ISO 19115-2 metadata for the Group for High Resolution SST (GHRSST)

Ed Armstrong
Senior Data Scientist
JPL Physical Oceanography DAAC
California Institute of Technology
27 Nov 2012

GHRSST *(http://www.ghrsst.org)*

* Group for High Resolution Sea Surface Temperature (formerly GODAE High . . . .)

Accessibility | Contacts | Enquiry Form | Login

Search: GO

GHRSST

Group for High Resolution Sea Surface Temperature

- Home
- Latest SST map
- What is GHRSST?
- Site Map
- Data Access
- Applications
- User Requirements
- Data Products
- Product Validation
- Science Team & Groups
- GHRSST articles
- Operational Announcements
- Contact us
- GHRSST Metrics Dashboard
- New Satellite programs
- GHRSST Data Tools
- What's New
- Calendar
- Documents

Your location: Home / About GHRSST

What is the GHRSST ?

The aim of the GHRSST is to 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 c

ollaboration and scientific innovation.

The Global Ocean Data Assimilation Experiment (GODAE) is an international collaboration for ocean forecasting activities which, in 2002, initiated a GODAE High Resolution SST Pilot Project GHRSST-PP to address an emerging need for accurate high resolution sea surface temperature (SST) products. The Pilot Project was replaced by the Group for High Resolution SST (GHRSST) in 2009. SST is required by operational ocean and atmospheric forecasting systems to constrain the modelled upper ocean circulation thermal structure and for exchange of energy between the ocean and atmosphere.

Today the GHRSST is a truly international project with over $18 Million US invested across all of the project activities. It orchestrates a wide variety of GHRSST data streams input and output data that must be shared, indexed, processed, quality controlled, analysed and documented within an international framework called the Regional/Global Task Sharing (R/GTS) framework. Large volumes of data and data services must be harnessed together to deliver the new generation of global coverage high resolution SST data sets that meet the GHRSST-PP GHRSST User Requirements. A basic 'version-1.0' of the R/GTS framework has

What is GHRSST ?

* An international consortium founded in 2001 for the production of improved and consistent SST records coordinated across four continents, providing SST data from nearly every SST measuring satellite in common data and metadata formats
  * Over 60 unique SST products produced to date for 15 unique SST sensors
    * Only satellite product that provides uncertainty estimates on a pixel-by-pixel basis
  * Distributes 4 TB/month
  * Previously used a netCDF 3 data model (with internal CF metadata) and pseudo NASA DIF external metadata model
  * Next generation products (2012):
    * netCDF 4 data model
    * ISO 19115-2 metadata model

Why Comply with Metadata Standards ?

* Improved interoperability via
  * Data Discovery
    * For others to know that your data exists
  * Data Usage
    * For others to access and read your data
    * Useful to automated tools (e.g., for mashups)
  * Data Understanding
    * For others to understand the meaning, context, and limitations of your data, including journal articles. Legacy preservation
    * Repurposing – applying data for different purpose than originally envisioned

About ISO

* International Standards Organization
* Largest developer of int'l standards
* NGO
* 163 countries participate
* HQ in Switzerland
* Coordinated by Technical Committees (TC)
* TC 211 (Geographic Information) coordinates the 19XXX series
* Int'l Standards are more universal than national standards

Who is using ISO 19115?

* ISO 19115 is being adopted in several places:
  * WMO, GOES-R, GEOS portals, FAA (airspace redesign project), Europe, Australia.
  * NOAA is mapping existing holdings in NGDC and NCDC.
  * SMAP using ISO 19115
  * GCMD has mapped its DIF attributes to ISO (22 core elements)
  * GHRSST has adopted ISO 19915-2
  * NASA's Metadata Evolution for NASA Data Systems (MENDS) is developing recommendations and guidelines for the Earth Science Data and Information System (ESDIS) by Dec 2010.

