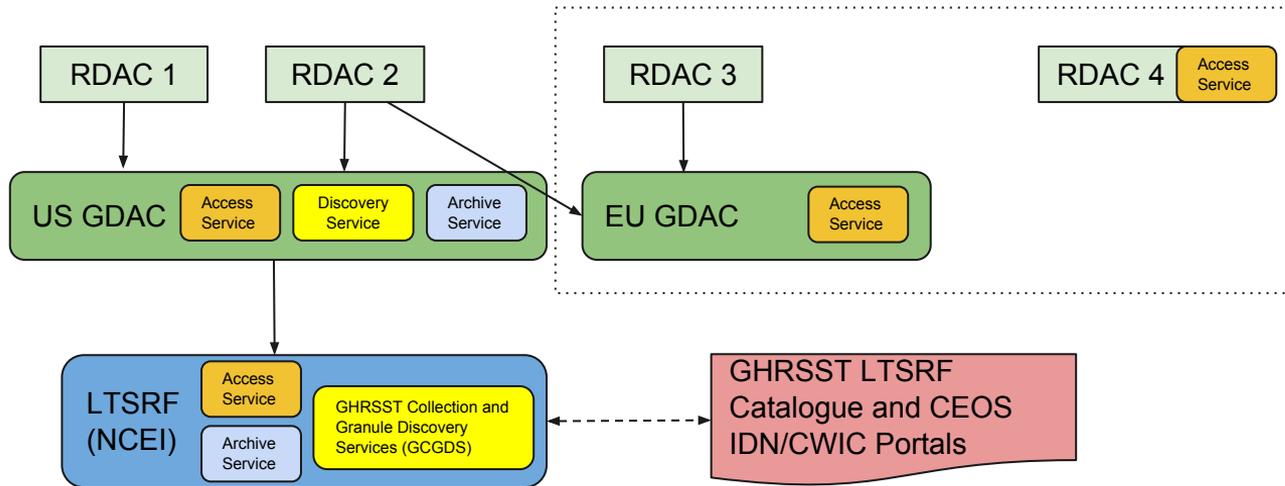# The new GHRSST Regional/Global Task Sharing

Current system design

Working well but:
- No single host for all datasets
- No one-stop catalogue and data access
- More providers distributing their own datasets (Eumetsat, CMEMS, Jaxa,...) without connection to a GDAC
- Somewhat outdated system with growing number of products and providers, difficult to maintain GDACs up-to-date
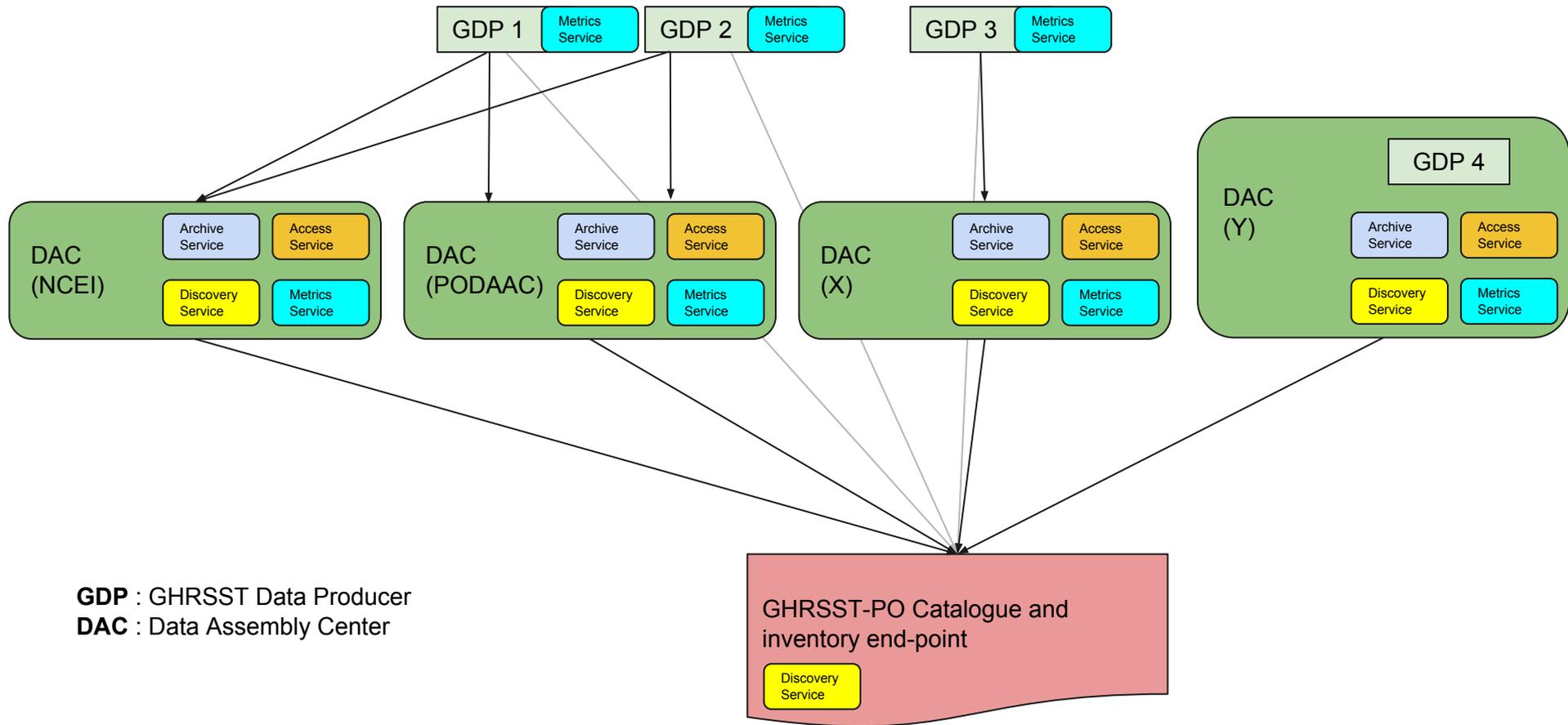
Agreed at GHRSST 17

