SOOS Data Guidelines

The Southern Ocean Observing System (SOOS) Data Guidelines define the standards for data inclusion within SOOSmap to promote consistency, accessibility, and interoperability. SOOSmap serves as a publicly accessible geospatial data platform supporting interdisciplinary research in the Southern Ocean. These guidelines outline best practices for data submission, formatting, metadata requirements, and geospatial criteria to support open and standardised data sharing, ensuring data follow the FAIR principles (Findability, Accessibility, Interoperability, and Reusability).

Scope

- SOOSmap will only provide access to and make visible **Open (publicly available) Data**. SOOSmap does not have the capacity to show private, limited access data (e.g., sensitive data or Personally Identifiable Information, PII).

- SOOSmap provides access to **observational scientific data**; experimental or modelled data are not within the scope of SOOSmap.

- SOOSmap **supports data across all scientific disciplines**, including physical, biological, chemical and meteorological data.

Data Formats and Structure

- Data must be within an **open-source, standardised format** (e.g., CSV, JSON, NetCDF). For data to be accessible and visible via SOOSmap, they must not be in proprietary formats requiring specific software to read (e.g., SAS7BDAT for statistical software SAS), or in non-machine-readable formats (e.g., PDFs).

- Data **variables must be identifiable and categorised** into one of the following:
    - String or Text
    - Numerical (i.e., discrete, or continuous)
    - Date or Datetime
    - Geometry (i.e., point, line, polygon)

- **Standardised terminologies and vocabularies should be utilised**, where possible, to improve the interoperability of the data. For example, NERC Vocabulary Server (NVS) terminologies, or NetCDF Climate and Forecast (CF) standard names.

Geospatial Requirements

- While SOOSmap has the capacity to provide access to many global datasets, the focus of SOOSmap is data that fall within the **geographic coverage of the Southern Ocean**. This, in turn, is defined as the waters covered by the SOOS Regional Working Groups map. Broadly, the SOOS Regions cover waters south of 50°S, though the northern boundary varies between regions.

✔ Data **must be georeferenced** to be accessible and visible within SOOSmap. Therefore, data must contain coordinates (e.g., Latitude and Longitude), or a geometry (e.g., point, line, polygon).

✔ Data must also **contain a Coordinate Reference System (CRS)** to enable accurate mapping.

Data Storage and Metadata

✔ SOOSmap **does not have the capacity to host data** or act as a data storage option. Observational data within SOOSmap are preferably hosted in a community-approved external data repository (e.g., UK Polar Data Centre, Australian Antarctic Data Centre, etc.).

  ○ If a suitable repository for the data is not known, SOOS can provide guidance to identify the most appropriate national or international repository or data aggregator.

✔ An **external metadata record describing the data must exist**, so that SOOSmap can link to and harvest the following:

  ○ Dataset Title

  ○ Data publisher/owner

  ○ Abstract and details on the methodology surrounding the collection and description of the data

  ○ Data update frequency (e.g., never, daily, yearly)

  ○ Digital Object Identifier (DOI) for the metadata/data, including a permalink for continued data access

  ○ Any licencing/limitations for publishing or using the data

Quality Control and Validation

✔ Data must undergo quality control checks before submission and inclusion to SOOSmap. This includes, but is not limited to:

  ○ Ensuring datetimes are provided in a proper and consistent format (e.g., ISO 8601 "YYYY-MM-DD").

  ○ Ensuring coordinates are all in the same format (e.g., decimal degrees) across the data.

  ○ Outliers and missing values should be documented and handled appropriately.

  ○ Numerical variables contain only numeric values.

  ○ Units are provided and standardised (e.g., The Unified Code for Units of Measure, UCUM), where required.

✔ Data processing steps and known limitations should be clearly stated in the metadata.

Remember:
Cite the data source and *__acknowledge SOOS__ when using data accessed from SOOSmap!*