ISO Standards

- ISO 19115
  - metadata standard for data
  - Provides descriptions of spatially referenced data
  - 400 attributes, 22 core attributes, many packages
- ISO 19115-2
  - Extension to 19115 for gridded remote sensing data and images (all spatially referenced)
- All use XML representations
- ISO 19139
  - XML representation or schema for encoding ISO 19115

ISO Standards

* ISO 19130
  * Description of locating satellite pixels to earth coordinates. Includes sensor and imaging models. Full orbit description. Support frame, pushbroom and whiskbroom passive sensors (and eventually active sensors like SAR).
* ISO 19119
  * Metadata standard for services
* ISO 19110
  * Feature types

What Information can be Represented in ISO 19115?

* File/directory structures
* Variables and their properties including quality information
* Space/time extent (bounding boxes, polygons, named regions, …)
* Measuring equipment used to acquire the data
* Geometry of the measuring process employed by the equipment
* Production process used to digitize the raw data
* Derivation of geophysical information from raw data
* Numerical methods and computational procedures used in the derivation
* Readme info (“Series”)
* And more. . .

ISO advantages

* For discovery purposes only modest improvements over current metadata models (e.g., NASA DIF, FGDC, CF etc.)
* Advantages come from elements the refer to:
  * Quality description
  * Processing lineage
  * Sensor/Instrument descriptions
  * Usage for automated tools
  * Interoperability at an international level

Significant metadata packages for earth science data and instruments

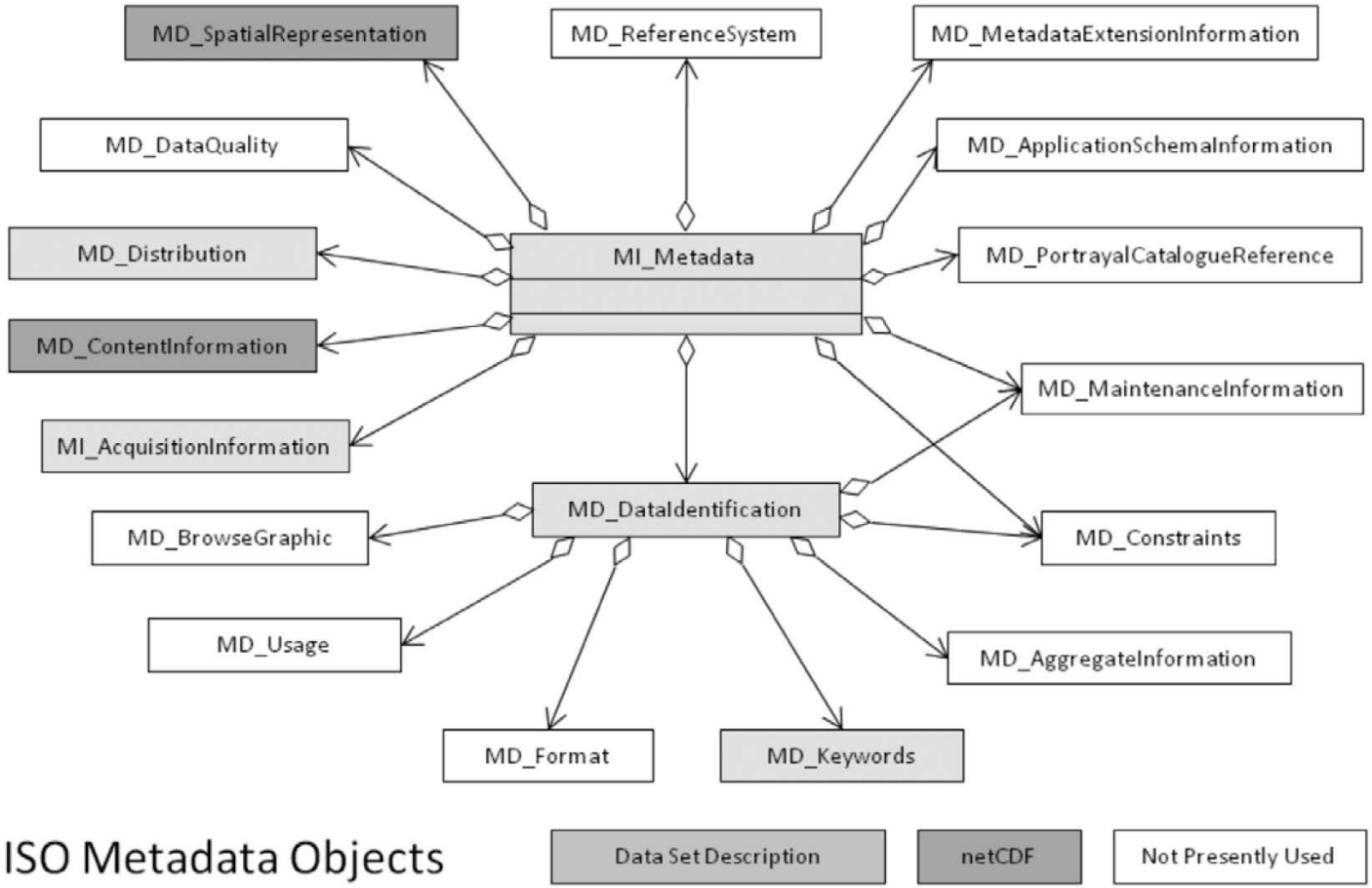
* DQ_ Data quality reports
* LI_ Lineage and provenance reports
* SD_ Sensor geo-positioning (ISO 19130)
* LE_Lineage Extended (ISO 19115-2)
* MI_Metadata for Imagery (ISO 19115-2)
* QE_ Data quality Extended (ISO 19115-2)

