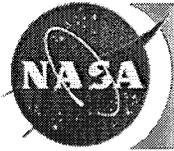


# **Managing Global Satellite Data: The GHRSSST-PP**

**Edward Armstrong, Jorge Vazquez and  
Andrew Bingham**

**JPL PO.DAAC**

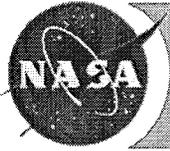
**California Institute of Technology**



# *The Metadata Perspective*

**JPL** Physical  
Oceanography DAAC

- Data organization, access, data discovery
  - GHRSSST-PP will produce ~100s daily L2P files
  - Files may reside in different locations -locally or on L2P server
  - Files maybe accessible in a variety of methods (e.g., FTP, OpenDAP etc.)
  - In situ matchup information necessary for the SSES generation
- Historical continuity; end of life-cycle transition



# *Data Discovery*

**JPL** Physical  
Oceanography DAAC

- Metadata should be searchable with user determined queries
  - Search of temporal and spatial boundaries
  - File locations and access methods
  - Points of contact
  - Accept automated requests ?
  - Subscriptions ?



# *MMR and MDB Review*

**JPL** Physical  
Oceanography DAAC

- Complete MMR and MDB structure reviewed in Feb 2004
  - Adheres to the NASA DIF format
  - Attributes streamlined
- XML formatting of metadata records
  - Document Type Definition (DTD) has been developed – formal description of XML



# *MMR metadata*

**JPL** Physical  
Oceanography DAAC

## Two types of metadata records:

### • Data Set Record:

- Information common to all files (i.e., same sensor)
- Project office will assist with preparing these

### • File Record:

- File specific information (e.g., parameter, location etc.)
- Prepared on a file-by-file basis by each RDAC

### • These will be linked by a common field (Entry\_ID)



# *File Record (FR) metadata*

**JPL** Physical  
Oceanography DAAC

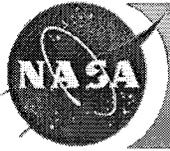
Field Name	Required	Type	Max length (bytes)
Entry_ID	yes	char	160
<i>Group: Temporal_Coverage</i>			
Start_Date	yes	char	10
Start_Time	yes	char	8
Stop_Date	yes	char	10
Stop_Time	yes	char	8
<i>Group: Spatial_Coverage</i>			
<i>Group: Related_URL</i>			
<i>Group: Personnel</i>			
<i>Group: Metadata_History</i>			



# *DSD metadata*

**JPL** Physical  
Oceanography DAAC

Field Name	Required	Type	Max length (bytes)
Entry_ID	yes	char	160
Entry_Title	yes	char	160
<i>Group: Data_Set_Citation</i>		<i>Group: Parameters</i>	
<i>Group: Temporal_Coverage</i>		<i>Group: Spatial_Coverage</i>	
<i>Group: Location</i>		<i>Group: Projection_Information</i>	
<i>Group: Data_Resolution</i>		<i>Group: Access_Constraints</i>	
<i>Group: Data_Center</i>		<i>Group: Related_URL</i>	
<i>Group: Reference</i>		<i>Group: Personnel</i>	
<i>Group: Metadata_History</i>			



# *MMR Database description*

**JPL** Physical  
Oceanography DAAC

- **Mysql database (version 4.0.17)**
  - Open source
  - Handles up to 4 Gb/table – millions of records per tables
  - Supports database transactions
  - Numerous APIs available (including Perl)
  - Relational
- **Installed on [coda.jpl.nasa.gov](http://coda.jpl.nasa.gov) (Linux server)**