- A new R/G TS shall be implemented, less centralized in terms of storage and data access, relying on DACs
- A minimum set of **consistent services** shall be implemented by each DAC and/or producer to guarantee homogeneous access and information for users:
    - Data access
    - Data discovery
    - Data archiving
    - Metrics
- Model for R/G TS to be agreed for GHRSST 18

Agreed at GHRSST 18

- Data access (DACs and producers acting as DACs)
    - Mandatory : HTTP or HTTPS
    - Strongly recommended : FTP and DAP
    - Recommended : WMS and WCS (for L3/L4)
- Data discovery
    - Catalogue : CSW (and possibly Opensearch)
    - Inventory : Opensearch
    - Each DAC implementing these services and complying to GHRSST minimum requirements for vocabulary and keywords (scenario 1)
- Archive and metrics are still not clearly specified and rough ideas for now
- No formal approval procedure took place yet, we are still in the proposal status

New R/GTS proposal

GDP 1 | Metrics Service
GDP 2 | Metrics Service
GDP 3 | Metrics Service

DAC (NCEI) — Archive Service | Access Service | Discovery Service | Metrics Service

DAC (PODAAC) — Archive Service | Access Service | Discovery Service | Metrics Service

DAC (X) — Archive Service | Access Service | Discovery Service | Metrics Service

DAC (Y) — GDP 4 | Archive Service | Access Service | Discovery Service | Metrics Service

GHRSST-PO Catalogue and inventory end-point — Discovery Service

GDP : GHRSST Data Producer
DAC : Data Assembly Center

What happened since / objectives for GHRSST 19

- Further analysis by UDS-TAG team : https://docs.google.com/document/d/1-CDw1eBDTEKv_AeF3nfsy4WCJsqRmeg8t81wq1VOhv0/edit?usp=sharing
- Converge and agree on implementation scenario for discovery services
- Main priority for specifications and implementation are data access and discovery/search services, starting with a pilot project
- Get formal approval and implementation roadmap for access and discovery services
- Metrics can be implemented as an external service for now by GHRSST-PO or other, e.g. monitoring of access services availability / user statistics by DACs as now - homogenized definition of this service would come later
- Archiving is usually addressed by organization in their own way and does not seem critical at this stage
- Commitment by DACs to make accessible full archive of datasets

Discovery and search services

Catalogue : dataset (or collection) level metadata

Inventory : granule level metadata - description and localization of each stored file

Discovery : see what is available

Search : select based on some user filters

**Human readable interface** : access through a web browser, typically web pages

Machine-to-machine interfaces : access through scripts, federated queries, typically webservices