Case Study: applying an ISO 19115-2 metadata model to products from the Group for High Resolution Sea Surface Temperature

Details

* New ISO metadata requirements driven by European space agency institutions
* ISO 19115-2 metadata model (remote sensing extensions of 19115)
  * MI_Metadata Object as its root
  * Metadata contained in an external file

GHRSST ISO objects



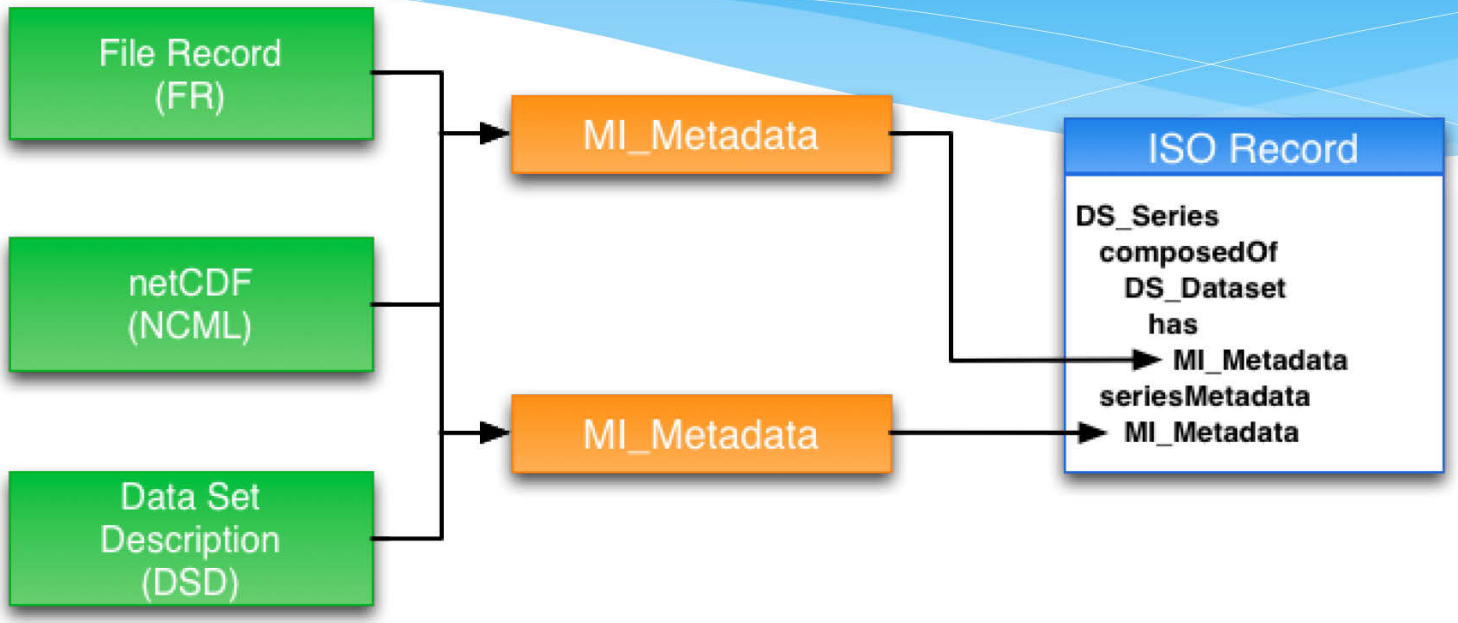
Contents of Metadata Conventions Document

