Dataset Lifecycle Policy

Edward Armstrong, Gary Corlett, Eric Tauer… ..DAS-TAG

14th GHRSST Science Team Meeting

Woods Hole, MA, USA

19 June 2013

National Aeronautics and
Space Administration

Jet Propulsion Laboratory
California Institute of Technology
Pasadena, California

Target: End User

* Mutual Goals:

  Provide the best quality sea surface temperature data for applications in short, medium and decadal/climate time scales in the most cost effective and efficient manner through international collaboration and scientific innovation.
  (GHRSST: https://www.ghrsst.org/ghrsst-science/what-is-ghrsst/)

  Preserve the integrity of the data, regardless of its circumstances, and make it available and usable to a wide audience.
  (PO.DAAC)

Target: End User

- Mutual Goals:

  Provide the ***best quality*** *sea surface temperature data for applications in short, medium and decadal/climate time scales in the most* ***cost effective*** *and* ***efficient*** *manner through international collaboration and scientific innovation.*

  Preserve the integrity *of the data, regardless of its circumstances, and to* ***make it available*** *and* ***usable to a wide audience****.*

The Goal

Bring in Data…

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.

Get it into the hands of the Users…

The Goal

Bring in Data…

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.

Get it into the hands of the Users…

Findability / Usability

The Balancing Act

Bring in Data…

Not forgetting part of directing users to the "right" data could include understanding their level of tolerance regarding "quality"…

