A Hybrid-Cloud Science Data System Enabling Advanced Rapid Imaging & Analysis for Monitoring Hazards

AGU Fall Meeting 2012, San Francisco, CA

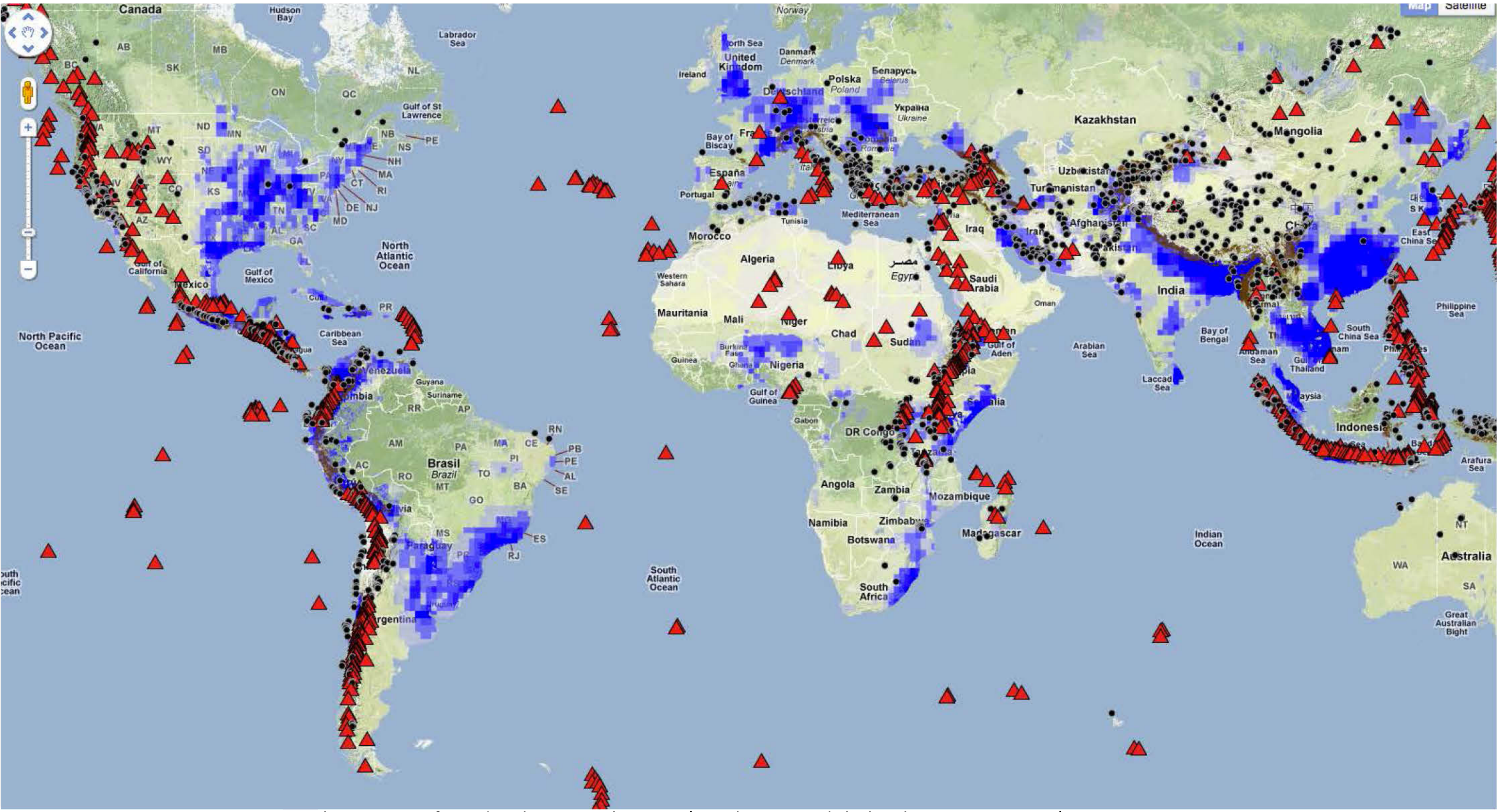
Oral IN33D-02 in session IN33D "Technology Enabling Earth Science From Big Data to Small Satellites”

