

# Reuse and republication of MOSAiC data - guideline

This document provides a guideline to the MOSAiC project researchers on the best practice for creating **derived data products** based on previously published MOSAiC data. For authorship of such data products the principles of Good Scientific Practice apply, as well as arrangements from the section *Authorship and Acknowledgment* of the [MOSAIC data policy](#).

## Licenses

- The open access data publication is recommended under [the Creative Commons Attribution 4.0 International \(CC BY 4.0\)](#) or [the CC0 1.0 Universal \(CC0 1.0\) Public Domain Dedication](#)
  - CC BY 4.0: this license requires the attribution to be given to the creator
  - CC0 1.0: CC0 is not a license but a public domain dedication tool, so it does not require attribution in its terms. Nevertheless, giving credit or citing the source is required under principles of Good Scientific Practice.
- It is necessary to consider the licensing of the source datasets, i.e. whether the licenses under which the source data were published allow reuse under upper mentioned licenses. Usually it is possible to publish the derived data products under the same or a less restrictive license than the source data.

## Indicating source data

- Follow the journal article model for listing citations under the references. For example, you can provide specific citations on your derived data product description for each source data set or data stream under the References. This will allow users to cite both the derived product as well as the source data set.
- During a submission of a derived data product, indicate the source data sets to the data center editors as such. They will make sure the source data are correctly assigned at the metadata level.
- In data repositories (PANGAEA, ARM, Arctic Data Center) it is possible to add source DOIs in the source metadata as the related identifiers. Use the relations `IsDerivedFrom`, `IsSourceOf`, or `isPartOf`.
- When you relate all of the datasets in this manner, the metadata gets passed along to DataCite and Scholix, which persistently links those data products through DOI metadata and can help show reuse. Also OSTI is supporting this concept.
- All this will make the data citation metrics possible.

## Documentation of provenance

- The Arctic data center provides a scheme for provenance documentation.
  - Use the web tool on a dataset landing page to document provenance *within* and dataset (ie: this script uses file-a as input and produces file-b)
  - To track relationships between datasets:
    - include a citation to the dataset(s), with DOI, in your narrative metadata (abstract and/or methods)
    - highlight the citation in correspondence with the Arctic Data Center, so the relationship can be established in a way that is computationally readable. If possible, the relationships should be established at the file level. However, in the case of a derived dataset with hundreds or thousands of source files it will be more practical to establish the provenance at the dataset level as opposed to file level.
- Archiving code along with data
  - Is welcome and encouraged in Arctic Data Center. We also accept citations to Zenodo or other repository DOIs.
  - In PANGAEA, we accept and encourage references to externally published code (e.g., GitHub releases with Zenodo DOIs).

## Citing data

- Independent if you refer to raw data, primary data or derived data products, use full data references to acknowledge all data sources.
  - Make use of citation tools at landing pages of data sets or data streams in data repositories. These allow to copy and export citations in various formats.
-