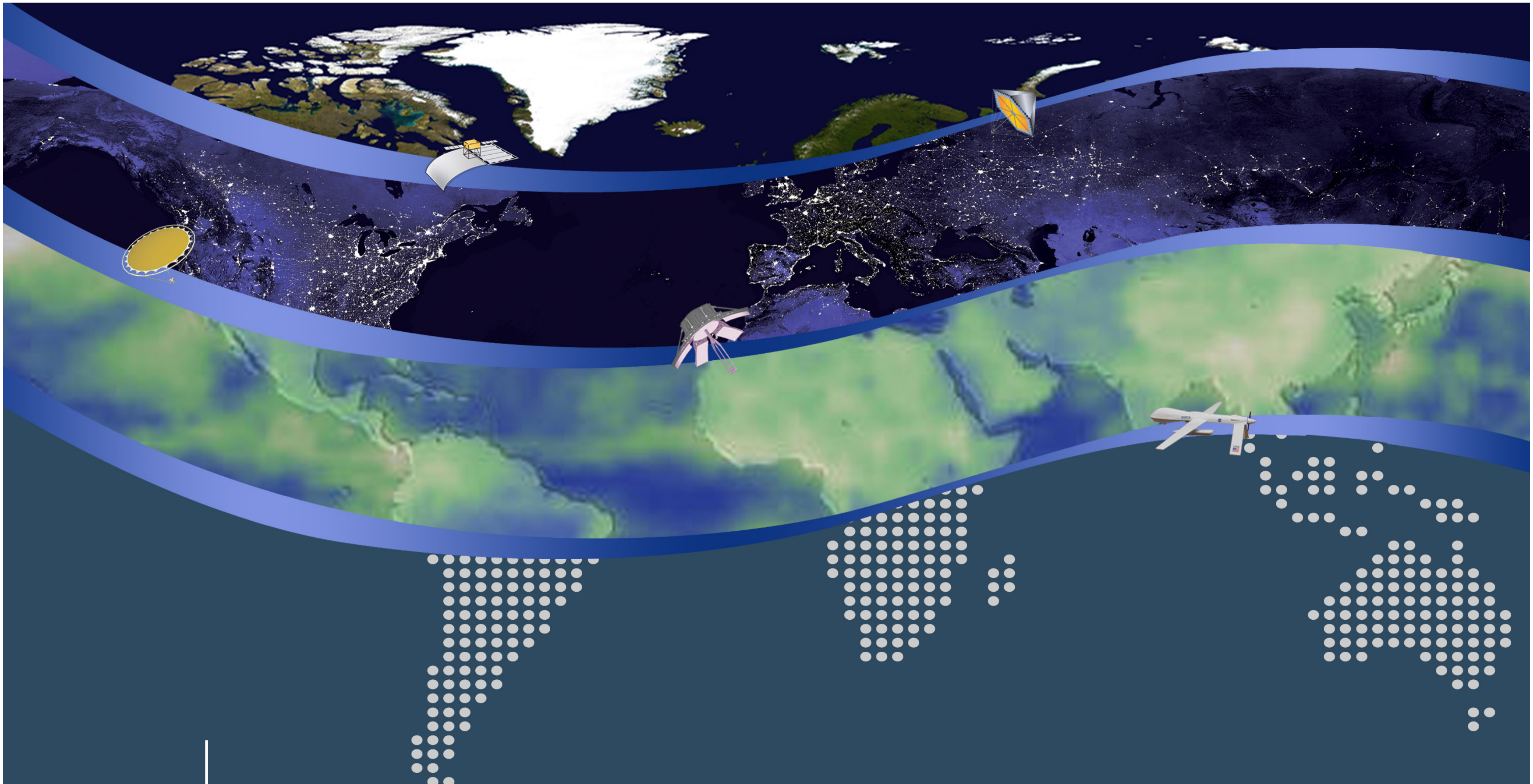
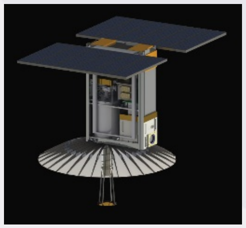


NASA AIST Projects: Cloud Computing

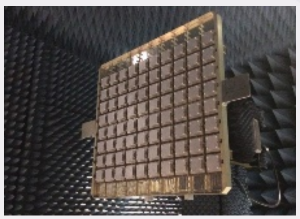
Michael.M.Little@nasa.gov

Earth Science Technology Program

ESTO manages, on average, 120 active technology development projects. Most are funded through the primary program lines below. Nearly 800 projects have completed since 1998.



