Draft

Draft

Web-based Altimeter Service
Annual Review

ACCESS 07
Phil Callahan (334), PI
Brian Wilson (335), Co-I
Rob Raskin (388), Co-I

June 30, 2009

Annual Review Outline

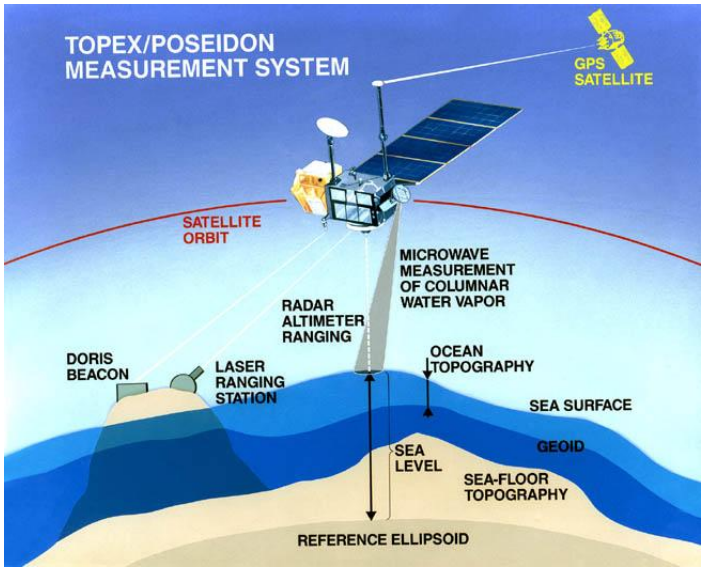
- Review Quad Chart
- Objectives
- Altimetry Background
- SciFlo Background
- Accomplishments
  - Altimetry Data
  - Computer System
  - Web Implementation
- Plans
  - Near-term Tasks
  - Long-term Tasks
  - PO.DAAC Transition
- Budget

Web-based Altimeter Service

Phil Callahan, JPL

Objective

- Develop a web-based tool for subsetting and updating altimeter data
  - Altimeter data consist of several specialized "components" that are updated by different groups at irregular intervals.
  - Specialized data exist for localized areas.
- Work with providers to get tool access to data and models as they are updated.
  - Locate specialized models for coastal areas.



Altimeter Sea Surface Height Measurement Components for Updating:

- Orbit
- Tides
- Radiometer
- Atmospheric – range, inverse barometer
- Geoid
- Range processing, corrections

Approach

- Build on SciFlo system for user interface, data access, algorithm control
- Modularize Geophysical Data Record (GDR) update algorithms to provide processing functionality
- Integrate SciFlo and GDR algorithms
- Develop data subsetting (localization)
- Work with scientists, data centers to get access to models and data sets

Co-Is

- **Rob Raskin (JPL), Brian Wilson (JPL)**

Key Milestones (2 year task)

• Develop local processing, modular GDR algorithms	03/09
• Integrate SciFlo and modular algorithms	04-08/09
• Develop data subset capabilities	04/09
• Develop user-specified output formatting	08/09
• Develop collection of external data and algorithms	03-12/09
• Set up web presence and beta test with selected users	12/09
• Revise service based on tests and work with PO.DAAC	04/10

TRL_in = 5-6; TRL_out = 7-8

Prototype Demo in June '09

Altimetry Background

- Series of successful science missions from TOPEX/POSEIDON through Jason-1, 2; Jason-3 being planned
- Large international community uses data for oceanographic studies (original intent) and increasingly other applications, particularly coastal and inland water
- Extreme accuracy (~1 cm absolute, << 1 mm/yr change) is key to success, particularly for climate change applications
- Measurement concept is simple but requires numerous corrections, “components” to reach required accuracy
- Accuracy is ensured, improved by
  - New orbits
  - Calibration, correction of errors, drifts in components
  - New components for new applications
- New uses may also require new, specific (re)processing

SciFlo Background

- SciFlo Distributed Dataflow System
  - Large-scale, loosely-coupled, distributed computing using Web (Simple Object Access Protocol - SOAP) and Grid services (on-demand virtualization)
    - Specify a processing stream as an XML document
  - Leverage Web Services standards, open source
  - Dataflow engine for automated execution and load balancing – parallel, asynchronous processes
- Automate large-scale, multi-instrument science processing by *authoring* a dataflow document that specifies a *tree* of *executable* operators.
  - iEarth Visual Authoring Tool (VizFlow)
  - Distributed Dataflow Execution Engine
  - Move data "granules" to the operators using FTP, HTTP, or OpenDAP URLs.
  - Move operators (executables) to the data.
  - Built-in reusable operators provided for many tasks such as subsetting, co-registration, regridding, data fusion, etc.
  - Custom operators easily plugged in by scientists.
  - Leverage convergence of Web Services (SOAP) with Grid Services