Get it into the hands of the Users…

Findability / Usability

The Quality Gate

Bring in Data… → Quality Check

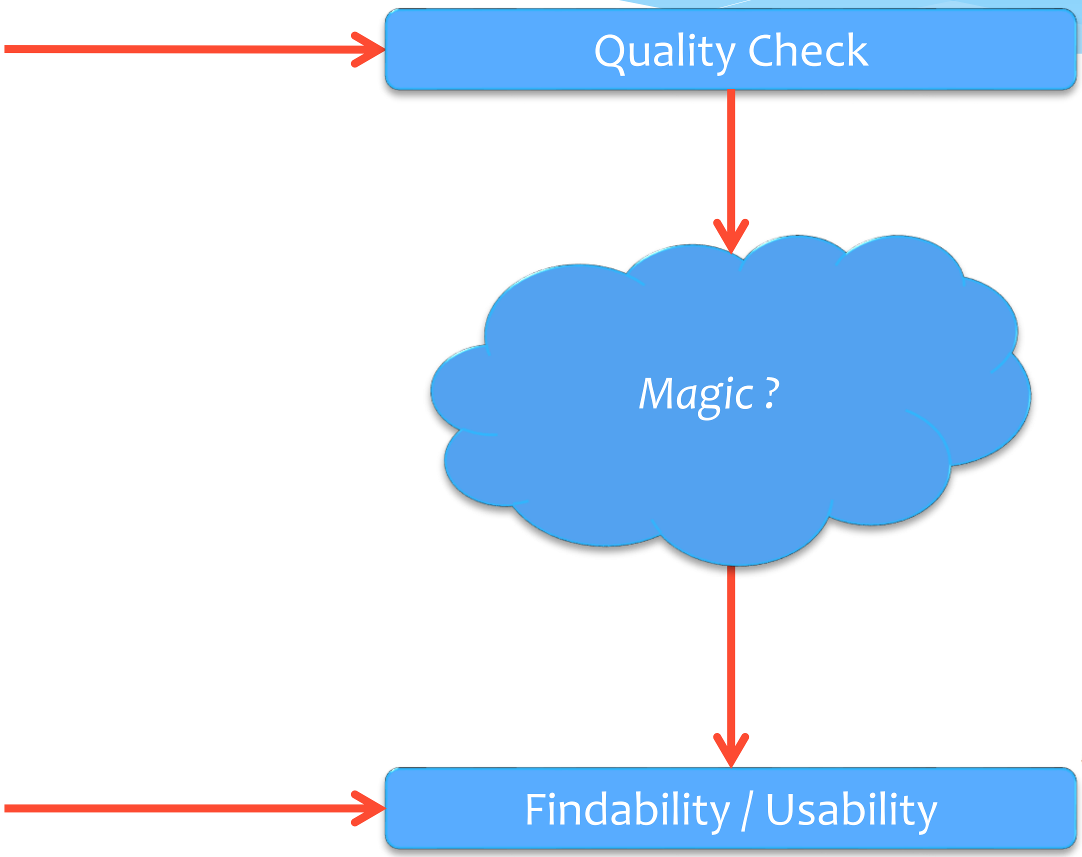
↓

Get it into the hands of the Users… → Findability / Usability

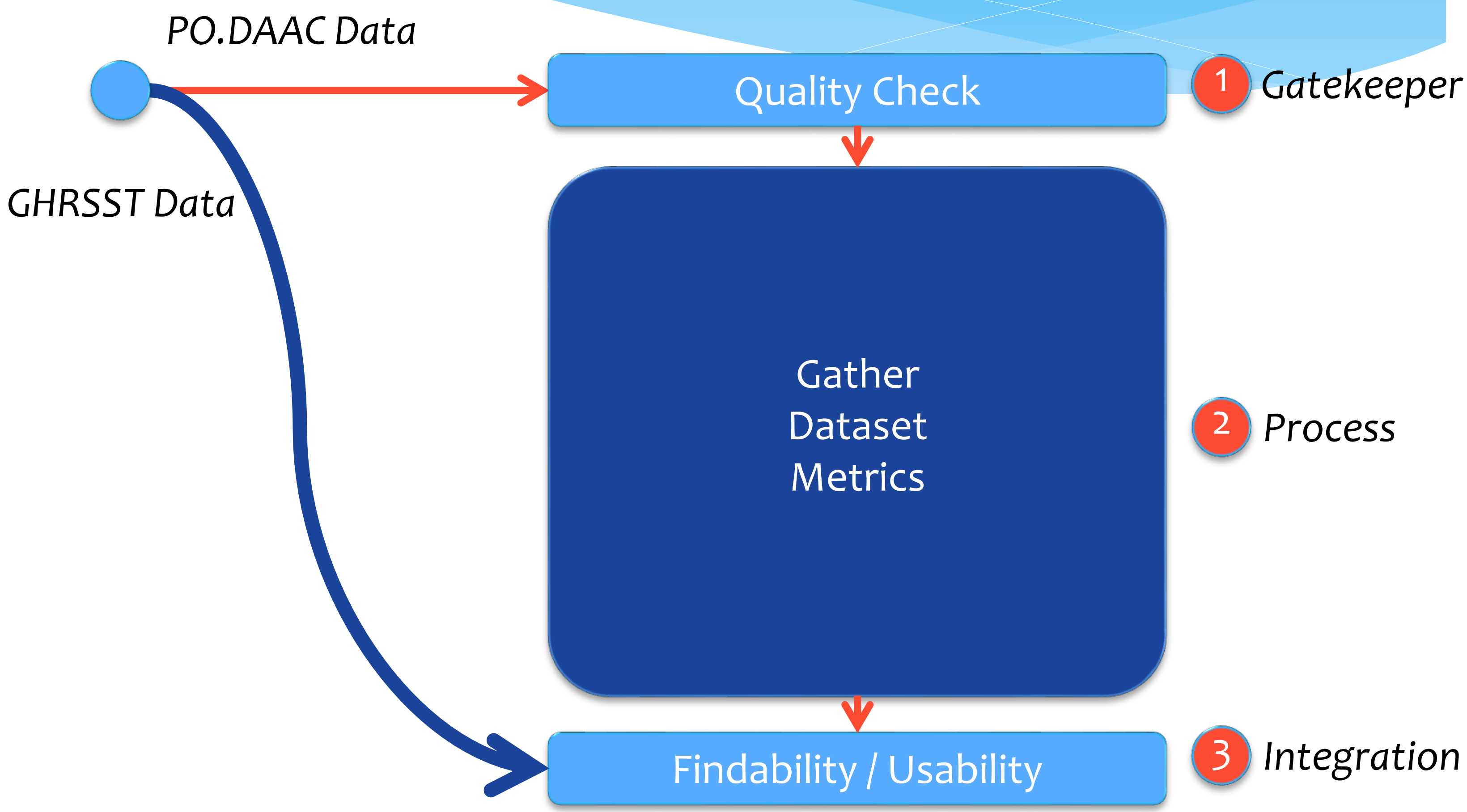
How do you get there from here?