In-Space Validation of Earth Science Technologies (*InVEST*)
on-orbit technology validation and risk reduction for small instruments and instrument systems that could not otherwise be fully tested on the ground or airborne systems
(average award: $1-1.8M per year over three years)



Advanced Component Technologies (*ACT*)
critical components and subsystems for advanced instruments and observing systems
(average award: $300K per year over two/three years)

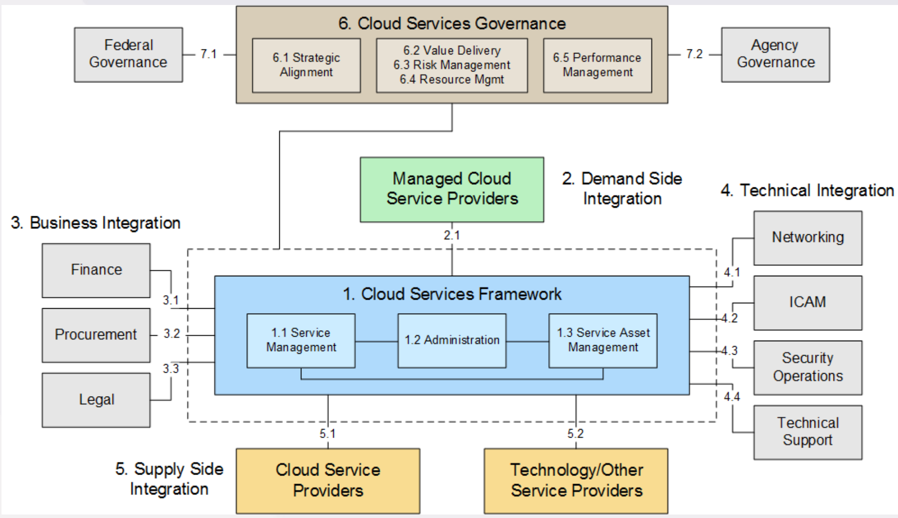
Instrument Incubator Program (*IIP*)
innovative remote sensing instrument development from concept through breadboard and demonstration *(average award: $1.5M per year over three years)*

Advanced Information Systems Technology (*AIST*)
innovative on-orbit and ground capabilities for communication, processing, and management of remotely sensed data and the efficient generation of data products
(average award: $750K per year over two years)

What's the Role of AIST?

- The objectives of the AIST Program are to research, develop, and demonstrate advanced information system technologies that:
  - Reduce the risk, cost, size, and development time for Earth science space-based and ground-based information systems,
  - Increase the accessibility and utility of science data, and
  - Enable new observations and information products.
- The AIST Program Focuses on three Technology Areas:
  - Operations Technologies – broadly support the future challenges of operating NASA's Earth science space-based, airborne or ground-based systems
  - Computational Technologies – operate directly on Earth Science data produced by sensors (real or simulated) to improve/enhance information extracted from the data stream or model outputs, measurements to be acquired by a new mission or science campaign, or researchers' tools for analytics (including development of quantum annealing algorithms)
  - Data-Centric Technologies – broadly support the science and applications communities in conducting the sequence of activities needed to transform Earth science observational data to improve information re-use, facilitate collaboration within the research community, and increase the speed with which results are produced and published

NASA's Enterprise Approach



Cloud Computing at NASA is provided through the Computing Services Program in the Office of the NASA CIO

Key Elements of an Enterprise Approach

- Focus on consumption of commercial Cloud Services instead of building Agency Clouds
  - Expend limited resources on highest risk
- Standardized Agency governance
- Standards and guidance for technical integration with Agency infrastructure, processes, services
  - Networking
  - Security operations
  - Authentication services
- Integrated hierarchical approach to FedRAMP compliance
- Common procurement vehicles with proper terms, conditions, best practices
- Payment system to facilitate “pay as you go” within Agency constraints
- Integration with Agency IT service catalog and help desk

What Problem is AIST Solving

JPL Experience

The best way to learn Cloud Computing is to use it

- AIST is working to overcome inhibitors to Cloud use in Science
  - Lack of experience with Cloud computer security by almost everyone
  - ATO process is labor intensive, expensive and requires experience
    - NASA authorization to use Public Cloud Computing has focused on Moderate
  - Maze of procurement restrictions on acquisitions and data centers
    - Unable to share compute work space with International collaborators
    - Network connectivity restrictions overkill for most science research
  - Difficult to understand the Technology without a way to try it
    - Working out the Business Model by NASA impeded early adopters
    - Agency Policy precluded Users from buying AWS computer time on P-cards
    - Lack of practical assistance in getting into AWS or Azure
    - These have been largely corrected by the diligence and foresight by Ray O'Brien
- AIST is demonstrating AWS value to Science PIs

