IN43B– 1734

The Synthetic Aperture Radar Science Data Processing Foundry Concept for Earth Sciences

Paul A. Rosen[1], Hook Hua[1], Charles Norton[1], Michael M. Little[2]
[1]Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109
[2]National Aeronautics and Space Administration, Earth Science Technology Office, Washington DC

Abstract

Since 2008, NASA's Earth Science Technology Office and the Advanced Information Systems Technology Program have invested in two technology evolutions to meet the needs of the community of scientists exploiting the rapidly growing database of international synthetic aperture radar (SAR) data. JPL, working with the science community, has developed the InSAR Scientific Computing Environment (ISCE), a next-generation interferometric SAR processing system that is designed to be flexible and extensible. ISCE currently supports many international space borne data sets but has been primarily focused on geodetic science and applications. A second evolutionary path, the Advanced Rapid Imaging and Analysis (ARIA) science data system, uses ISCE as its core science data processing engine and produces automated science and response products, quality assessments and metadata. The success of this two-front effort has been demonstrated in NASA's ability to respond to recent events with useful disaster support. JPL has enabled high-volume and low latency data production by the re-use of the hybrid cloud computing science data system (HySDS) that runs ARIA, leveraging on-premise cloud computing assets that are able to burst onto the Amazon Web Services (AWS) services as needed. Beyond geodetic applications, needs have emerged to process large volumes of time-series SAR data collected for estimation of biomass and its change, in such campaigns as the upcoming AfriSAR field campaign. ESTO is funding JPL to extend the ISCE-ARIA model to a "SAR Science Data Processing (SDP) Foundry" to on-ramp new data sources and to produce new science data products to meet the needs of science teams and, in general, science community members. An extension of the ISCE-ARIA model to support on-demand processing will permit PIs to leverage this Foundry to produce data products from accepted data sources when they need them. This paper will describe each of the elements of the SAR SDP Foundry and describe their integration into a new conceptual approach to enable more effective use of SAR instruments.

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 objectives
- 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 (typically) 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

SAR SDP Foundry Funding Model

- Major Ongoing Costs
  - SDP Processing
  - Repository of Data Products
- Instrument development team
  - Produces L0 and L1 data products conforming to Foundry interface
  - On-ramping of instrument by the Foundry team at JPL
    - Foundry creates configuration model, selectable configuration file, processing pattern
    - Acceptance by Instrument Team in conjunction with Appropriate Science Team
- Research or Applied Sciences Community Science Team defines, tests and accepts and funds data products
  - Funds JPL for Implementation in the Foundry
  - Funds Community acceptance tests of output products
- EOS-DIS funds Repository and Stewardship Functions
  - Presumably ASF Repository Functions
- Foundry technology development competes for AIST funding
- Ongoing operations should not be funded by individual PI projects
  - Maintenance of Foundry, including Help Desk, training, software maintenance

- ***However, data processing costs themselves would be paid for by customers as part of their Project***
  - ***PI controls what processing to pay for***

Will It Work?

- ESTO is supporting this effort as technology demonstration
- Goal is to demonstrate that the Foundry is:
  - Scientifically valuable
  - Technically feasible and efficient
  - Cost effective
  - Supported by the user and application development community
- We look forward to working with the community to explore the Foundry's potential

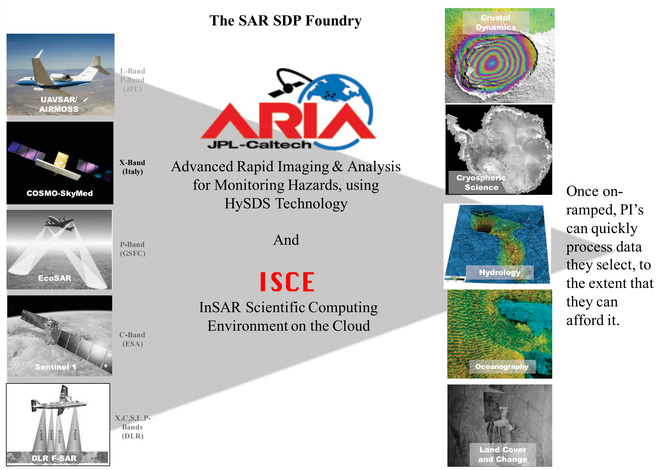
Synthetic Aperture Radar Science Data Processing (SAR SDP) Foundry Concept