Bring in Data…

Get it into the hands of the Users…

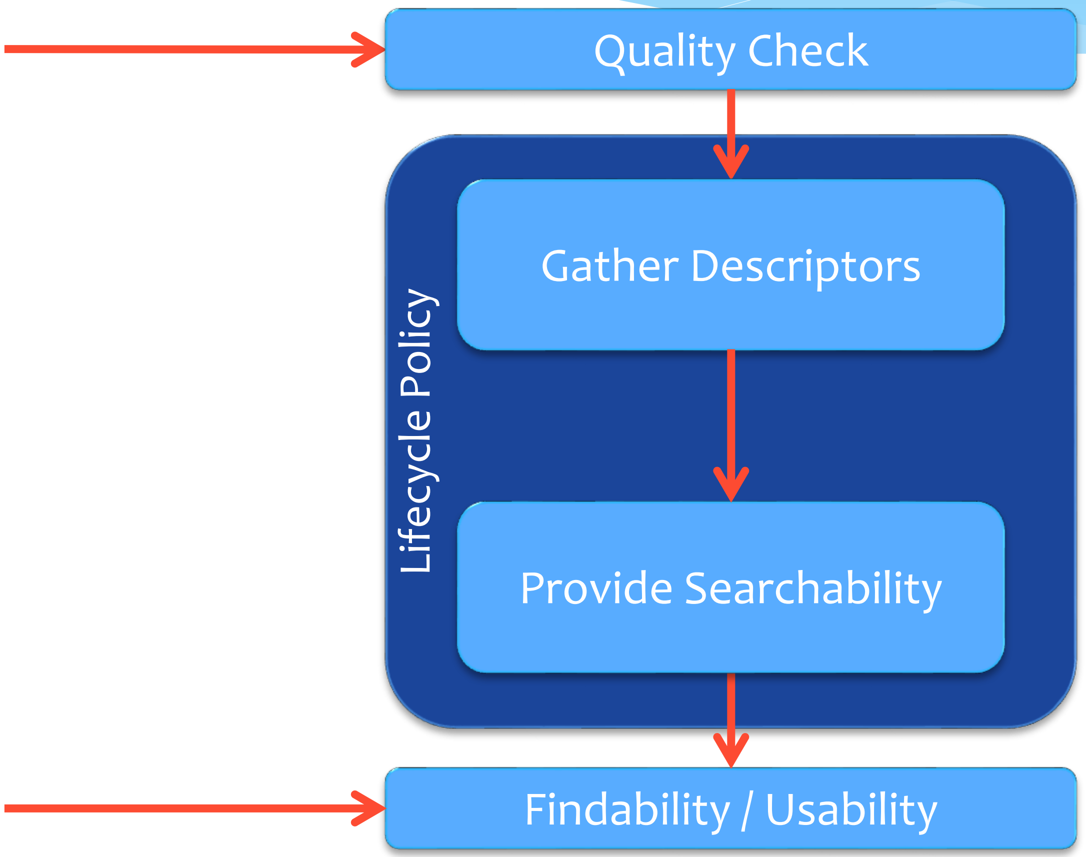


The Challenges



PO.DAAC's Approach

Bring in Data…

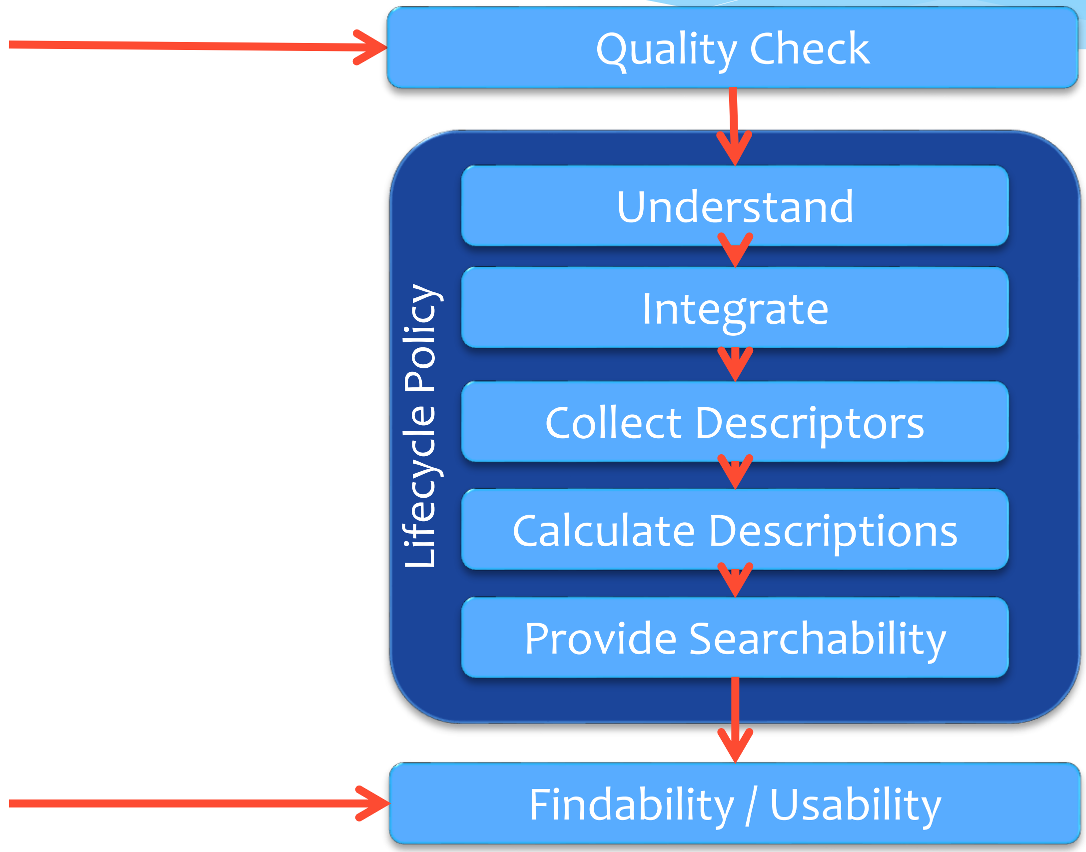


Get it into the hands of the Users…

The Dataset Lifecycle: Conceptual

Bring in Data…