AIST Funds Three Major Cloud Efforts

- AIST Managed Cloud Environment (AMCE)
  – Collaboration with CIO's
  – Enable science PI's to leverage NGAP investments in data access
- Technology developments to enhance Cloud Computing usage
  – SAR Science Data Processing Foundry
    - Enable use of JPL's ISCE and ARIA Projects by SAR PI's
  – Object Store-Based Data Service for Earth System Science
    - Collaboration with EOS-DIS – ACCESS16-0031 (PI: Readey, HDF Group)
- ATO Clearance for Arc-GIS Online Portal (Esri)
  – Software-as-a-Service for GIS tools
  – AIST collaborating with GSFC CIO overseen by NASA OCIO
    - ATT/ATO for Agency Use to solve Computer Security
  – Business model managed by CIO's Enterprise License Team @ NSSC
    - Usage Credits provided under Enterprise License Agreement

AMCE: Overview

Science Team Management and Use of AWS

- Ensure PI-led Teams have access to essential AWS capabilities
  - Fast-track AWS capabilities as soon as FEDRAMP approved
- Avoid Financial Management and Procurement Problems
  - Goal: As much compute as you need within 24 hours of initial request
  - Pay as you go
  - Ensure resource consumption is properly tagged (automatically)
  - Multi-level monitoring of planned vs. actual consumption
- Minimize Computer Security Problems
  - Provide an Authorization to Operate for public information with minimum restriction
  - Restrictions are clearly defined to the Users
- Minimize management overhead costs through automation

AMCE: Implementation

- Objective:
  - Supply AWS cloud resources to decrease hardware/compute costs distributed at PI's business location(s)
  - Demonstrate a model for Science users to leverage AWS accelerate science, advance flexibility of team work, reduce computing costs.
- AIST Project Users of Compute Resources
  - Awardees receive an allocation of funding for AIST support cloud, based on their proposal estimates
    - NASA awardees
    - University awardees without NASA credentials
    - Industry (non-contractor) awardees without NASA credentials
    - Other Government Agency awardees without NASA credentials
    - Non-US citizens are expected to be on most teams
  - During initial phase, all SBU computing will be handled by PI on their own
  - HEC Program handles Awardees requiring supercomputing resources
- AIST Overhead – Practical support to derive cost experimentally
  - Cost of cloud management software (DC2, NESSUS, Splunk) and instances to run them
  - Computer Security instance to handle log segregation
  - Cloud SysAdmin, Cybersecurity and Business Management labor

AMCE: Concept of Operations

- Analogous to running an apartment building
  - Projects needing to be in a single family dwelling are given individual AWS accounts (same ATO, same SEWP funding)
- Fundamental Principles
  - Provide Project PI with maximum flexibility in determining what to use in AWS
  - Ensure that projects are protected from each other
    - Orientation includes how to avoid interfering with other projects
  - Monitor compliance with Computer Security restrictions
    - Easy common-sense rules and training
  - Ensure that projects do not over run their budget
    - Tools and training to self-monitor and alarm
    - Automate overrun notices: alert, warning and shutdown
    - PI, PI's privileged users, PI's Users
    - AIST Project Manager and AIST Technical Manager get notifications
    - AIST PM and PI's notified before account lockout

AMCE: Project Research to Ops Transition

- Reduce cost of refactoring to run in a NASA environment
  - Use accepted Operating Systems (OS), libraries, security measures
    - Pre-scanned
  - Central authority for reviewing/approving image updates
    - Schedule for de-authorizing environment images
  - Build in log distribution to CSO Instance to detect incidents
  - Appropriate forensic data for post-incident analysis
- Potential for avoiding re-installation when apply technology under Software as a Service Model
  - Integration with other Science assets, like NEX or NCCS
- Encourage use of source control, documentation tools
- Improve ability to conduct independent testing of projects
- GAP: Ability to accurately estimate and project timing of cloud costs
  - Contrast to classic ownership cost estimates
  - Some tools exist but are hard to use