Discovery and search services

service	Usage	Human interface implementation	Machine-to-machine interface implementation
Central catalogue (GHRSST-PO)	Single up-to-date list of all known GHRSST datasets Main access point for users with no prior knowledge of what and where datasets are available - explicit redirection to DACs Edited remotely by producers (dataset description) and DACs (access services)	Catalogue web page with dataset descriptions	**CSW webserviceI** **ISO 19139-3 metadata** Integration of GHRSST catalogue (or subset) in tools or other portals without duplicating metadata Already implemented by many data centers
Federated inventory (DACs)	Single end-point at GHRSST-PO / local end-point at each DAC No knowledge required by users of where data are physically stored when querying from GHRSST-PO - federated query with implicit redirection Several URLs returned for granules accessible at different DACs or through different access services	Virtual repository (folder and file tree)	**Opensearch webservice** Search query arguments : dataset, time frame, bounding box Usage by automatic scripts Service already implemented in some data centers

R/G TS refined data discovery, search and access system
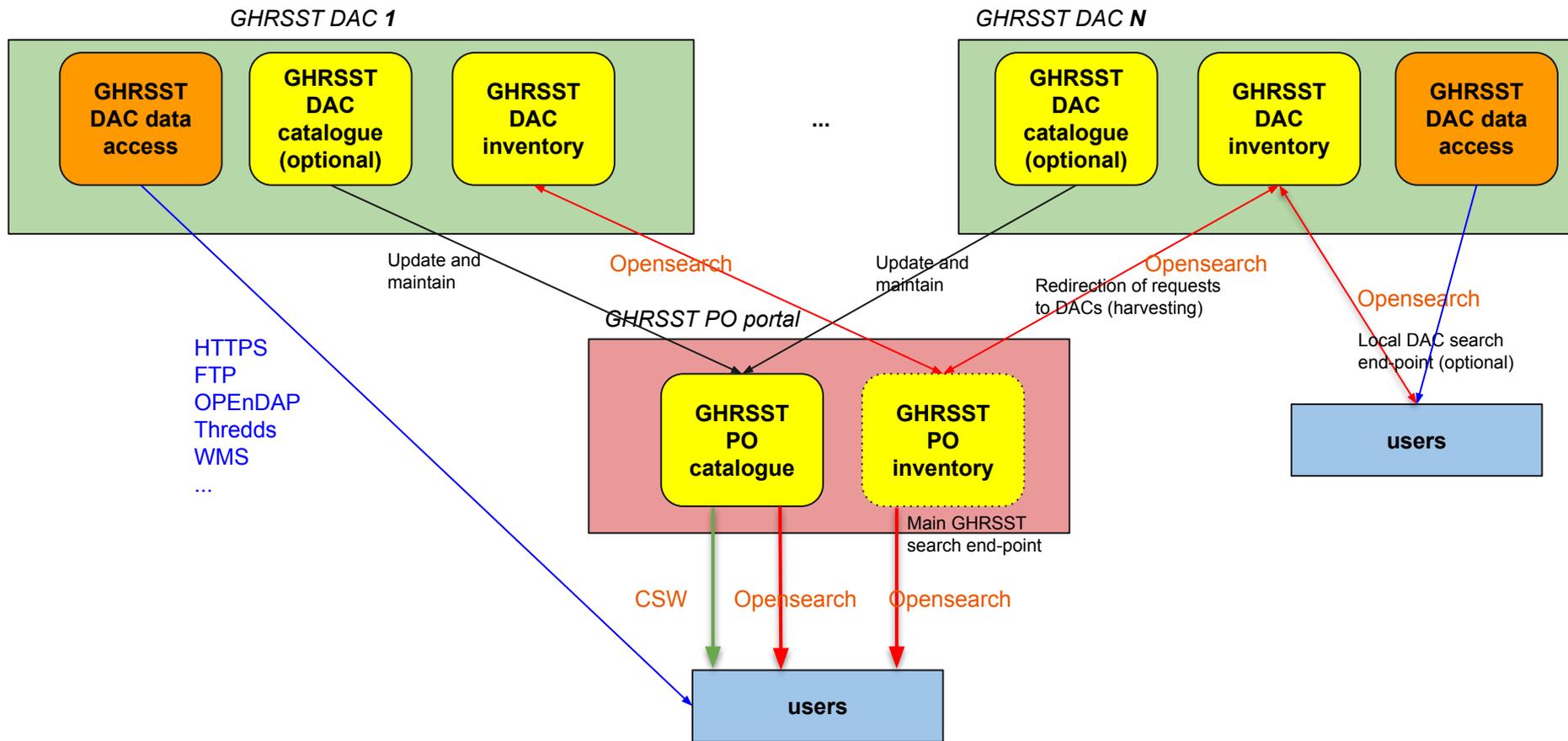
Organization	If a GDP, available products	If a DAC, related GDPs
Eumetsat	Sentinel-3A & S3B L2P	self
JPL PODAAC	JPL JPL_OUROCEAN	self REMSS NAVO CMC NCEI
NCEI	X	self OSPO STAR (future DAC) ABOM UFRJ
STAR	X	self
Ifremer	L4 Medspiration Med Sea L4 Medspiration Brazil L4 Medspiration South-Africa	self OSI SAF NEODAAS NAVO REMSS
CMEMS	L3S Global L4 Global OSTIA L4 North Western Shelves L4 Baltic L4 Arctic L4 Med L4 Black Sea	self
JAXA	Himawari	