Get it into the hands of the Users…



What methodology is right for GHRSST?

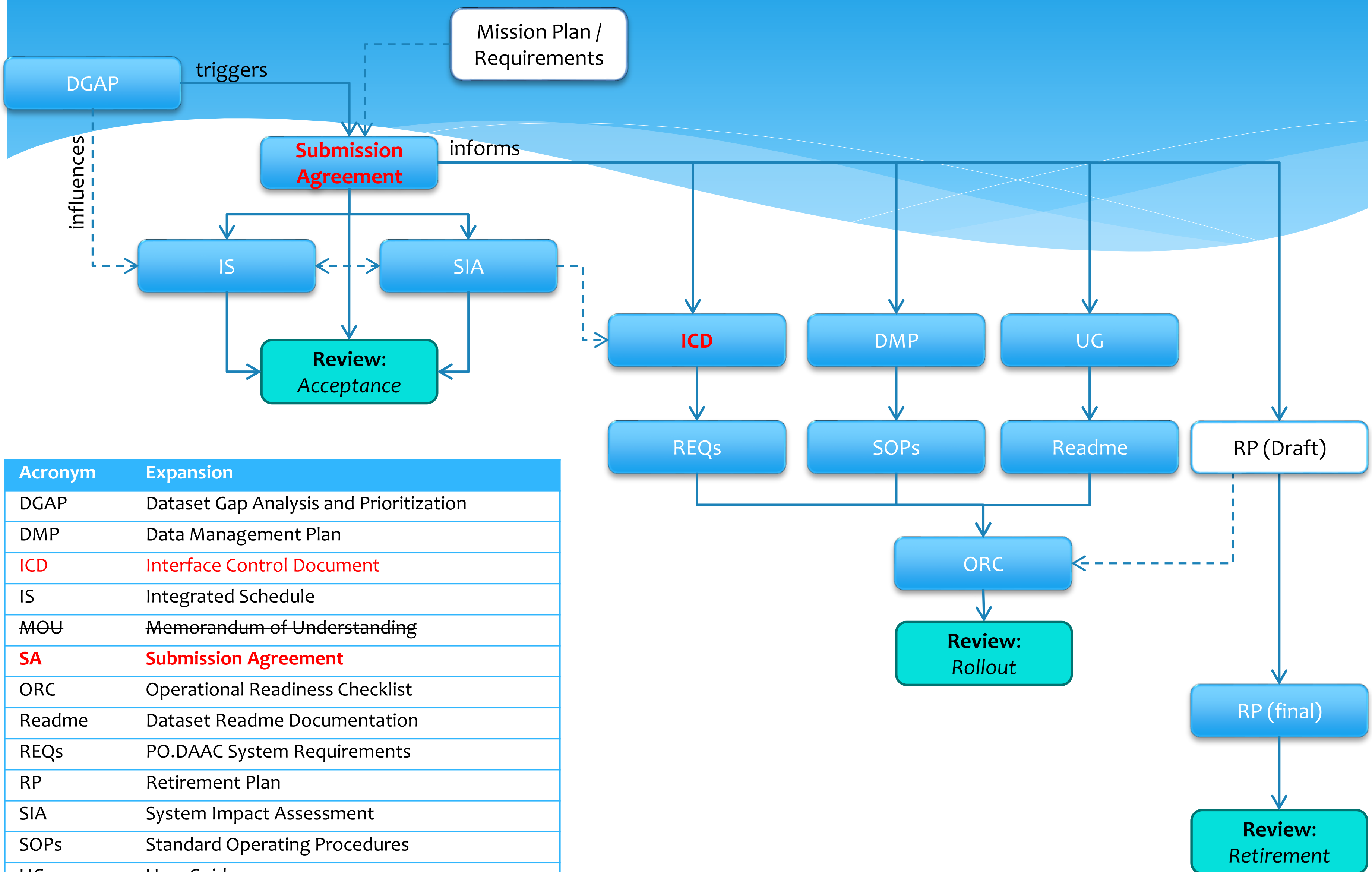
Provide the best quality SST data in a cost-effective, efficient manner, preserving its integrity so that it will be available and usable to a wide audience.