AMCE Benefits

For the PI

- Enables quick and easy access to a lot of data
  - NGAP
- Supports elasticity of computing resources
  - 1 to 10,000 nodes in a few minutes
- Accelerates project research
  - Procurement lead time
- Facilitates cost tracking
- Facilitates technology infusion & transition to ops
  - Easy end-user implementation

For the Program

- Lowers program costs
- Accelerates Tech Infusion
  - Eases Demonstration/Test Drive
  - Software as a Service avoids re-installation by adopters
- Accelerates Project startup
- Encourages collaboration
- Demonstrates value of cloud to science community
- Accelerates integration into NASA software archive, data products, software

- Status:4 projects onramped, 5 in pipeline
  - 2 months of operation, ramping up slowly
  - 2 more interested non-AIST Projects

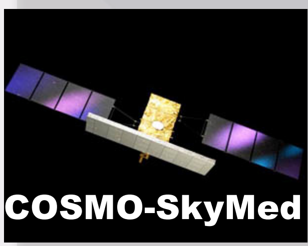
Backup

- SAR Science Data Processing Foundry
- ACCESS16 Storage Object Project

Technology Highlight

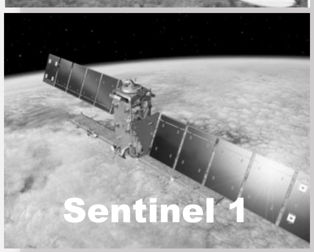
NASA SAR Science Data Processing Foundry Enables Science

L-Band
P-Band
(JPL)



X-Band
(Italy)

P-Band
(GSFC)



C-Band
(ESA)

X,C,S,L,P-Bands
(DLR)

PIs: Paul Rosen (ISCE) and Hook Hua (ARIA-MH)

The SAR SDP Foundry

ARIA
JPL-Caltech

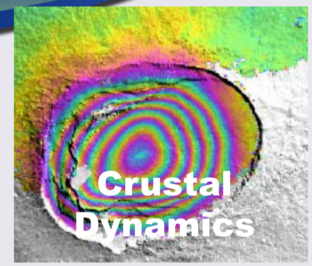
Advanced Rapid Imaging & Analysis for Monitoring Hazards

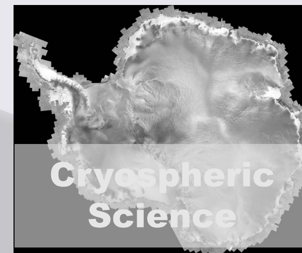
ISCE

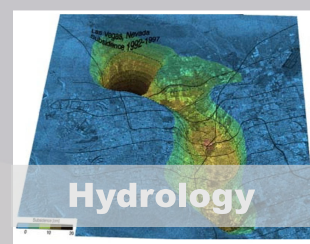
InSar Scientific Computing Environment on the Cloud

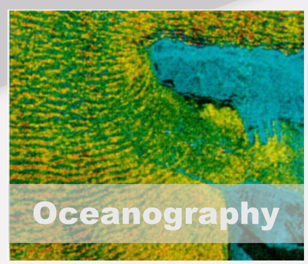
amazon web services

Bottom Right: EcoSAR acquired image over Andros Island, Bahamas in March 2014. Top 4 right: from InSAR Workshop Summary Report, 2004, Oxnard, CA.









SAR data from NASA and other sources can be on-ramped into the cloud-based processing data system, the **SAR Science Data Processing (SDP) Foundry.** The Foundry, consisting of ESTO projects ARIA science data system and ISCE, as enabled by a hybrid cloud computing environment, is being created to accept L-0 and L-1 SAR data to produce quality tags, metadata and data products defined and approved by Science Teams. Processing improvements will be shared across instruments and disciplines.

Once on-ramped, PI's can quickly process data they select, to the extent that they can afford it.

SAR SDP: Key Considerations

- Any **SAR mission** can produce thousands to millions of images
  - Orbital global mapping produces data continuously
  - Sub-orbital flight missions and campaigns support specific objective
- **Science Data Products** vary depending on system design, domain and intended purpose
  - Radar frequency (Band)
  - Scanning strategy (multi-pass, single pass, etc)
  - Platform operations artifacts (orbital vs. aircraft)
- **Science Data Processing** has some common characteristics
  - High volume of embarrassingly parallel processing jobs
  - Quality Assurance, metadata and registration of images
  - Cloud Computing offers scalable, if not affordable, solution
    - Prioritization for scheduling
    - Event Triggers
    - Low latency processing
  - Create metadata for provenance, geolocation, temporal, quality
- A big pool of money is easily targeted for budget reductions