- Definition
  - A set of user-selectable components 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
    - provides provenance, metadata, quality control, registration and workflow
  - 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
- Processing environment is published and community-accepted
- Clearly defined processes for on-ramping 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

The SAR SDP Foundry



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

L0 or L1 SAR data from NASA and other sources, as well as new processing workflows, can be on-ramped into the Foundry. Figure shows some of the planned data sets (COSMO-SkyMed implemented).

Acknowledgement

The authors would like to thank the Earth Science Technology Office and High End Computing Program at NASA for support. This work was performed at the Jet Propulsion Laboratory, California Institute of Technology under a contract with NASA.

NISAR Big Data Handling

- Average Input Data Volume to SDS: 3 TB/day
- Average Daily Production Volume: 104 TB / day
  - L0 (6 TB per day)
  - L1(32 TB per day, L-band)
  - L2 (66 TB per day, L-band)
- Aggregate Data Generated per day
  - Sustained mean 1.2GB/s – Forward processing
  - Sustained mean 4.8GB/s – Bulk reprocessing (optional)
- New architectures needed to support data handling and throughput
  - SDS processing at JPL on-premise
  - SDS processing at collocated with DAAC
  - SDS processing at public cloud
  - Hybrid cloud: on-premise and public cloud for processing and data storage

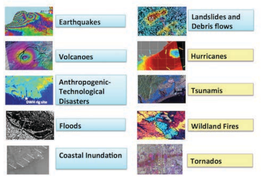
Expected high data throughput needs of modern SAR missions

- Sentinel 1A/1B
  - C-band SAR
  - 1.8TB/day raw data
- NISAR (2020)
  - L-band SAR
  - Deliver >100TB data products per day to DAACs
- SWOT (2020)
  - Bulk reprocessing requires >3GB/second mean sustained data throughput to DAAC

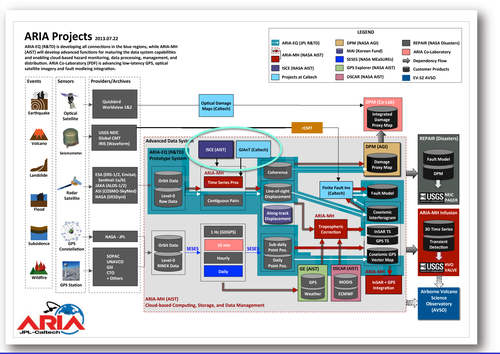
HySDS Option for the Proposed NISAR Mission

The NISAR mission will launch in 2020 and will collect large volumes of SAR data to measure Earth's changing ecosystems, dynamic surfaces, and ice masses, providing information about biomass, natural hazards, sea level rise, and groundwater.

Our Hybrid cloud Science Data System (HySDS) is currently being considered as a part of the solution for dealing with the flood of data in need of processing. The project will process the raw Level-0 data into Level-2 products using ISCE and provide these low level products directly to the scientists. HySDS would control the processing to these levels and could carry data to higher level science products.