* Process has to be light enough that it will not deter contributors
* Consistent methodology → Consistent Results
* Is PO.DAAC's methodology appropriate? (*Perhaps lighter weight, but conceptually equivalent…*)

Lifecycle Policy: Purpose

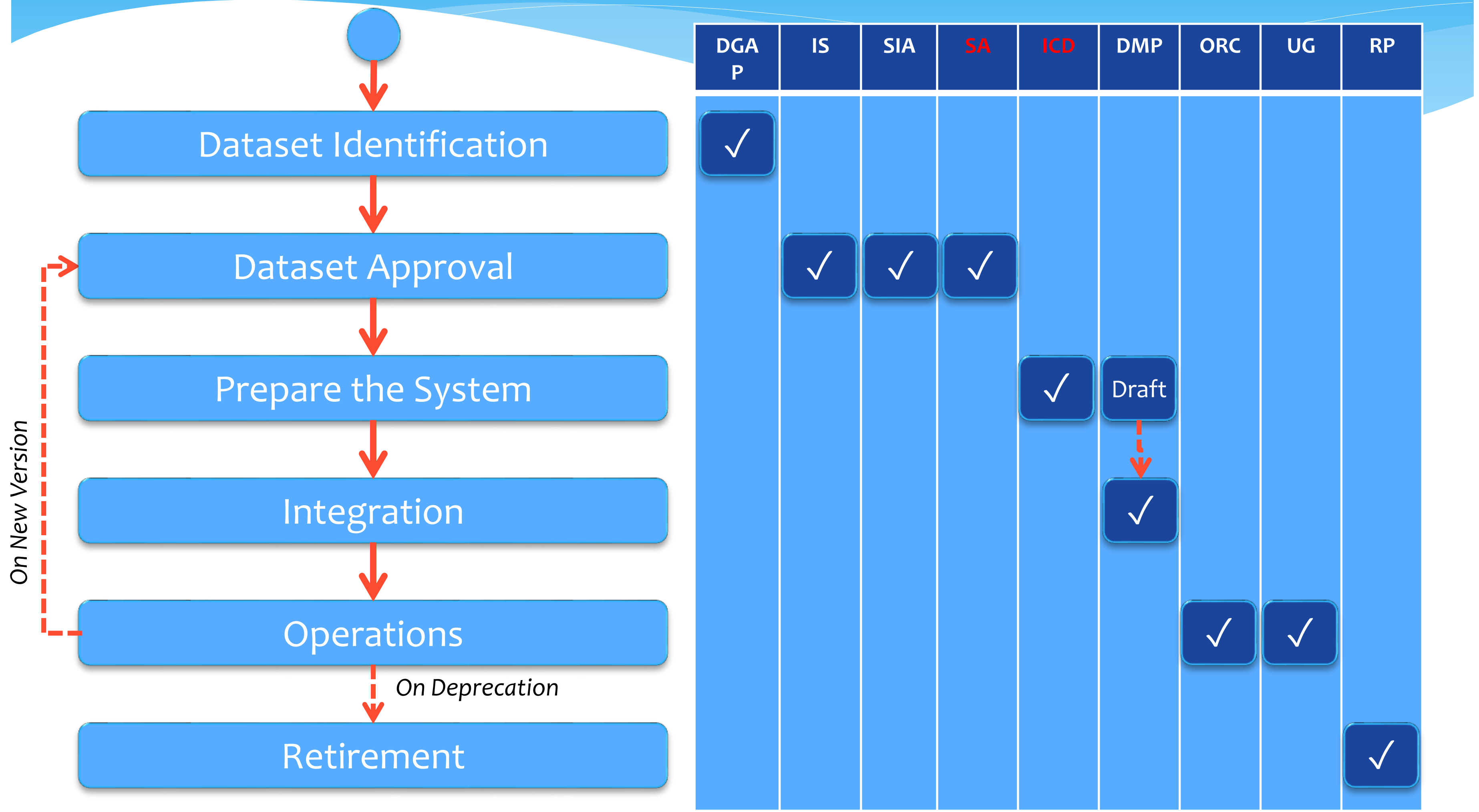
* Document driven:
  * Simplifies the methodology
  * Supports end-goals:
    * Ability to report on maturity and accountability
    * Enable searchability and usability
  * Metrics are required in order to achieve those goals
  * ***The "policy" – generation of key artifacts in the proper form – ensures the collectability of those metrics.***