Functional Capabilities

- Provide system to allow users to access and combine various parts of altimeter GDRs with new components (data, corrections, models) on demand.
  - Projects generate fundamental data record from telemetry with time, orbit, instrument information and basic corrections
  - Producers of other "components" of the altimeter record register with Altimeter Service
    - Components can be data or data+operator(s)
    - Producers need to provide some documentation to guide use

- Use Scenario

➢ **Login** to the Altimeter Service portal
➢ **Select** a **time** range (or orbit cycle range) and a latitude/longitude **region** (or the globe)
➢ **Select** the desired choices of **components** from a menu, some of which are remote data or model components that have been registered with the Altimeter Service
➢ **Register** their own **component** corrections and models by providing a few items of information and URL's pointing to the data files
➢ **Reprocess** (update) altimeter GDR data on demand, using the selected components
➢ **Retrieve** the updated GDR containing the parameters of interest in useful formats (netCDF, perhaps GDR binary, ASCII for small requests)
➢ **Visualize** the key parameters SSH and SSHA on a map or via Google Earth or as a time series
➢ **Compare** SSH or SSHA from their GDR update to prior runs or known GDR versions

Some Usage Scenarios

- Generate basic altimeter data records: Project produces initial IGDR; all subsequent updates done with service – add POD, improved atmospheric models (observed replaces predict), any other improvements available on ~1 month time scale

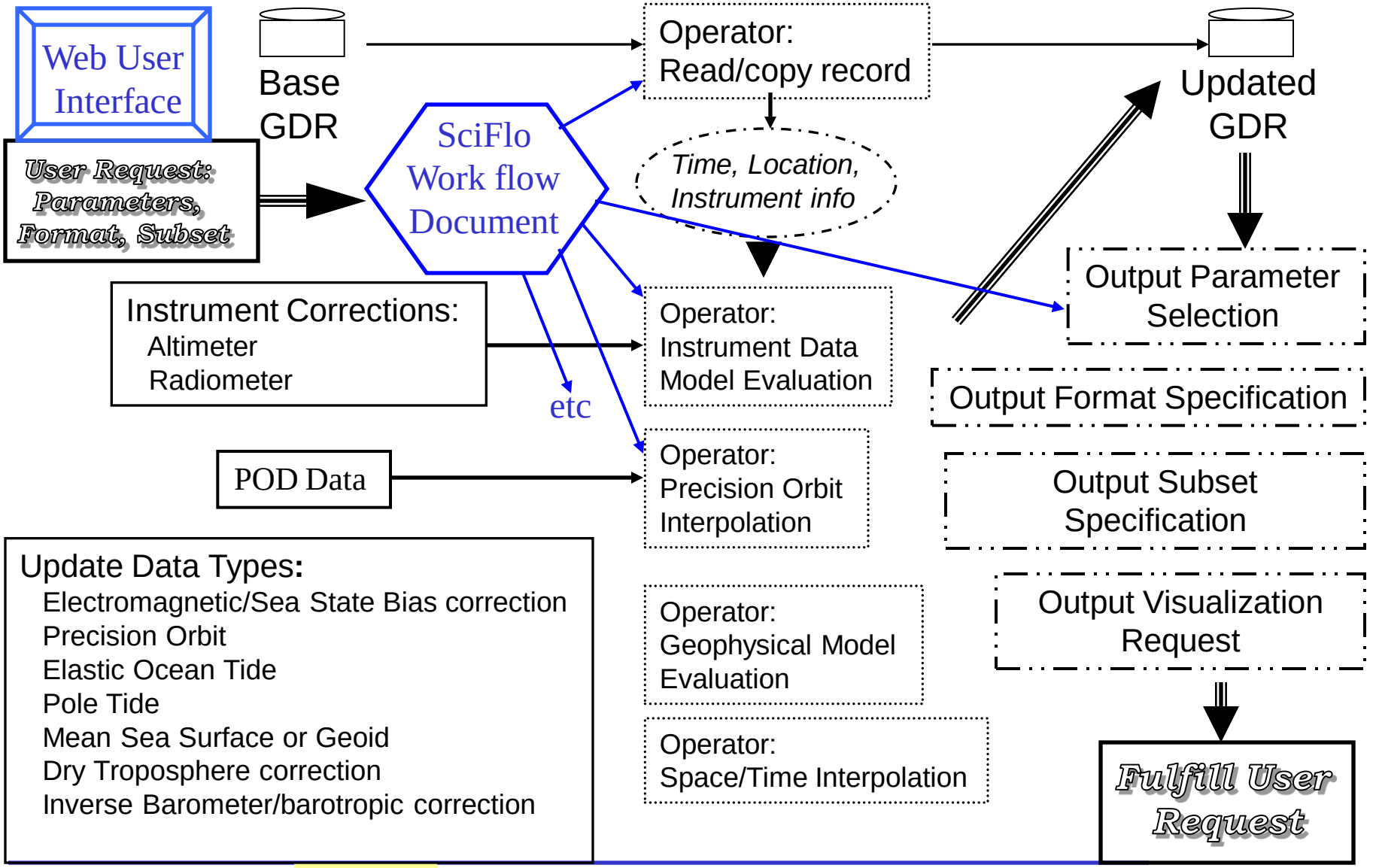
- Update existing data with improved components: Particularly revised orbits, improved tides, radiometer calibrations, barotropic corrections, geoid, mean sea surface
  - Faster, independent of Project update cycle
  - Test versions of components