Wednesday, December 5, 2012

Hook Hua[1], Susan Owen[1], Sang-Ho Yun[1], Paul Lundgren[1], Angelyn Moore[1], Eric Fielding[1], Gian Franco Sacco[1], Costin Radulescu[1], Mark Simons[2], Timothy M. Stough[1], Brian Wilson[1], Chris Mattmann[1], Peter F. Cervelli[4], Michael Poland[3], Jennifer Cruz[1]

[1] Jet Propulsion Laboratory
[2] California Institute of Technology
[3] USGS Hawaiian Volcano Observatory
[4] USGS Menlo Park Science Center

Natural Hazards



▲ known or inferred Holocene volcanoes (Smithsonian Global Volcanism Program)

● shallow earthquake epicenters (>50 km depth) on land with a magnitude of 6.5 or higher since 1976

■ extreme flood events (Dartmouth Flood Observatory , from 1985-2003)

■ subject to landslides (Norwegian Geotechnical Institute and UNEP-Grid Geneva).

Credit: Sang-Ho Yun (JPL)

Advanced Rapid Imaging and Analysis (ARIA) Overview

JPL-Caltech

Integrating Space Geodesy Seismology Modeling

Seismological Laboratory California Institute of Technology

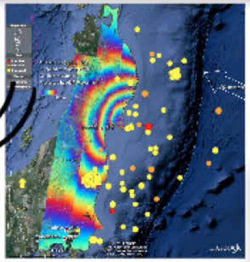
Radar Spacecraft

GPS Networks

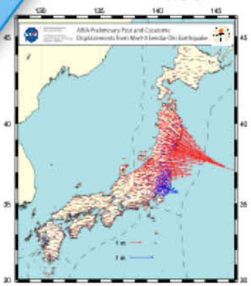
Seismic Networks

Optical Sensors

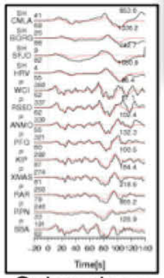
Automated Data Collection and Processing



Radar



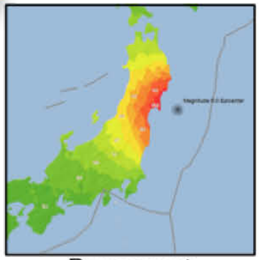
GPS



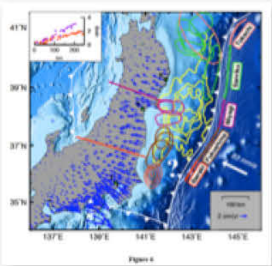
Seismology

Monitoring and Near Real-time Assessment

Damage Proxy Inundation



Permanent Ground Deformation



High Resolution Hazard Assessment from Fault Models

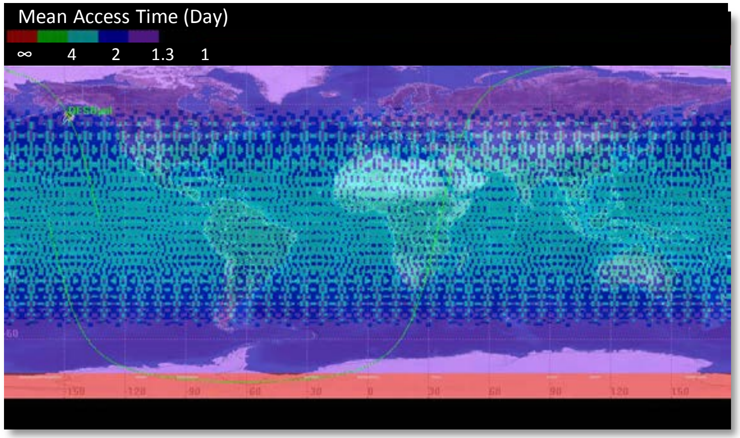
Examples from the 2011 M9.0 Tohoku-Oki (Japan) earthquake

Potential Partners

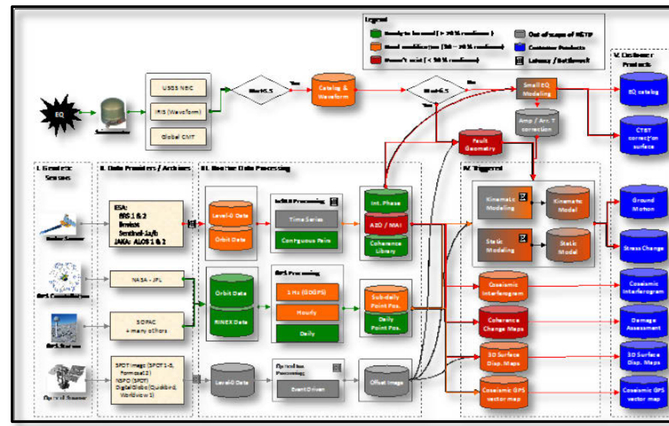
NASA · USGS science for a changing world · U.S. Department of Homeland Security · NSF · Cal EMA California Emergency Management Agency · World Bank