DSLP Artifact Relationships



Acronym	Expansion
DGAP	Dataset Gap Analysis and Prioritization
DMP	Data Management Plan
ICD	Interface Control Document
IS	Integrated Schedule
~~MOU~~	~~Memorandum of Understanding~~
SA	**Submission Agreement**
ORC	Operational Readiness Checklist
Readme	Dataset Readme Documentation
REQs	PO.DAAC System Requirements
RP	Retirement Plan
SIA	System Impact Assessment
SOPs	Standard Operating Procedures
UG	User Guide

Lifecycle Phases and Documents



Lifecycle Policy: Purpose

* The purpose of describing the lifecycle policy is to ensure our approach is:
  * Consistent across our holdings
  * Follows best practices, and
  * Is adequate to meet our requirements.

Lifecycle Approach

* Main Components:
  * ***Quality Gate***: Ensure quality gets into the system and control resources.
  * ***Policy***: Provide consistency of treatment, ensure metadata is solid, and verify the integrity of the files.
  * ***Users***: Facilitate selection process and provide tools.

Mutual Goals:

* PO.DAAC would like to follow the same, consistent approach with GHRSST datasets:
  * Would like to align our metrics with GHRSST's metrics
  * Would like to align our maturity approach with GHRSST's approach

Dataset Lifecycle Submission Agreement

~~Memo of Understanding~~

Submission Agreement

Between the NASA Physical Oceanography Distributed Active Archive Center (PO.DAAC) and [Institution X] for cooperation in support of [Dataset X]

Introduction:

Identify the parties of the MOU including technical point of contacts for provider and PO.DAAC.

Purpose and Scope

Identify purpose of the MOU. Which dataset does it target? What are the responsibilities of each signature party? For example:

- [Institution X] will provide reasonable timely access to [Dataset X] via the protocols specified in the Interface Control Document (ICD)
- [Institution X] will provide data and metadata in compliance with HDF/netCDF/CF/ACDD
- [Institution X] will assist in the preparation of a User Guide
- [Institution X] will provide retirement and end of mission support
- PO.DAAC will archive and distribute [Dataset X] as directed by [Institution X]
- PO.DAAC will prepare [Dataset X] in full compliance of the PO.DAAC Dataset Lifecycle Policy including metadata, documentation, read software and user support
- PO.DAAC will maintain dataset documentation and metadata.
- PO.DAAC will expose [Dataset X] to its tools and web services suite based on recommended data and metadata model compliance. Services include FTP, THREDDS, and OPeNDAP.
- New versions of dataset may recycle entire portions of this MOU
- [Institution X] will provide a Dataset Technical Description (see below)

Policy

- Describe any adjustments to the NASA data policy of free and open distribution (if any)

Restrictions

Any restriction or disclaimers of the data noted (if any)

Dataset Technical Description

Submission Agreement

Dataset Technical Description

Instrument Overview *(one or two paragraphs with one or two supporting figures)*

Describe the instrument science objective, capability, measurement principle, satellite and orbit characteristics. Provide the strengths and weaknesses of the instrument measurement or derived measurement.

Provide:

- *Platform(s):*
- *Instrument(s):*
- *Spatial Resolution:*
- *Temporal Repeat:*
- *Temporal Resolution Nominal:*
- *Temporal Resolution Max:*
- *Temporal Resolution Min:*
- *Temporal Sample Frequency:*
- *Swath Width:*
- *Spatial Coverage:*
- *Temporal Coverage:*

Lineage *(one to two paragraphs)*

Describe the origin of the dataset in terms of the measurement principle and actual on-orbit measurements made. Provide an observation map, if coverage is not global.

Describe the processing applied to the on-orbit observations that was used to generate the geophysical variable presented in this dataset. If the data processing involved is complex, provide only the gross overview, and references later for details. Note any models or a priori assumptions used in the processing. If this dataset merges measurements from more than one instrument, describe how that is done.