SAR SDP: Foundry Concept

- Definition
  - A set of user-selectable components which are implemented in a scalable processing environment to leveraging a common framework for producing community-accepted Science Data Products from SAR instruments
  - Support multiple research and applied science communities
  - Community review/acceptance of processing model and subsequent improvements
  - Community defined science data products
- Components
  - Interface to Instruments which have been on-ramped
  - Production Processing Codes for Community defined Science Data Products which have been on-ramped
  - ISCE – Processing environment for instrument output
  - ARIA SDS – end-to-end SDS for SAR processing and data management
  - Hybrid Cloud – Provides scalable processing environment, including AWS
  - Foundry User Interface
    - Implements Business Model
    - Permits user selection of instrument, scenes, standard data products
  - EOS-DIS designated repository provides common destination for output products

SAR SDP: Foundry Benefits

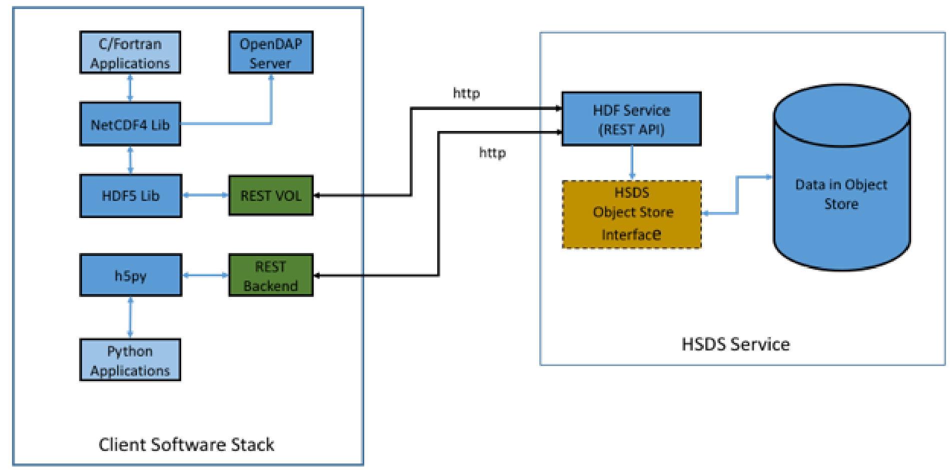
- Processing is under the control of the customer with data and funding
  - JPL can leverage their cloud interface
  - NASA can leverage OCIO SEWP Acquisition and simply use a WBS instead of a PR
  - Non-NASA collaborators, through agreement, can buy their own processing on AWS
- Hybrid cloud helps to keep cost as low as possible as the technology evolves
  - AWS Spot Pricing
  - Could move processing closer to data sources
  - Processing continues expensive, but could harvest idle local resources
- Processing environment is published and community-accepted
- Well defined processes for onramping instruments and data product specifications
  - Interface Control Documents publish requirements for L0 and L1 to permit processing
  - Instrument Team can account for high volume processing at initial product design
- Processing improvements are shared among the science communities
  - Example: Reliable use of Spot-pricing
- Science Data Products can become available to the communities regardless of who funded their production
  - Consistent with 2004 InSAR Working Group Workshop Summary Report (10/20/2004)
  - Can also deliver to an optional destination for immediate use

Object Store-Based Data Service (OSDS) for Earth System Science

PI: John Readey, The HDF Group

Objective

- Improve the performance of data delivery from cloud storage systems to cloud computing platforms
  - Reduce latency and data transfer volume
  - Reduce pre-processing to move data onto cloud processors
- Demonstrate functionality and performance of HDF5-based Object Store on Amazon Web Services (AWS)
  - Reduce effort to re-use existing DAAC assets
- Implement technology in OpenNEX under EOS-DIS funding



Object Store Data Service Architecture

Approach

- Produce science-based system design
  - Define science use cases
  - Define OSDS high level design and Core System design
  - Align HDF components to Object Store concepts
- Implement OSDS Core System on AWS and test
- Develop enhancements to Core System and test
- Release Beta Version to OpenNEX
- Develop Client Libraries to permit re-use of OpenNEX tools

Co-Is/Partners: P. Votava, A. Michaelis, ARC

Key Milestones

• Define Science use cases and OSDS design	08/16
• Complete Core System development and test	05/17
• Complete Core System integration into AWS	08/17
• Release Beta version of Core System	02/18
• Complete Client Library testing	06/18

TRL_{in} = 5