Organization	If a GDP, available products	If a DAC, related GDPs
OSI SAF	L2P Metop-A AVHRR L2P Metop-A IASI L3C Metop-A AVHRR Global L3C Metop-A AVHRR NAR L2P Metop-B AVHRR L2P Metop-B IASI L3C Metop-B AVHRR Global L3C Metop-B AVHRR NAR L3C VIIRS AVHRR NAR L3C MSG L3C GOES L3C MSG reprocessed L3C GOES reprocessed	
NEODAAS		
OSPO		
ABoM	L2P Himawari-8 L2P AVHRR HRPT L3U AVHRR HRPT L3C AVHRR HRPT L3S AVHRR HRPT L2P AVHRR HRPT reprocessed L3U AVHRR HRPT reprocessed L3C AVHRR HRPT reprocessed L3S AVHRR HRPT reprocessed L3C VIIRS L3S VIIRS+AVHRR L3C VIIRS reprocessed L3S VIIRS+AVHRR reprocessed L3U MTSAT-1R reprocessed L4 RAMSSA L4 GAMSSA	
UFRJ		
NAVO	VIIRS L2P	
CMC	CMC L4	
REMSS		

Proposed roadmap

By end 2018:

- Pilot project demonstrating central catalogue management and inventory federated queries
    - Editable catalogue with CSW webservice and ISO 19139-3 datasets form (proposed by Ifremer)
    - Federated queries based on existing Opensearch granule search services at NOAA and NASA
    - Demonstrate feasibility and user usage and benefit
- Finalize specifications and software survey for implementation

Progressive implementation by 2019

The GHRSST Science Team is requested to endorse the proposed R/G TS update and to proceed to demonstrate its feasibility through a pilot project.

Example of remotely editable catalogue, ISO 19139-3 compliant

Summary of UDS TAG discussion

- Agreement on centralized catalogue
- Agreement on federated inventory
  - **Opt.1 : Based on existing webservice => restricted to the core set of search arguments shared by these webservices**
  - Opt.2 : based on common interface specifications => require implementation of new webservices
- Available software for DACs being investigated and assessed
- Implementation roadmap
  - DACs (inventory search) based on existing services (ex: at NOAA and NASA)
  - GHRSST-PO (central catalogue and federated search) based on existing ISO19139/Inspire existing service (e.g. Sextant at Ifremer)
  - Pilot project to demonstrate feasibility and user benefit, and finalization of requirements and specifications by end 2018

Other points

- GDS 2.0rev7 updated by GHRSST PO
  - Removing all active tables (see attached) and moving them to the GHRSST website
  - Clarification and update to GHRSST Product Levels (possibly leading to GDS2.1)
  - Correction of solar zenith angle CDL (see email discussion between Ed and Andy Harris)
  - Update metadata to CF-1.6 and ACDD-1.3 and add additional parameters agreed in Qingdao
- Revision of geostationary specs and format
  - Discussion on product level non conclusive at GHRSST-18 - way forward proposed by P. Cornillon
  - https://docs.google.com/document/d/1JYhGv90RIxvMKALDGk1B1a7Rf1GF8kHd2sH5ChDD-QY/edit?usp=sharing
  - New issue raised by Sasha : lat/lon in geostationary files with satellite projection
    - Replace with formula (CF compliant) or ancillary location file
  - Proposal on L2C by Sasha to be considered or matched to an already existing product level

EOSDIS product level

Level 0
Reconstructed, unprocessed instrument and payload data at full resolution, with any and all communications artifacts (e.g., synchronization frames, communications headers, duplicate data) removed. (In most cases, the EOS Data and Operations System (EDOS) provides these data to the data centers as production data sets for processing by the Science Data Processing Segment (SDPS) or by a SIPS to produce higher-level products.)