- Produce regional/coastal products: Select regional data, apply local tide models, radiometer corrections (processed to remove land effects), local barotropic models
  - Special retracking could be linked to original points

- Produce storm products: Time/space subsetting, special tropo models

Altimeter GDR Update Process Flow

Web-based Altimeter Service



Development Approach

- Create web user interface to select data range, which components to update, output format
- Connect user interface to SciFlo system to generate work flows to carry out user requests: data access, algorithm control, output generation
- Modularize GDR and other algorithms to provide update, processing functionality
  - Algorithms for altimeter components such as orbit, tides, sea state bias, atmospheric corrections
- Integrate SciFlo and GDR algorithms
  - "Register" external models and data sets
- Develop data subsetting (space/time localization)
  - Key short term application to coastal applications
- Provide output formatting, mainly netCDF, parameter selection
- Work with scientists, data centers to get access to models and data sets
  - Key to long term utility is enlisting providers of new, improved components
  - Coordinate with PODAAC, AVISO, other data providers on International Altimeter Service

Binding Algorithms into Python

- Reuse existing algorithms
  - Modularize codes into callable Fortran or C subroutines
  - No globals: All file names and config. parameters are passed in arguments
- Bind algorithms into python
  - Use f2py or SWIG to generate glue code, or write by hand
  - Compiled into shared object library
  - Library dynamically linked into python using 'import'
  - Precise Orbit Estimation (Fortran algorithms):
    - from poe import poe_interp_pt, poe_interp_highrate_pt
- Service Workflows
  - SciFlo document invokes python routines as processing steps
  - Algorithms can also be published as Web Services
  - Regional Tide Model: could call remote service exposed at university

Accomplishments (1 of 2)

- Computer System
  - Acquired and configured Sun dual processor, 1TB system
  - Configured identical system acquired for TOPEX Retracking in same way to act as a development machine later in project
- Software
  - Installed SciFlo and all required software support packages
  - Modularized Precision Orbit interpolation software for use in system
  - Modularized tide model (GOT4.7)
  - Wrote preliminary version of software to convert TOPEX RGDR data to netCDF (needs additional CF-compliant names)
  - Developed method for simple spatial subsetting for repeating orbit

