# Why we need to highlight FAIR and Open Data and how to do it with EASYDAB

**Anette Ganske**[1], Angelika Heil[2], Hannes Thiemann[2], and Andrea Lammert[2]

[1]TIB – Leibniz-Informationszentrum Technik und Naturwissenschaften, Forschung und Entwicklung, Hannover, Germany (anette.ganske@tib.eu)

[2]Data Management Department, German Climate Computing Center (DKRZ), Bundesstr. 45a, 20146 Hamburg, Germany

The FAIR[1] data principles are important for the findability, accessibility, interoperability, and reusability of data. Therefore, many repositories make huge efforts to curate data so that they become FAIR and assign DataCite[2] DOIs to archived data for increasing the findability. Nevertheless, recent investigations (Strecker[3], 2021) show that many datasets published with a DataCite DOI don't meet all aspects of the FAIR principles, as they are missing important information for the reuse and interoperability in their metadata. Further examinations of data from the Earth System Sciences (ESS) reveal that especially automatic processability is suboptimal, e.g. because of missing persistent identifiers for creators and affiliations, missing information about geolocations or time ranges of simulations. As well, many datasets either don't have any licence information or a non-open licence.

The question arises of how datasets with open licences[4] and high-quality metadata can be highlighted so that they stand out from the crowd of published data. One solution is the newly developed branding for FAIR and open ESS data, called EASYDAB[5] (Earth System Data Branding). It consists of a logo that earmarks landing pages of those datasets with a DataCite DOI, an open licence, open file formats[6], rich metadata, and which were quality controlled by the responsible repository. The EASYDAB logo is protected and may only be used by repositories that agree to follow the EASYDAB Guidelines[7]. These guidelines define principles on how to achieve high metadata quality of ESS datasets by demanding specific metadata information. Domain-specific quality guidelines define the mandatory metadata for a self-explaining description of the data. One example is a quality guideline for atmospheric model data - the ATMODAT Standard[8]. It prescribes not only the metadata for the files but also for the DOI and the landing page. The atmodat data checker[9] additionally helps data providers and repositories to check whether data files meet the requirements of the ATMODAT Standard.

The use of the EASYDAB logo is free of charge, but repositories must sign a contract with TIB – Leibniz Information Centre for Science and Technology[10]. TIB will control in the future that datasets with landing pages highlighted with EASYDAB indeed follow the EASYDAB Guidelines to ensure that EASYDAB remains a branding for high-quality data.

Using EASYDAB, repositories can indicate their efforts to publish data with high-quality metadata.

The EASYDAB logo also indicates to data users that the dataset is quality controlled and can be easily reused.

1: https://www.go-fair.org/fair-principles/

2: https://datacite.org/

3: https://edoc.hu-berlin.de/bitstream/handle/18452/23590/BHR470_Strecker.pdf?sequence=1

4: https://opendefinition.org/licenses/

5: https://www.easydab.de

6: http://opendatahandbook.org/guide/en/appendices/file-formats/

7: https://cera-www.dkrz.de/WDCC/ui/cerasearch/entry?acronym=EASYDAB_Guideline_v1.1

8: https://doi.org/10.35095/WDCC/atmodat_standard_en_v3_0

9: https://github.com/AtMoDat/atmodat_data_checker

10: https://www.tib.eu/en/