"We have high hopes for ARIA", K. Saito, World Bank

The Challenge of Leveraging Remote Sensing for Disaster Response

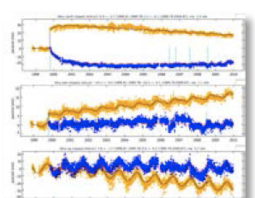


Automated data system are required to analyze large quantities of data from LSM (formerly DESDynI), other satellite missions, and rapidly expanding GPS networks.

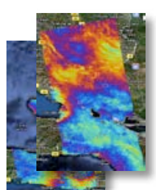


Going from Artisan to Automation: Use system engineering approach to translate specialized data analysis into operational capability.

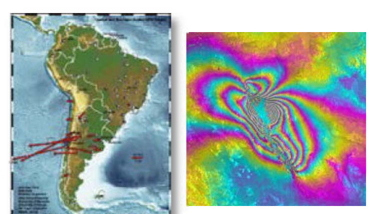
Demonstrate response to hazards with standardized set of data products for decision & policy makers.



Temporal Records of Ground Deformation

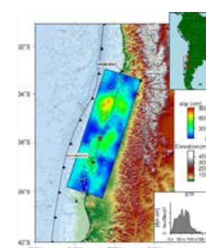


Spatial Maps of Ground Deformation



Coseismic Ground Deformation

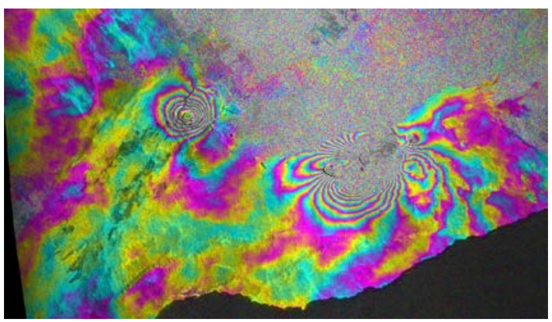
Coseismic Damage



Earthquake Models

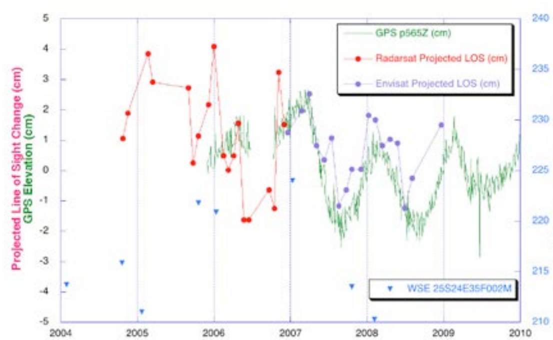
Hazard Response

- Volcanic eruptions, landslides, and levee failures
  - Hazards that can be more accurately forecasted with sufficient monitoring of precursory ground deformation
- Lava flows, mudslides, tornadoes, floods, and other natural and man-made disasters
  - Surface change detected by coherence and reflectivity change maps
- Use of high-resolution geodetic measurements from GPS and InSAR



The recent Kamoamoa Eruption occurred from March 5-9, 2011. The figure above is a near real-time interfergraom from Cosmo-SkyMed data, produced on March 7, ~12 hours after acquisition (courtesy of Paul Lundgren / JPL)

Damage Proxy Map of flooding in Ishinomaki, Japan caused by Tohoku Tsunami. Envisat (ESA, C-band) ASAR 2011.01.31-2011.03.02-2011.04.01 (courtesy of S-H Yun)