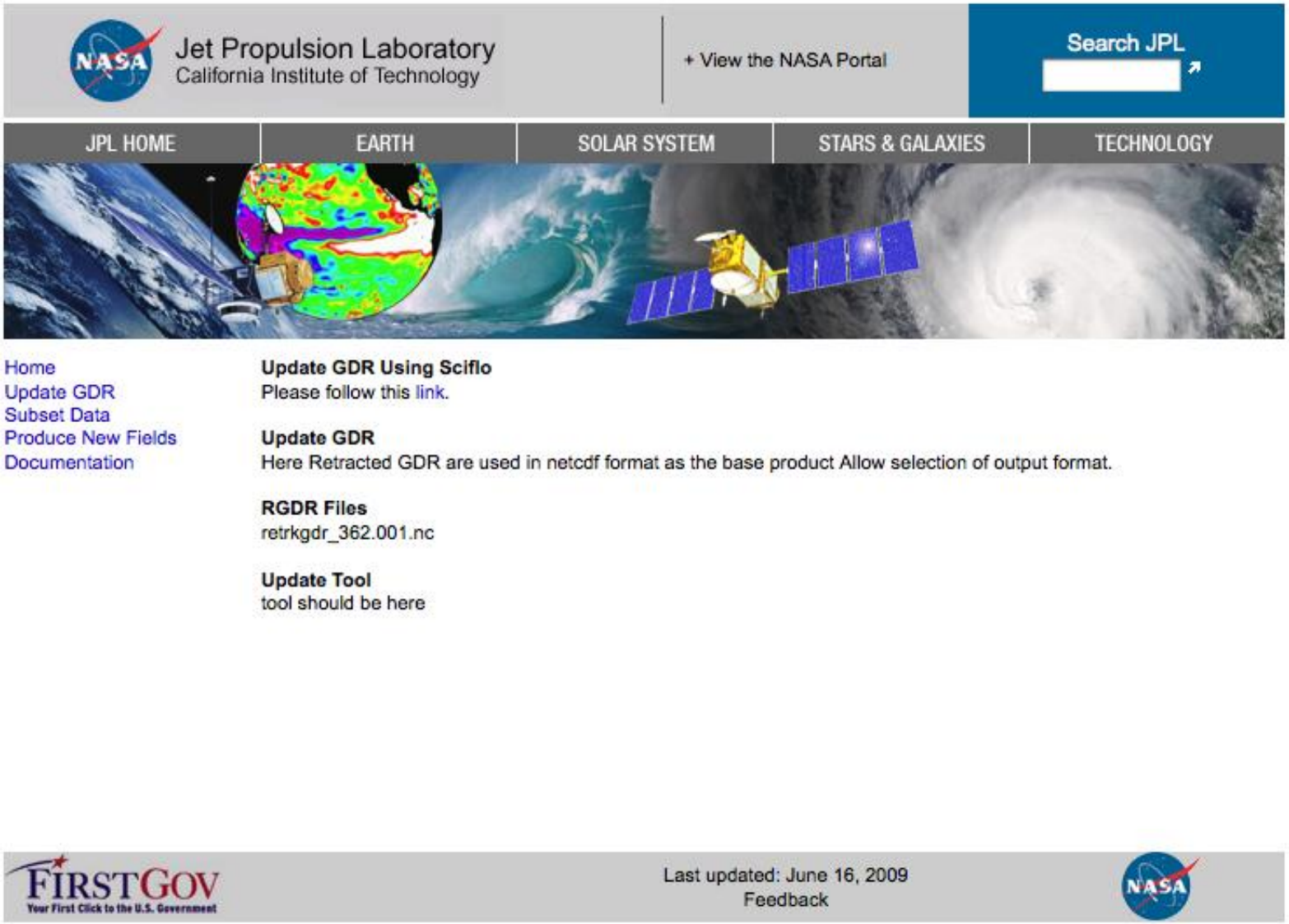
Accomplishments (2 of 2)

- Web Implementation
  - Created beginning of web user interface (see examples)
  - User subsetting by time (spatial subsetting not connected yet)

- Altimetry Data
  - Acquired all TOPEX Retracked RGDRs as base data
  - Converted data to initial version of netCDF
- Additional Data sources contacted or currently available
  - Updated radiometer wet tropo computation, specifically for coastal areas
  - Updates of orbits from GSFC: ftp://dirac.gsfc.nasa.gov
  - AVISO (French data center): new version high resolution of Dynamic Atmospheric Correction (DAC) in netCDF available by ftp: ftp://ftp.cls.fr/pub/oceano/AVISO/auxiliary/dac/
  - Baltic, North Sea models from Danish Meteorological Institute
  - Coastal altimetry projects: PISTACH (CNES, CLS), ALTICORE (ESA, several agency consortium)
  - Tide Gauge calibration check (G. Mitchum, U. South Florida)