- ISO 19115-2 metadata model (remote sensing extensions of 19115)
- Based on the MI_Metadata Object as its root
- Combines both granule and collection metadata information
  - DS_Dataset object: granule metadata
  - seriesMetadata object: collection level metadata

Aggregation

- ISO allows possibilities of aggregating collection level and granule metadata
- GHRSST will deliver records in a single file
  - Combine both granule and collection metadata information under DS_Series objection
    - DS_Dataset object: granule metadata
    - MI object: additional collection level metadata
- A school of thought suggests packaging ISO records into self describing files like netCDF and HDF. This is possible but not trivial to do.

GHRSST METADATA TO ISO



GHRSST Data Processing Specification version 2 on Metadata Conventions depicting the workflow of metadata translation for both data set and granule (file) level metadata to ISO 19115-2.

Lineage and Quality implementation for GHRSST

* Only MI_CoverageDescription classes implemented for coverage/array information including those GHRSST variables deemed “quality information” such as flags, error statistics etc.
* DQ_, LI_, QE_ classes not implemented yet

Example of coverage types

5 GHRSST Coverage Types

Table 10-1. L

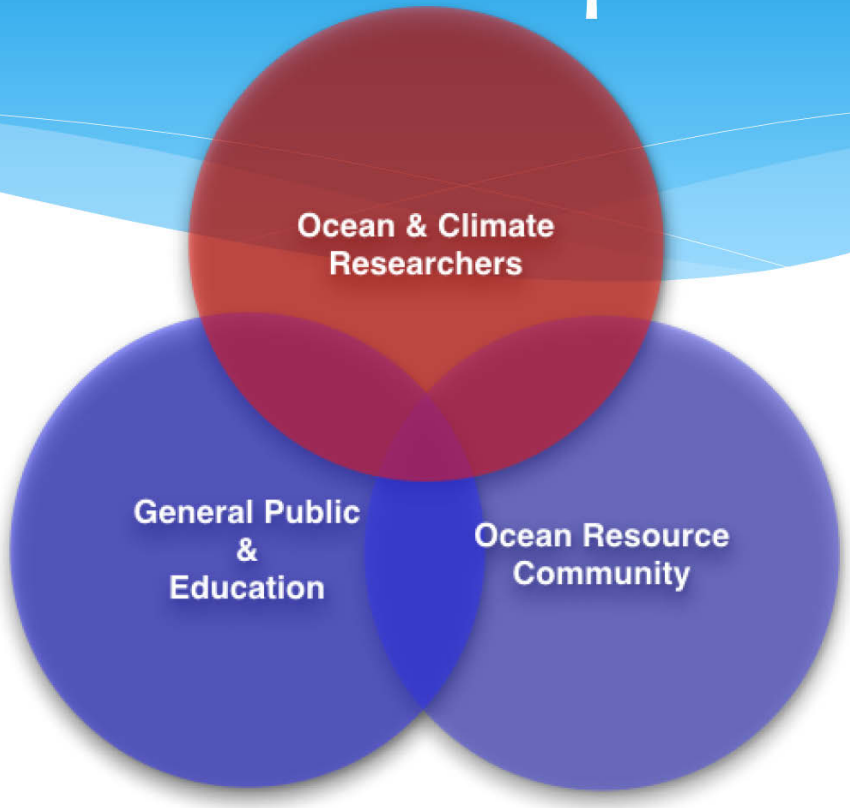
Variable	CoverageType
L2P and L2P_GRIDDED	
time	referenceInformation
lat	referenceInformation
lon	referenceInformation
sst_dtime	referenceInformation
sea_surface_temperature	physicalMeasurement
polar_stereographic	referenceInformation
sea_ice_fraction	auxilliaryData
sources_of_sea_ice_fraction	qualityInformation
DT_analysis	qualityInformation
rejection_flag	qualityInformation
proximity_confidence	qualityInformation
confidence_flag	qualityInformation
SSES_bias_error	qualityInformation
SSES_standard_deviation_error	qualityInformation
aerosol_optical_depth	auxilliaryData