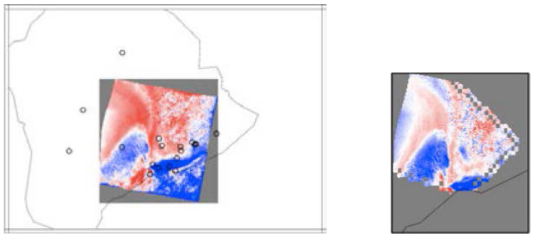
GPS and InSAR time series measurements of surface elevation changes induced by pumping in California's Central Valley. InSAR derives from Radarsat and Envisat, with no projection difference corrections. Combining results from multiple satellites can extend the time series, providing additional insights on seasonal and secular changes (courtesy of T. Farr)

Motivation

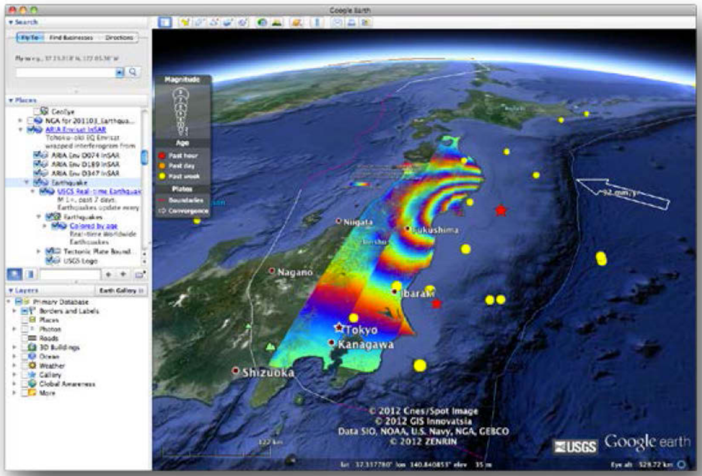
- Automation barriers
  - Analyses of InSAR and GPS data sets are currently handcrafted following events
  - Not generated rapidly and reliably enough for use in operational response to natural disasters.
  - Difficult for many volcano observatories and other monitoring agencies to process GPS and InSAR products in an automated scenario needed for continual monitoring of events.
- Interoperability barriers
  - Multi-sensor observation data access, preparation, and fusion to create actionable products.
  - Interoperable services, data, and metadata
- Improving science and hazards monitoring
  - Combining high spatial resolution InSAR products with high temporal resolution GPS products
  - Automating data preparation & processing across global-scale areas of interests

Objectives

- Develop a service-oriented hazard/disaster monitoring data system enabling both science and decision-support communities to monitor ground motion in areas of interest with InSAR and GPS.
- Enable high-volume and low-latency automatic generation of NASA Solid Earth science data products (InSAR and GPS) to support hazards monitoring.
- Enable improved understanding through visualization, mining, and cross-agency sharing of results.
- Enable interoperable discovery, access, and sharing of derived actionable products for hazards monitoring.