First User Screen

Web-based Altimeter Service



Second User Screen – Data Selection

Web-based Altimeter Service

AWS

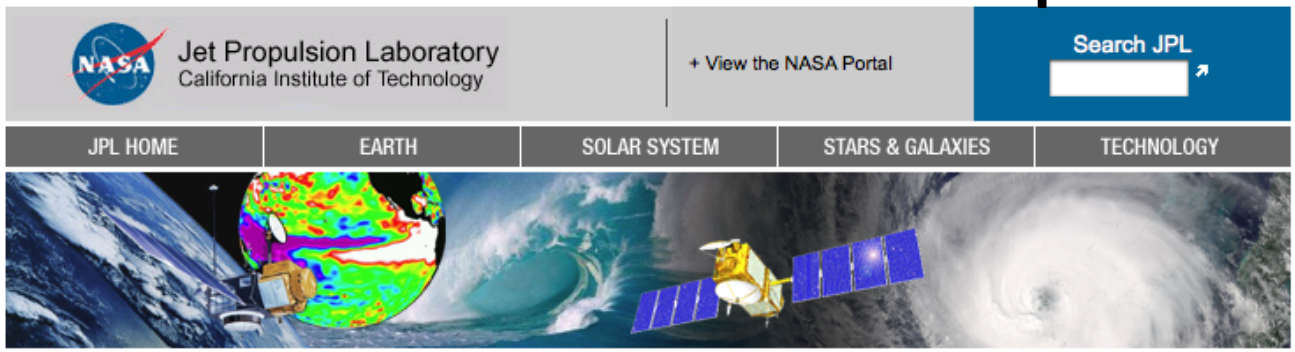
Altimetry Web Service

Sciflo Inputs

cycleNumber:	362
pathNumber:	123
startTime:	2001-01-27 00:00:00
endTime:	2001-01-27 23:59:59

submitSciflo | submit w/ no cache

Third User Screen – SciFlo Output



AWS

`Altimetry Web Service`

Sciflo Inputs

tag	value
cycleNumber	362
pathNumber	123
startTime	2001-01-27 00:00:00
endTime	2001-01-27 23:59:59

Execution monitoring:

Work Unit Status/Color Legend: waiting, ready, staging, working, finalizing, done, cached, exception, cancelled, paused

Status of this sciflo is: done

Execution time:
1.185

Work Unit Monitoring

index	procId	type	dependencies	status	execution time	results
0	do_poe	python function		cached	0.000	results: xml

Work Unit Status/Color Legend: waiting, ready, staging, working, finalizing, done, cached, exception, cancelled, paused

Execution log: log

Sciflo Outputs

tag	value
outPath	str [download]

Plans

- Near-term Tasks
  - Complete regional subsetting function
  - Develop netCDF element subsetting
  - Improve CF compliance of netCDF format data. Rerun RGDR ➔ netCDF
  - Produce documentation, help for users
  - Expose capability with time, space, netCDF subsetting and orbit, tide updating to selected members of OSTST for testing (Oct ’09)
  - Present posters on Altimeter Service at Third Coastal Altimetry Workshop, OceanObs09, both in mid-Sept in Italy
- Long-term Tasks
  - Refine user interface for subsetting
  - Improve notification of completion
  - Develop caching, advertising of datasets done by users
  - Add capabilities for interpolating along-track corrections
  - Acquire additional models, data sets
  - Transition to PO.DAAC

Plans – PO.DAAC Transition

- Made presentations at PO.DAAC Users Working Group, Coastal Altimetry Workshop in Oct '08. Received positive feedback
- Presented AltiServ demo to PO.DAAC in June 2009
- Working with Thomas Huang (Infrastructure Lead), Charles Thomas (Tools and Services Lead) and Andy Bingham (Manager) to ensure smooth integration of AltiServ into PO.DAAC infrastructure
- Will make presentation at next PO.DAAC Users Working Group, to elicit feedback on progress, features
- Hardware and server will be directly turned over to PO.DAAC.
  - Hardware currently sits in PO.DAAC subnet with other PO.DAAC equipment.
- User interface will be migrated to PO.DAAC web site on 4/15/10
- Documentation and testing to be carried out during early 2010, and completed 4/15/10

Budget

- Progress has been slower than planned because of staff (un)availability
- Spending is in line with progress
- Spending
  - Apr – Sep ’08: $34k (original plan: $149k)
  - Oct – May ’09: $253k (plan of Jan ’09: $313k)
  - Jun – Sep ’09 plan: $155k (Total FY’09 from Jan’09 plan: $468k)
  - Spent to date: $287k
  - Total Request: $790k
  - Remaining to spend: $503k
  - Expected balance Oct 1, 2009: ~$350k (about 9 months at typical spending rate)

Backup Material

Details

SciFlo Example

Web-based Altimeter Service

GPS & AIRS Level-2 Space/Time Matchup