OCSI

* Oceanographic Common Search Interface
  * Restful-based tool provided by the PO.DAAC to provide metadata and search web services to PO.DAAC data holdings

Our Users Need Our Help

* Discover/Identify the relevant data
  * Deliver information (metadata) that our user communities can understand
    * What to package?
    * How to package?
* Retrieve the relevant data
* Use and Understand the data content

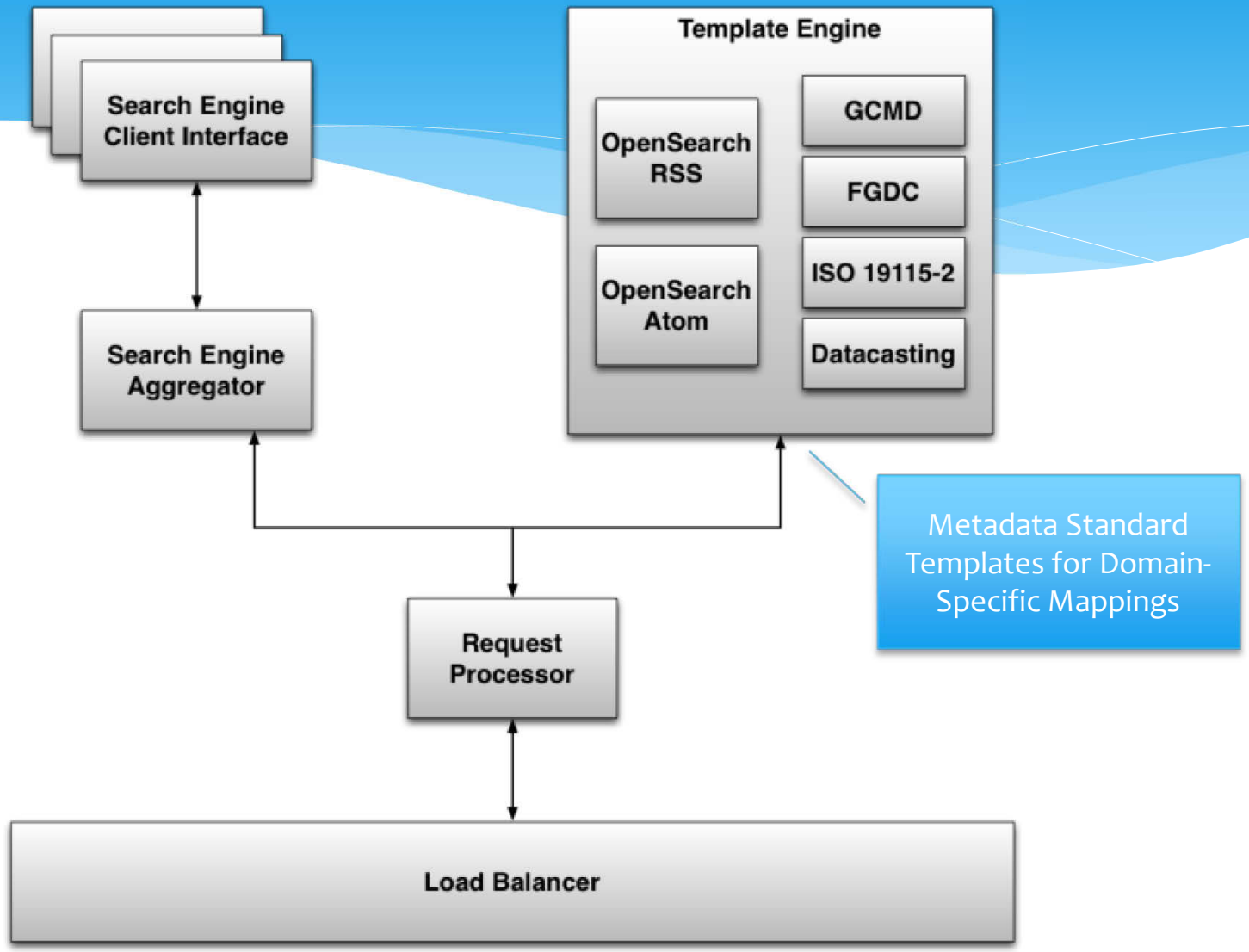


PO.DAAC User Communities

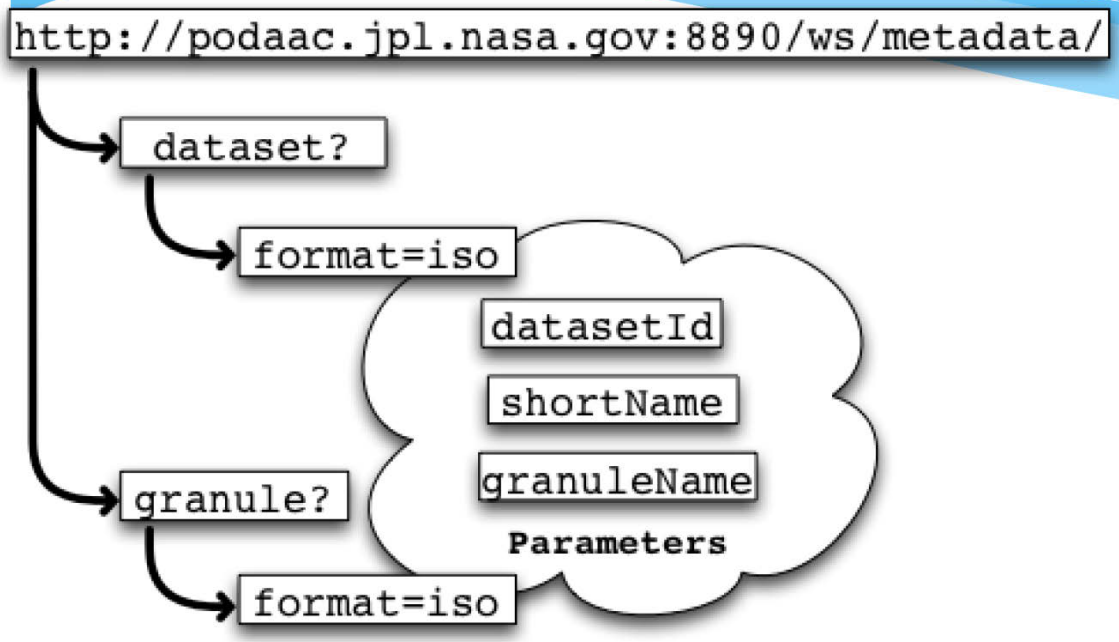
What We Provide

* Search and Discovery
  * OpenSearch and Solr/Lucene
  * FTP crawling
* Metadata packaging - our users speak different (metadata) languages
  * FGDC
  * GCMD
  * ECHO
  * ISO 19115-2
  * and many more…
* We guarantee to deliver the most up-to-date information available

