World Data Center for Climate (WDCC) – Quality Assurance of Data –

version v1.0; 13 December 2022

How to cite:
World Data Center for Climate (WDCC) (2022, December 13), Quality Assurance of Data, Version v1.0, World Data Center for Climate (WDCC) at the German Computing Center (Deutsches Klimarechenzentrum, DKRZ). https://www.wdc-climate.de/docs/WDCC_quality_assurance.pdf (last accessed: YYYY-MM-DD).

Quality Assurance (QA) of metadata and primary data plays an important role in the data publication process as well as for data re-use. The QA tasks and responsibilities are divided between **checks in preparation of data storage** and checks subsequent to the data storage in tape archive, as the **Technical Quality Assurance (TQA)** on data and metadata consistency, and the **Scientific Quality Assurance (SQA)**, which has to be done and appraised by the data creator.

The results of the QA for the data will become part of the describing metadata during the publication process.

Checks in Preparation of data storage in WDCC
Part of the editorial process is the advice of the data creator regarding file formats and commonly used standards. For NetCDF data this includes the usage of the CF-Checker for all data and communication of errors back to the data creator. We are using the latest stable Version of CF-Checker, provided at https://github.com/cedadev/cf-checker.
Automated checks for all data include:

- Check if files match the given format in the metadata. Check magic numbers inside the files and the file endings
- Check if file names follow the same pattern (different file name-patterns can indicate wrong files)
- Check number of files per dataset (varying numbers can indicate missing files)
- Check date information in file name (if available) - Heuristic approach to check if there are gaps
- Check if all datasets of the corresponding data collections will be archived or if datasets are missing
- Check for erroneous access metadata created in previous steps
- Check for empty files
- Check for file size per dataset (strongly varying file sizes can indicate errors)
- Check for correct checksum and Tracking-ID/PID as additional service for selected projects

In case of failed checks, the necessary edits to the scientific data are only made by the data creator.

Technical Quality Assurance at WDCC

During the cross- and double-checks of WDCC's publication agent at least the following criteria are checked:

1. Number of data sets is correct and > 0
2. Size of every data file is > 0
3. The data sets and corresponding metadata are accessible
4. The data sizes are controlled and correct
5. The spatial-temporal coverage description (metadata) is consistent to the data, time steps are correct and the time coordinate is continuous
6. The format is correct as described in the metadata
7. The description of variables in metadata and data is consistent

Scientific Quality Assurance at WDCC

Contents of scientific quality checks as well as definitions of quality levels and the overall quality procedures vary significantly between data types and scientific disciplines. The SQA has to be done and appraised by the data creator as part of the editorial process and in consultation with the editor, if necessary.