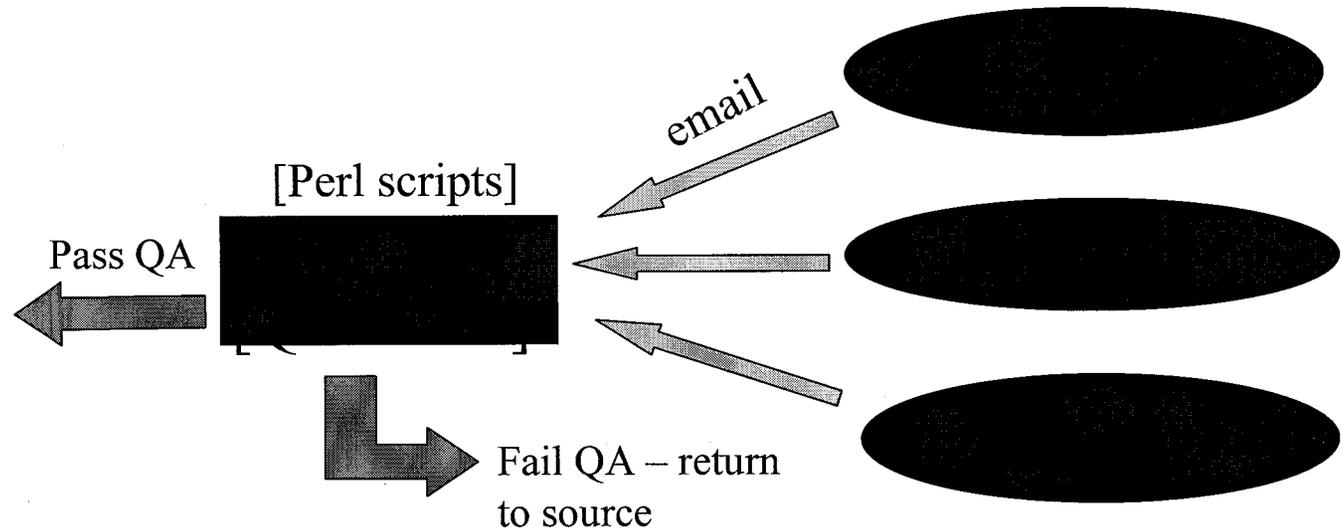


# *First MMR delivery model*

**JPL** Physical  
Oceanography DAAC

**metadata repository**

Create/update mySQL  
tables [SQL]

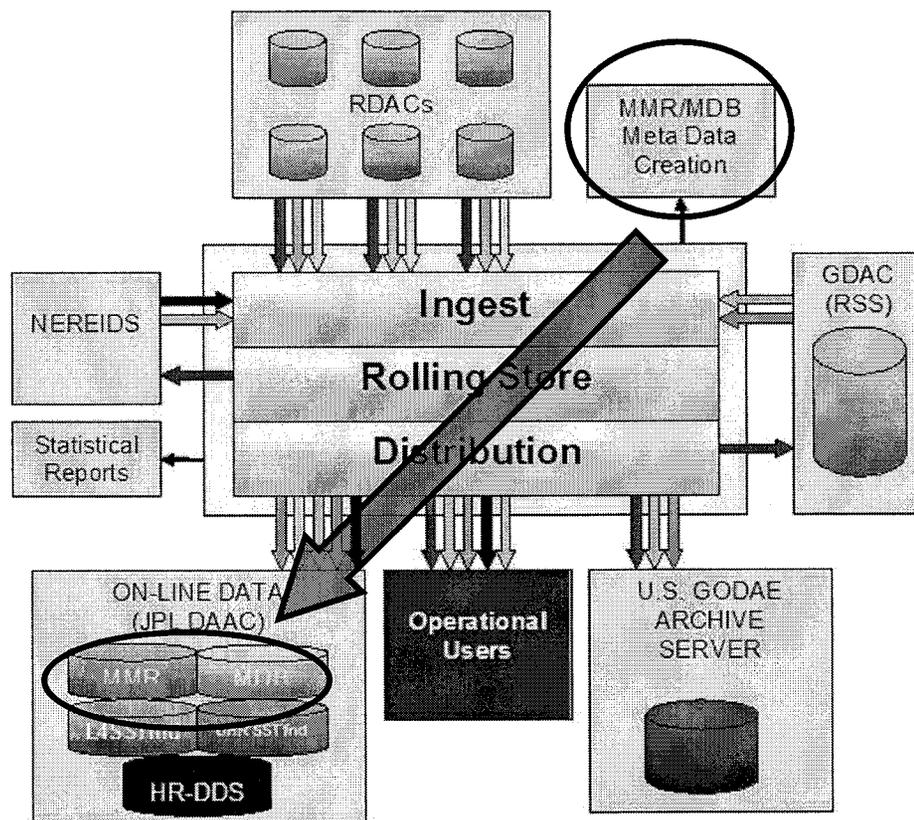


.... but email may result in security/access  
issues....