Give a quantitative description of the sampling used in creating the gridded product, and its spatial variation (as necessary).

If this dataset is a new version of an existing dataset describe what has been updated.

Validation and Uncertainty Estimate *(one or two paragraphs with one supporting table or figure)*

Summarize the validation work done for this dataset and/or the dataset from which this product is derived (e.g. the Level 2 or 3 product).

Provide an error/uncertainty estimate and describe any systematic biases. If the uncertainty/and or bias varies spatially or temporally, provide an appropriate description of that variation. Use a table or figure to illustrate the main result if appropriate. Provide references later for more details.

Provide the typical range of the variance of the data used to construct this product.

Submission Agreement

If the uncertainty estimate and/or variance have significant per datum variation, provide these in a separate file that can be associated with this dataset, and note the file name(s) here.

References

Add references for details that may be of interest to the more sophisticated user. Use this space to describe some details that are not appropriate for the main text.

PO.DAAC will provide dataset citation information including DOI.

Definitions

Any definition or acronyms identified

Signature

This MOU is acknowledged as a non binding cooperative agreement between the following parties on the ________of________, 20__
Digital signatures are acceptable.

For the PO.DAAC:

____

For [Institution X]:

____

backups

Exposing quality info

4. Validation and Uncertainty Estimate

Buoy measurements from the Tropical Atmosphere Ocean / Triangle Trans-Ocean Buoy Network (TAO/TRITON) were used to validate the AMSR-E SSTs. The mean difference, satellite minus buoy SST, and standard deviation (STD) were calculated using

1 May 2002 through 9 June 2005.The TAO array shows AMSR-E to have very small biases (-0.03 C) and STD (0.41 C) [2].

5. Considerations for Model-Observation Comparisons

5.1 Time sampling bias

The sun synchronous orbit of AMSR-E yields retrievals at a local time of approximately 1:30 AM and 1:30 PM [5].The diurnal cycle is not resolved or well measured by polar orbiting satellites because their sun-synchronous design results in the measurement of only 2 points of the diurnal cycle. In the case of AMSR-E, the diurnal cycle is measured at 1:30 AM and 1:30 PM (local time). To avoid the diurnal warming present in the daytime measurements, only the nighttime, or 1:30 AM, observations were included in the monthly averages.

5.2 Inhomogeneous sampling

The sampling is not globally homogeneous because of missing SST data, primarily due to rain. Some rain contaminated SSTs exist in AMSR-E data. At the edges of rain cells, there may be undetected rain that causes a biased SST retrieval. Version 7 AMSR-E SSTs has enhanced rain flagging to minimize this occurrence. Figure 1 below shows a lower number of observations along the equator. These low values are due to the Intertropical Convergence Zone (ITCZ) where strong convection leads to a region of increased precipitation. There is also an area in the Arctic where there is extremely high number of observations. At this latitude and above, the polar region is measured by almost every orbit. A similar occurrence happens at the southern pole but it is not seen because of land. There are also lower numbers of retrievals near the coastline of North America and around the European/Mediterranean regions. These are due to radio-frequency-interference from geostationary cable satellite transmissions, which can reflect off the ocean, back to the satellite, and contaminate the data. When this happens, data are flagged and removed from the data set.

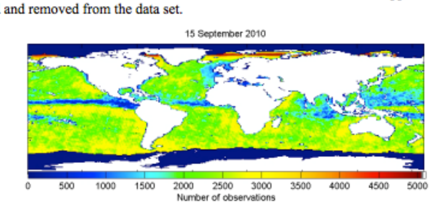


Figure 1: Number of SST observations included in the monthly average of September 2010

Quality info: Obs4MIPS, CMIP5 intercomparison

AMSR-E

Sea Surface Temperature

1. Intent of This Document and Point of Contact (POC)

1a) This document is intended for users who wish to compare satellite derived observations with climate model output in the context of the CMIP5/IPCC historical experiments. Users are not expected to be experts in satellite derived Earth system observational data. This document summarizes essential information needed for comparing this dataset to climate model output. References are provided at the end of this document to additional information.

This NASA dataset is provided as part of an experimental activity to increase the usability of NASA satellite observational data for the modeling and model analysis communities. This is not a standard NASA satellite instrument product, but does represent an effort on behalf of data experts to identify a product that is appropriate for routine model evaluation. The data may have been reprocessed, reformatted, or created solely for comparisons with climate model output. Community feedback to improve and validate the dataset for modeling usage is appreciated. Email comments to HQ-CLIMATE-OBS@mail.nasa.gov.