METADATA TRANSLATION ARCHITECTURE



ISO 19115-2

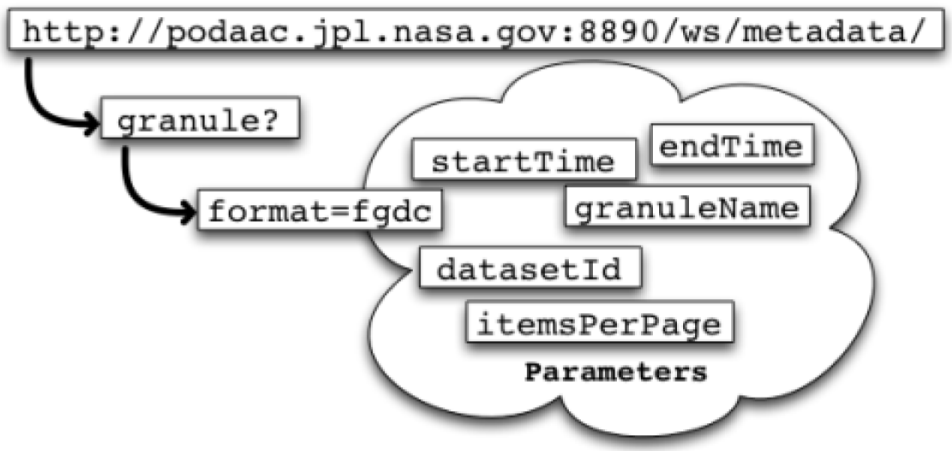


* Reference the GHRSST GDS 2.0 Specification
* Supports both Dataset and Granule metadata

Terminal

```
% curl -X GET \
? "http://podaac.jpl.nasa.gov:8890/ws/metadata/dataset?format=iso&shortName=OSDPD-L2P-MSG02"
```

Federal Geographic Data Committee



- Speedup of FGDC record generation
  - For 300+ granules, record generation went from taking hours to seconds
- Use of template engine supports reuse
  - Generate FGDC record for all datasets, not just GHRSST datasets
- Retrieve FGDC record as needed via web service call
  - No longer require pre-generate FGDC record files

Terminal

```
% curl -X GET \
? "http://podaac.jpl.nasa.gov:8890/ws/metadata/granule?format=fgdc&
shortName=JPL-L2P-MODIS_A&
startTime=2011-02-01T00:00:00Z&endTime=2011-02-01T23:59:59Z&itemsPerPage=500"
```

OCSI Summary

* There will always be
  * Metadata standards or recommendations
  * Different (may be better) ways to look for data
* Why PO.DAAC decides to invest in OCSI?
  * No need to redo the plumbing for each new standard
  * Allows us to focus on the domain – metadata standard and metadata resources
* Users expect more intelligent search results
  * Semantic Web has a lot to offer

Conclusions (1)

- ISO 19115 and its variants are powerful new ways for capturing critical metadata for satellite products. Goes beyond just data discovery. Key features:
  - Quality reports
  - Lineage and provenance documentation
  - Sensor/Instrument descriptions
  - Aggregation
- International standardization: my data system can talk to your data system

Conclusions (2)

* Disadvantages:
  * Pitfalls of the early adopters
  * Steep learning curve. Complexity is high.
  * Many ways of expressing the same thing
  * Not many resources yet, or tools
  * Expensive documentation !
* But organizations including NASA are gravitating to ISO 19115. Core requirement for decadal survey missions.

Conclusions (3)

* Much more capabilities are possible with metadata
  * Semantics is the ultimate objective
* Emerging standards require early adopters to implement and demonstrate value
* Who mandates standards?
  * Can voluntary standards work?