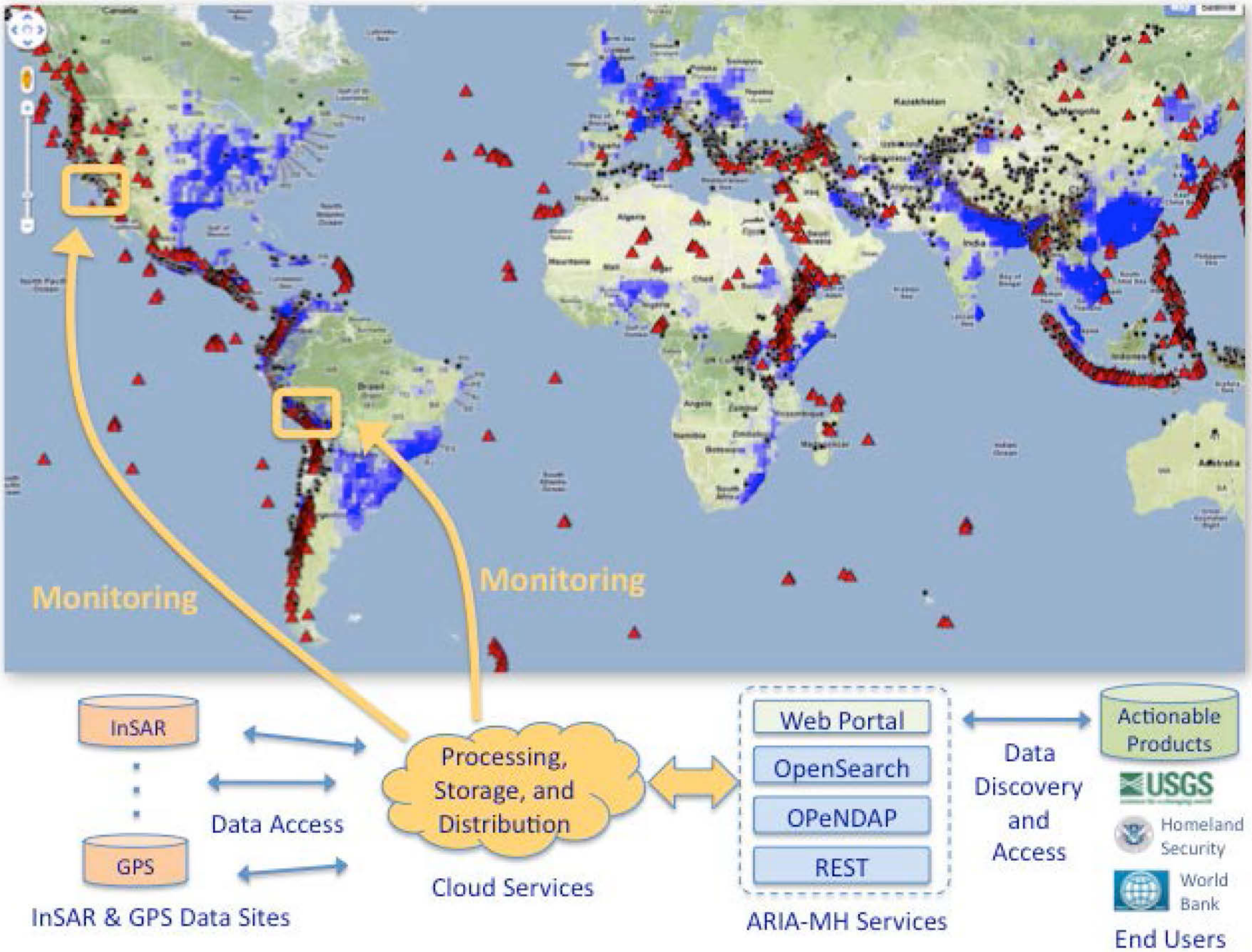


GPS-based troposphere corrected interferograms