Dataset File Name (as they appear on the Earth Science Grid, ESG) are:

tos_AMSRE_L3_v7_200206-201012.nc

tosNobs_AMSRE_L3_v7_200206-201012.nc

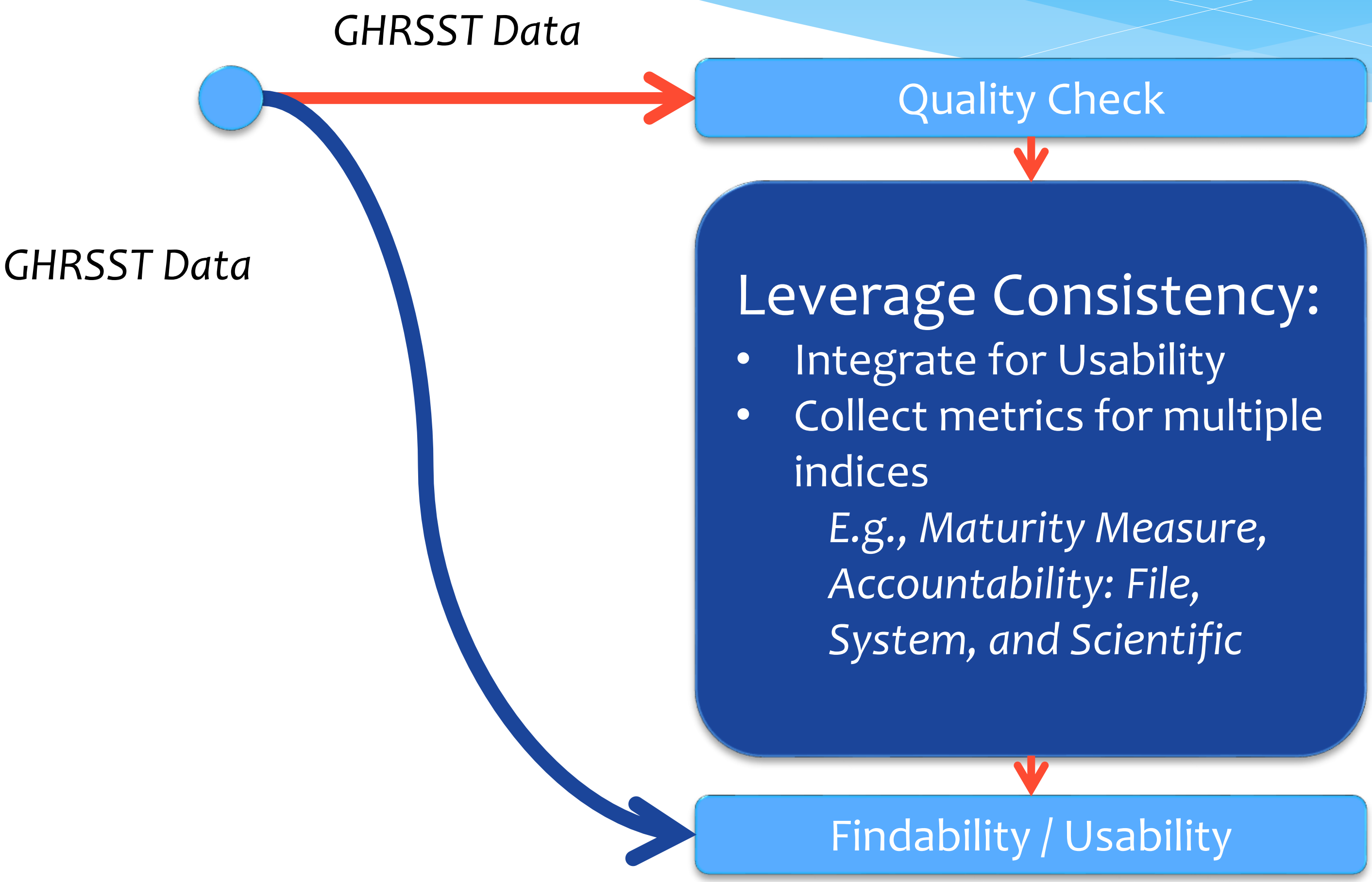
tosStderr_AMSRE_L3_v7_200206-201012.nc

1b) Technical point of contact for this dataset:

Chelle L. Gentemann, gentemann@remss.com, also email:

support@remss.com

The Opportunities



How do you get there from here?

Bring in Data…

Quality Check

Not Magic:
Gather Dataset
Metrics

Get it into the hands of the Users…

Findability / Usability