ABSTRACT

The presentation focused on describing a new dataset lifecycle policy that the NASA Physical Oceanography DAAC (PO.DAAC) has implemented for its new and current datasets to foster improved stewardship and consistency across its archive. The overarching goal is to implement this dataset lifecycle policy for all new GHRSST GDS2 datasets and bridge the mission statements from the GHRSST Project Office and PO.DAAC to provide the best quality SST data in a cost-effective, efficient manner, preserving its integrity so that it will be available and usable to a wide audience.

1. Dataset Lifecycle Policy

The primary motivation for the PO.DAAC with respect to the implementation of the policy is to ensure consistency across the data holdings with regard to metadata and formats, data quality and maturity, and to ensure requirements for internal data management best practices are followed. Impacts on data, operations, tools and distribution are assessed through the collection of various metrics. The primary components of the lifecycle are defined by a series of documents designed to collect these lifecycle policy metrics (Fig. 1).

Some of the metrics are related to internal procedures to document system requirements such as impacts on operations, and tools and distribution (e.g., the System Impact Assessment document), but of fundamental importance to the data provider is a document known as the Submission Agreement. This document is part of the lifecycle “quality gate” designed to improve the capturing of data quality and descriptions. Although the document contains sections to establish the respective expectations between the data provider and the PO.DAAC with regards to data latency, tools and services availability, support and distribution requirements, it more importantly contains sections for the provider to document and improve the quality characterization of their dataset including data uncertainty assessment and validation results, and as well as the processing lineage and algorithm description. Components of these sections could come from published literature, project validation results, or project algorithm description documents.

Figure 1. Example of the various facets and documents in the hierarchy of the PO.DAAC dataset lifecycle policy (DSLP). Of importance to a data provider is the Submission Agreement to document data quality.

An example of the populated data quality components in the Submission Agreement is seen in Figure 2 for a Oceansat-2 scatterometer dataset:
2. Conclusion

This Submission Agreement as part of the dataset lifecycle policy is meant to be a first step to assess the dataset quality and can be eventually leveraged to improve GHRSST ISO 19115 metadata records (using data quality DQ_ and lineage LE_ objects) as well. It can also potentially be used to improve dataset selection from the user perspective. After plenary discussion it was agreed that new GHRSST datasets should strive to adopt this lifecycle approach including the Submission Agreement.

Acknowledgements

This work was carried out at the NASA Jet Propulsion Laboratory, California Institute of Technology. Government sponsorship acknowledged. Copyright 2013 California Institute of Technology. Government sponsorship acknowledged.