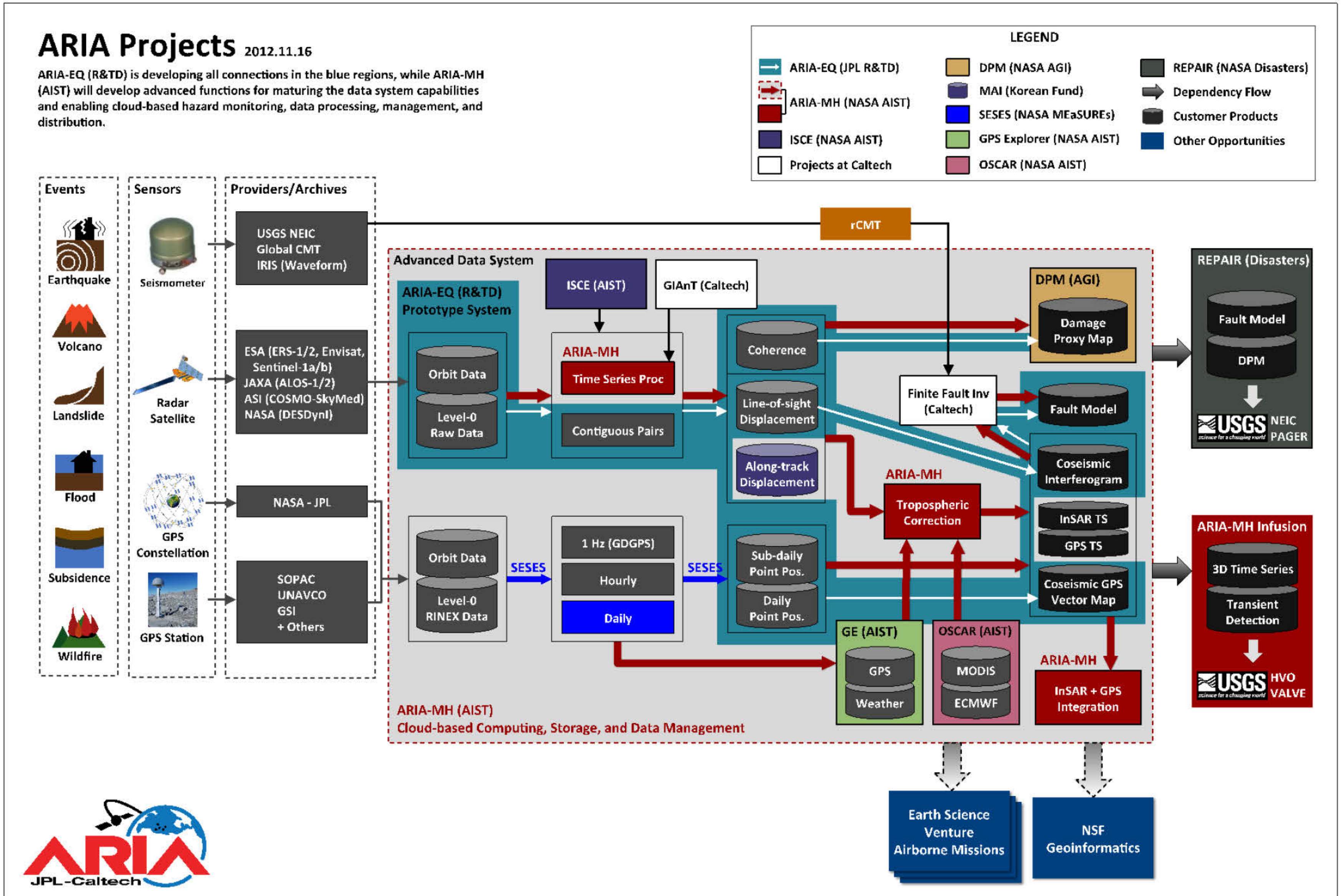
Visualization of geocoded interferogram in Google Earth