# Integration with OCEANIDS

JPL Physical Oceanography DAAC



→	L2P	→	HR-DDS	→	L4SSTfnd
→	MMR	→	Browse Map-Images	→	UHRSSSTfnd
→	MDB	□	OCEANIDS	→	OCEANIDS Trigger



# *MMR End-to-end test*

**JPL** Physical  
Oceanography DAAC

- Complete file record metadata created for MODIS L2 end-to-end test
  - Automated creation
  - > 400 FR metadata files ingested without error or user intervention over 3 days
  - Email driven (no Oceanids)
- Perl program parses the XML formatted metadata and check validity of each field
  - Email sent to provider with list of errors if improperly formatted metadata is detected
  - Perl API allows direct connection to database
- Database “transaction” methodology allows for handling of multiple simultaneous metadata requests.



# *Metadata creation tools*

**JPL** Physical  
Oceanography DAAC

- FR automation
  - Unique temporal/spatial information contained in individual L2P file; only need to extract
  - Other parts of FR metadata are “static”
- Perl approach
  - NetCDF Perl module →  
use NetCDF;  
\$nid = NetCDF::open( \$L2P\_file, NetCDF::NOWRITE );  
NetCDF::attget( \$nid, NetCDF::GLOBAL, “start\_date”,  
  \$start\_date );
- Perl approach I took uses a combination of NetCDF reads and reading from a static template
- DSD metadata will probably only need onetime creation and periodic maintenance



# Database query

JPL Physical  
Oceanography DAAC

- SQL spatial query for eastern equatorial Pacific (for Modis Aqua)

```
seaeddy: /home/led
seaeddy: /home/led
mysql> select record_number, s_lat, n_lat, w_long, e_long from fr_spatial_coverage
-> where n_lat <= 20 AND s_lat >= -20 and w_long >= -150 AND e_long <= -50;
+-----+-----+-----+-----+-----+
| record_number | s_lat | n_lat | w_long | e_long |
+-----+-----+-----+-----+-----+
|          11  | -10.417 | 10.496 | -80.859 | -56.111 |
|          36  |  -8.632 | 12.282 | -119.923 | -95.174 |
|          44  | -12.626 |  8.288 |  -94.244 | -69.466 |
|          84  | -14.623 |  6.382 |  167.315 | -167.835 |
|         147  |  -6.424 | 14.494 | -106.576 | -81.783 |
|         156  |  -2.438 | 18.495 | -132.368 | -107.397 |
+-----+-----+-----+-----+-----+
6 rows in set (0.00 sec)

mysql> 
```

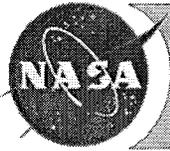
Record\_number can in turn be used to extract other file information from other tables



# *User Access to database*

**JPL** Physical  
Oceanography DAAC

- Front-end to database requirements
  - Accepts web-based searches and queries
  - Outputs to HTML or email
  - Others?
  - We can output total metadata records in a variety of formats (e.g., DIF or FGDC)
- Most likely candidate is PHP
  - Server-side scripting language
  - Returns simple HTML (or XML)
- Still on the "to do" list



# *Description of MDB*

**JPL** Physical  
Oceanography DAAC

- In situ records for satellite matchups
  - Primary source for SSES generation
- MDB records have own unique XML DTD and require their own ingest parsers
  - MDB records should not be aggregated into multiple files
- Clone of MMR
  - Separate database
  - Stores point data on 5x5 grids (or others)
  - Has not been yet been implemented – FY 05 activity
- Discussion:
  - Access for SSES (similar front-end as MMR?)
  - Subscriptions
  - Validation data withheld



# Summary

**JPL** Physical  
Oceanography DAAC

## • MMR

- MMR constructed and installed on server
- Ingest programs written
- Successfully completed 3 day end-to-end test
- To do: web interface; testing with individual RDACs

## • MDB

- First data access to in situ/matchup observations
- Only conceptualized but essentially a clone of MMR