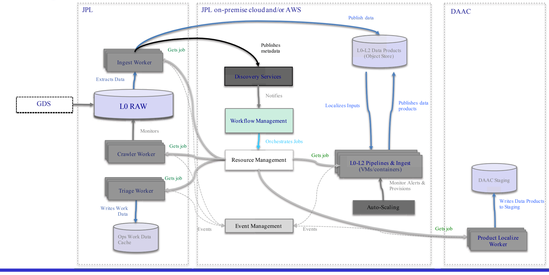


ISCE in ARIA (Advanced Rapid Imaging Analysis system)

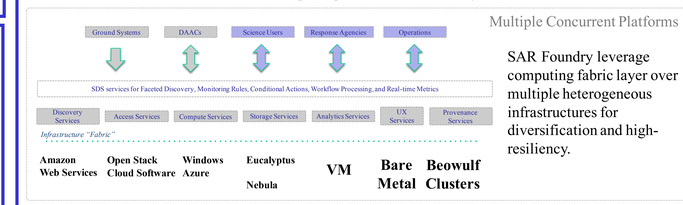


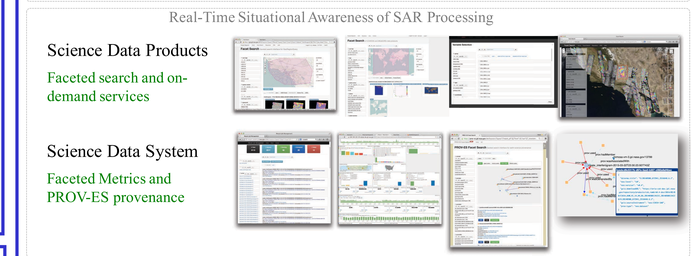
Hybrid Cloud Computing Science Data System (HySDS)

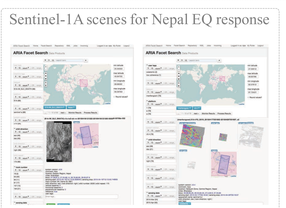
- Utilizes both on-premise and off-site infrastructure
  - PB-scale processing and storage purely in public cloud currently too expensive
- Hybrid Cloud data system architecture
  - Burst out to public cloud when demand exceeds on-premise resources
- Leverage Amazon GovCloud US to address export control and firewall security issues
- Processing at low-cost (up to 10X cheaper) using high-resiliency data system that can run in competitive AWS spot market
- Auto-scaling of science data system

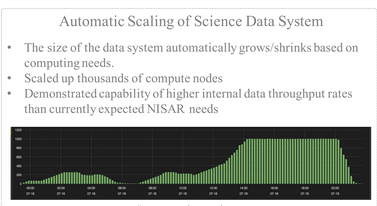


Cloud Computing-based SAR Foundry









SAR SDP Foundry Next Steps

- Characterize and develop on-ramping cost model
- Open HySDS adaptation for SAR SDP Foundry to beta users
- Develop charging protocols
- Define a sensor and workflow to on-ramp of interest to the community
- Exercise the Foundry with customer inputs

References

Gurrola, Eric, Gian Franco Sacco, Paul A. Rosen, and Howard Zebker, InSAR Scientific Computing Environment. *Earth Science Technology Forum* 2010.
H. Hua; G. Manipon; G. Sacco; S. Owen; E. Fielding; S. Yun; P. Lundgren; A. Moore; P. Milillo; P. Rosen; F. Webb; M. Simons; A. Smith; B. Wilson; "The Advanced Rapid Imaging And Analysis Data System: Automating SAR Data Analysis For Science And Hazard Response", IGARSS 2015, Milan, Italy, July 28, 2015
Zebker, Howard, Scott Hensley, Piyush Shanker, Cody Wortham, Geodetically Accurate InSAR Data Processor. *IEEE Trans. On Geoscience and Remote Sensing*, 48(12), 2010.
Shams, K. Riofrio, A. Hardman, S. et al. Enabling Earth Science through Cloud Computing. IEEE/AIAA Aerospace 2012.
PS Agram, R Jolivet, B Riel, YN Lin, M Simons, E Hetland, MP Doin, etal. New Radar Interferometric Time Series Analysis Toolbox Released. Eos, Transactions American Geophysical Union 94 (7), 69-70