Approach

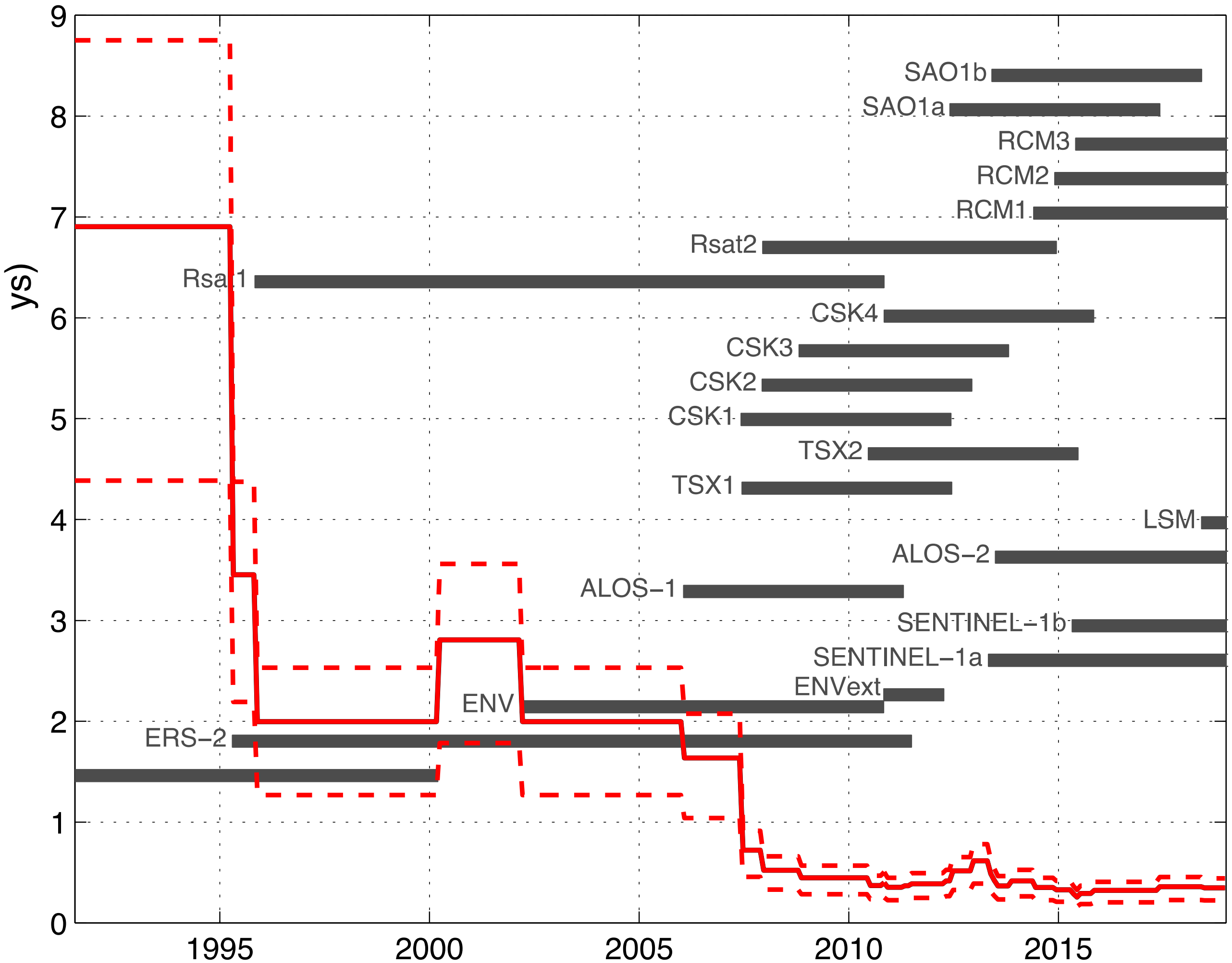


Monitoring
Monitoring
InSAR
GPS
Data Access
InSAR & GPS Data Sites
Processing, Storage, and Distribution
Cloud Services
Web Portal
OpenSearch
OPeNDAP
REST
ARIA-MH Services
Data Discovery and Access
Actionable Products
USGS
Homeland Security
World Bank
End Users

ARIA Projects and Data Products



Data Acquisition Latency of InSAR Missions



Expected wait time until the first SAR satellite to visit after an event

Ascending + descending orbit

Right-looking mode

Latitude of 38° N/S

Present: 15 hours

2020: 8 hours

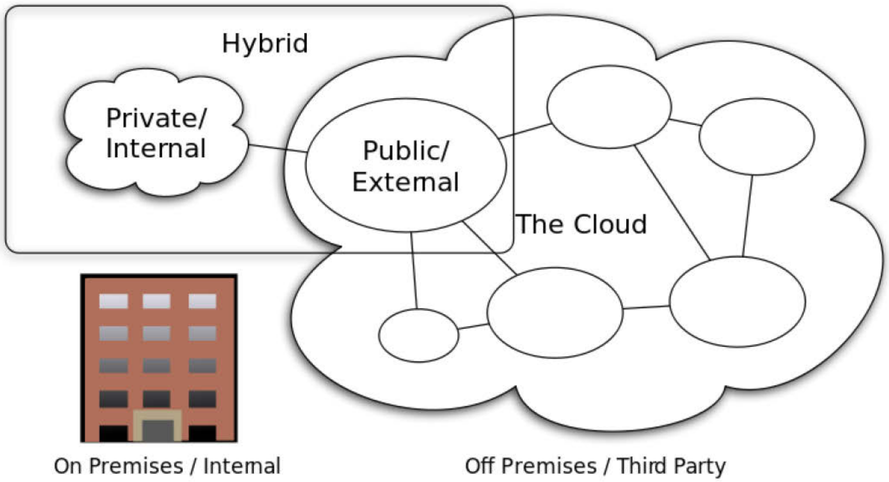
Credit: Sang-Ho Yun (JPL)

Motivation for Cloud Computing

- Incoming flood of data volume
  - Nominally ~700MB/scene, 100Ks of scenes, order of 10GBs-100GBs temp storage per product, PBs total storage of products
  - Example InSAR satellites
    - COSMO-SkyMed (CSK) and CSK second generation data from ASI
    - Sentinel 1A/1B
    - ALOS-2
    - Decadal Survey: US L-band SAR Mission
- Monitoring
  - User definable bounding box regions of interest for nominal background monitoring/processing.
- Elasticity of computing when responding to events
- Process migration to geographically disperse data centers
  - ESDIS DAACs (e.g. ASF)
  - UNAVCO SAR Archive
  - ASI for CSK
  - DLR for TerraSAR-X
  - JAXA
  - Various GEO Supersites

