

UK Polar Data Centre Data Quality Checklist:

- **Data value:** check for long-term preservation value using the [NERC Data Value Checklist](#)
- **Data version:** final version for a completed dataset, with all components of the dataset provided and quality checked; specified version for a continuous dataset with relevant components provided and quality checked.
- **Data format:** most suitable and preferably 'open data format' (note: Microsoft products are proprietary formats, where possible consider conversion to open access alternatives, e.g. csv or txt rather than Excel or Word). Data formats encapsulating both data and metadata (e.g. netCDF) are encouraged.
- **Data file names:** short, unique, meaningful, without spaces and special characters (e.g. \*, %), commas or quotes; use consistent naming through related files.
- **Data variable names:** short, unique, meaningful, without spaces and special characters (e.g. \*, %), commas or quotes; use consistent naming through related files. Tabular data should have a variable name in the first row and for each column. Where applicable, use standard names, such as the [Climate and Forecast conventions](#)
- **Data variable units:** defined for each measurable variable, use units as specified by [The Unified Code for Units of Measure](#)
- **Data variable values:** quality checked by data creators, no obvious errors, missing values should be distinguished from true zero values (e.g. designated as 'null' or NaN (not-a-number) or '-99999', where appropriate).
- **Data description:** create readme notes enabling to assess the dataset without opening the data files, explain any abbreviations and acronyms used and document any relevant:
  - data collection techniques
  - instrumentation
  - calibration
  - experimental design
  - sampling regime
  - analytical methods
  - software code, including version
  - quality controls applied
- **Data discovery:** provided all relevant metadata as specified in the [UK PDC metadata guidance](#)
- **Data confidentiality/ethics:** the data do not contain any personal information or other disclosure risks. Sensitive information must be anonymised or be of a sufficiently broad scale that sensitive location information is not made publicly available.