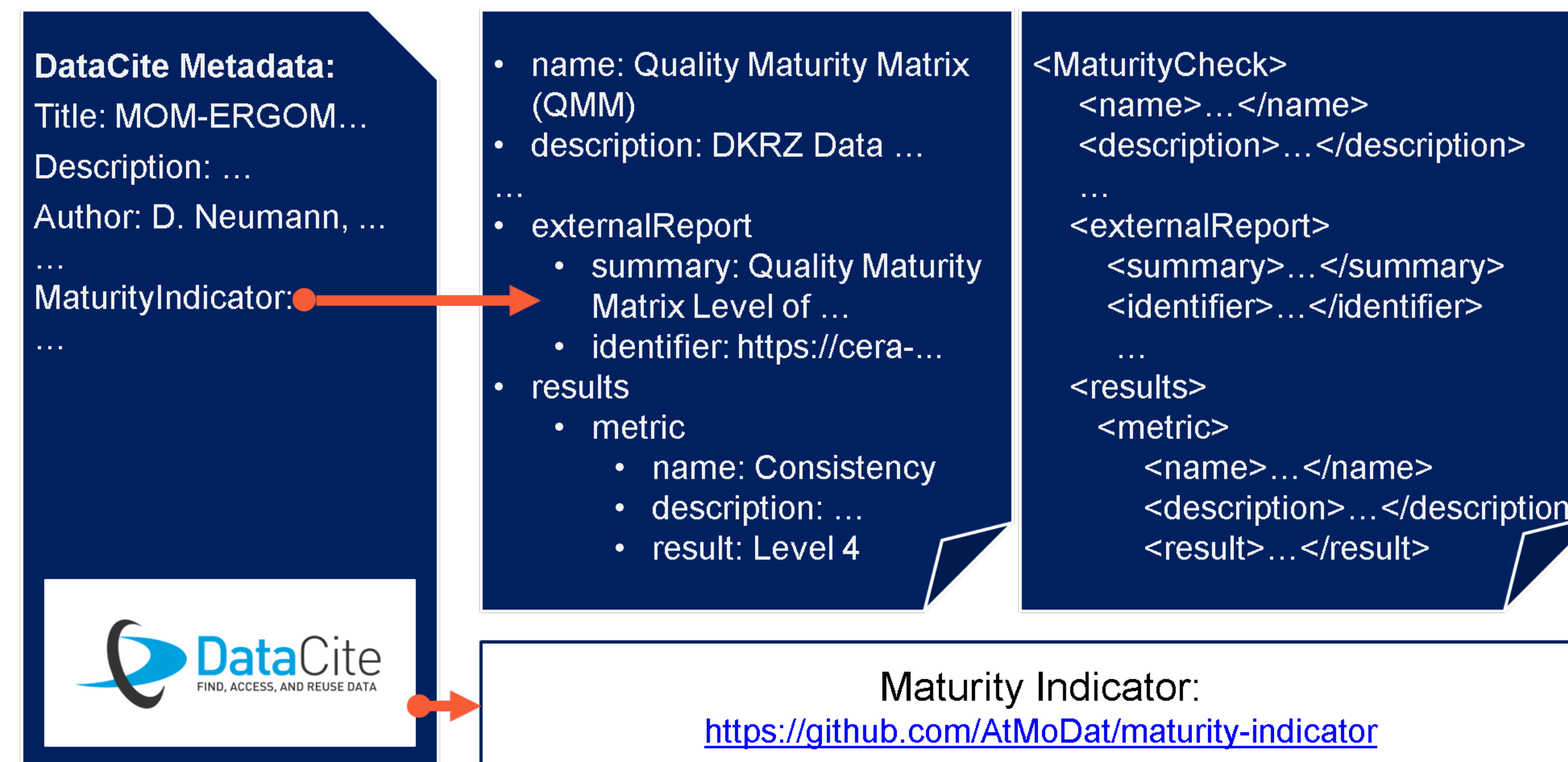https://doi.org/10.14454/f2wp-s162

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

DataCite Metadata:
Title: MOM-ERGOM…
Description: …
Author: D. Neumann, …
…
MaturityIndicator: …

- name: Quality Maturity Matrix (QMM)
- description: DKRZ Data …
…
- externalReport
  - summary: Quality Maturity Matrix Level of …
  - identifier: https://cera-…
- results
  - metric
    - name: Consistency
    - description: …
    - result: Level 4

```
<MaturityCheck>
    <name>…</name>
    <description>…</description>
    …
    <externalReport>
        <summary>…</summary>
        <identifier>…</identifier>
    …
    <results>
        <metric>
            <name>…</name>
            <description>…</description>
            <result>…</result>
```

Maturity Indicator:
https://github.com/AtMoDat/maturity-indicator

Examples of Maturity Metrics

- 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

Contact us: info@atmodat.de
Visit us: https://www.atmodat.de

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

Federal Ministry of Education and Research

anette.ganske@tib.eu; https://orcid.org/0000-0003-1043-4964
angelina.kraft@tib.eu; https://orcid.org/0000-0002-6454-335X

kaiser@dkrz.de; https://orcid.org/0000-0002-2756-9964
heydebreck@dkrz.de; https://orcid.org/0000-0001-8574-9093