Level 1A
Reconstructed, unprocessed instrument data at full resolution, time-referenced, and annotated with ancillary information, including radiometric and geometric calibration coefficients and georeferencing parameters (e.g., platform ephemeris) computed and appended but not applied to Level 0 data.

Level 1B
Level 1A data that have been processed to sensor units (not all instruments have Level 1B source data).

Level 2
Derived geophysical variables at the same resolution and location as Level 1 source data.

Level 3
Variables mapped on uniform space-time grid scales, usually with some completeness and consistency.

Level 4
Model output or results from analyses of lower-level data (e.g., variables derived from multiple measurements).

CEOS product level

Level 0
Reconstructed unprocessed instrument data at full space time resolution with all available supplemental information to be used in subsequent processing (e.g., ephemeris, health and safety) appended.

Level 1
Unpacked, reformatted level 0 data, with all supplemental information to be used in subsequent processing appended. Optional radiometric and geometric correction applied to produce parameters in physical units. Data generally presented as full time/space resolution. A wide variety of sub level products are possible.

Level 2
Retrieved environmental variables (e.g., ocean wave height, soil moisture, ice concentration) at the same resolution and location as the level 1 source data.

Level 3
Data or retrieved environmental variables which have been spatially and/or temporally re-sampled (i.e., derived from level 1 or 2 products). Such re-sampling may include averaging and compositing.

Level 4
Model output or results from analyses of lower level data (i.e., variables that are not directly measured by the instruments, but are derived from these measurements).

Data Level - CEOS Definition : http://www.ceos.org/images/WGISS/Documents/Handbook.pdf

Current Level [LEO]	Proposed change (if any) for variants	Definition
L2P		Geophysical variables derived from Level 1 source data at the same resolution and location as the Level 1 data, typically in a satellite projection with geographic information. These data form the fundamental basis for higher level GHRSST products and require ancillary data and uncertainty estimates. No adjustments to input SST have been made.
	L2P	Along swath products with downgraded resolution *VIIRS at 1500m resolution, AVHRR GAC*
	Is there any? L3?	Along swath products with some gap filling (interpolation,...) over cloudy pixels
L3U		L2 data granules remapped to a space grid without combining any observations from overlapping orbits
		Degraded resolution L2 (GAC) ?
L3C		SST measurements combined from a single instrument into a space - time grid. Multiple passes/scenes of data can be combined. Adjustments may be made to input SST data.
L3S		SST measurements combined from multiple instruments into a space - time grid. Multiple passes/scenes of data are combined. Adjustments may be made to input SST data.
L4		Data sets created from the analysis of lower level data that results in gridded, gap-free products. SST data generated from Multiple sources of satellite data using optimal interpolation are an example of L4 GHRSST products

Current Level [GEO]	Proposed change (if any) for variants	Definition
L2P		Geophysical variables derived from Level 1 source data at the same resolution and location as the Level 1 data, typically in a satellite projection with geographic information. These data form the fundamental basis for higher level GHRSST products and require ancillary data and uncertainty estimates. No adjustments to input SST have been made.
	L2P	Single geostationary snapshot in original satellite projection *Ex: Used by OSPO for GEO products at full resolution.* *Ex: Used by BoM for GEO products MTSAT-1R and Himawari-8 at full resolution on the original GEO projection.*
L3U		L2 data granules remapped to a space grid without combining any observations from overlapping orbits
	L3U	Single geostationary snapshot resampled on a new space grid, e.g. plate carree projection *Ex: Used by OSPO for GEO products full-disk at 0.02 degree grid spacing.* *Ex. Used by BoM for GEO MTSAT-1R full-disk products at 0.05 degree grid spacing.*
L3C		SST measurements combined from a single instrument into a space - time grid. Multiple passes/scenes of data can be combined. Adjustments may be made to input SST data.
	L3C	Combination of several geostationary snapshots in original satellite projection *New NOAA L2C*
	L3C	Combination of several geostationary snapshots on a new space grid, e.g. plate carree projection *Ex: OSI SAF MSG/Seviri hourly product* *Ex: BoM IMOS Himawari-8 hourly L3C product remapped to regular 0.02 degree grid over IMOS domain.*
L3S		SST measurements combined from multiple instruments into a space - time grid. Multiple passes/scenes of data are combined. Adjustments may be made to input SST data.
L4		Data sets created from the analysis of lower level data that results in gridded, gap-free products. SST data generated from Multiple sources of satellite data using optimal interpolation are an example of L4 GHRSST products