Hybrid Cloud Computing

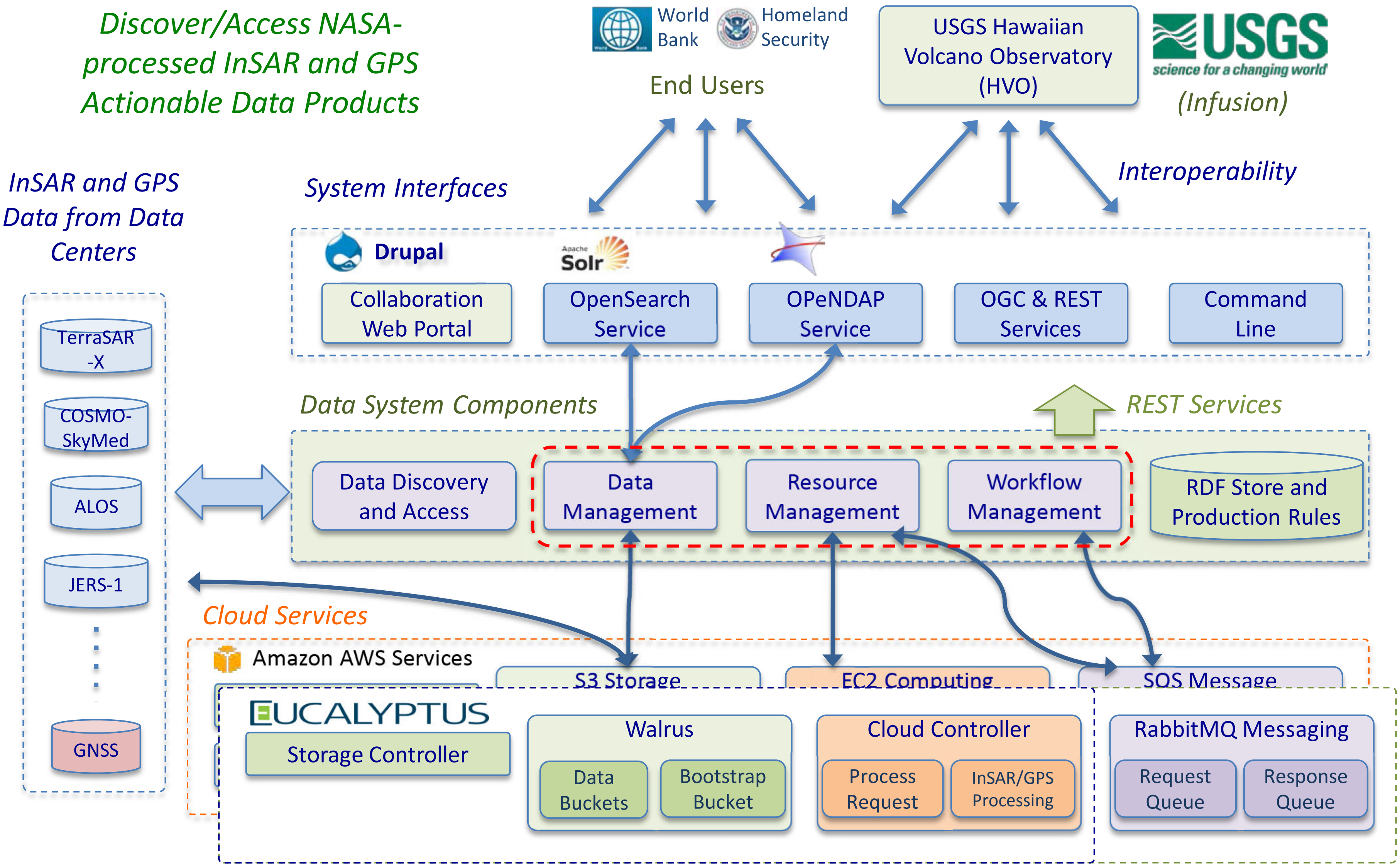
- Utilizes both *on-premise* and *off-site* infrastructure
- Leverage existing infrastructure investment
  - Sunk cost investment
- Hybrid data system architecture
  - Burst out to public cloud when demand exceeds on-premise resources
  - AWS-compatible Eucalyptus cloud stack
- *Vendor lock in*: decrease dependency on commercial vendors
- Recognize that not all resources need to be in the public cloud
  - ITAR / EAR99 software
  - Costs for cloud storage
- Degree of fault tolerance



Cloud Computing Types

AWS and Eucalyptus Compatibility

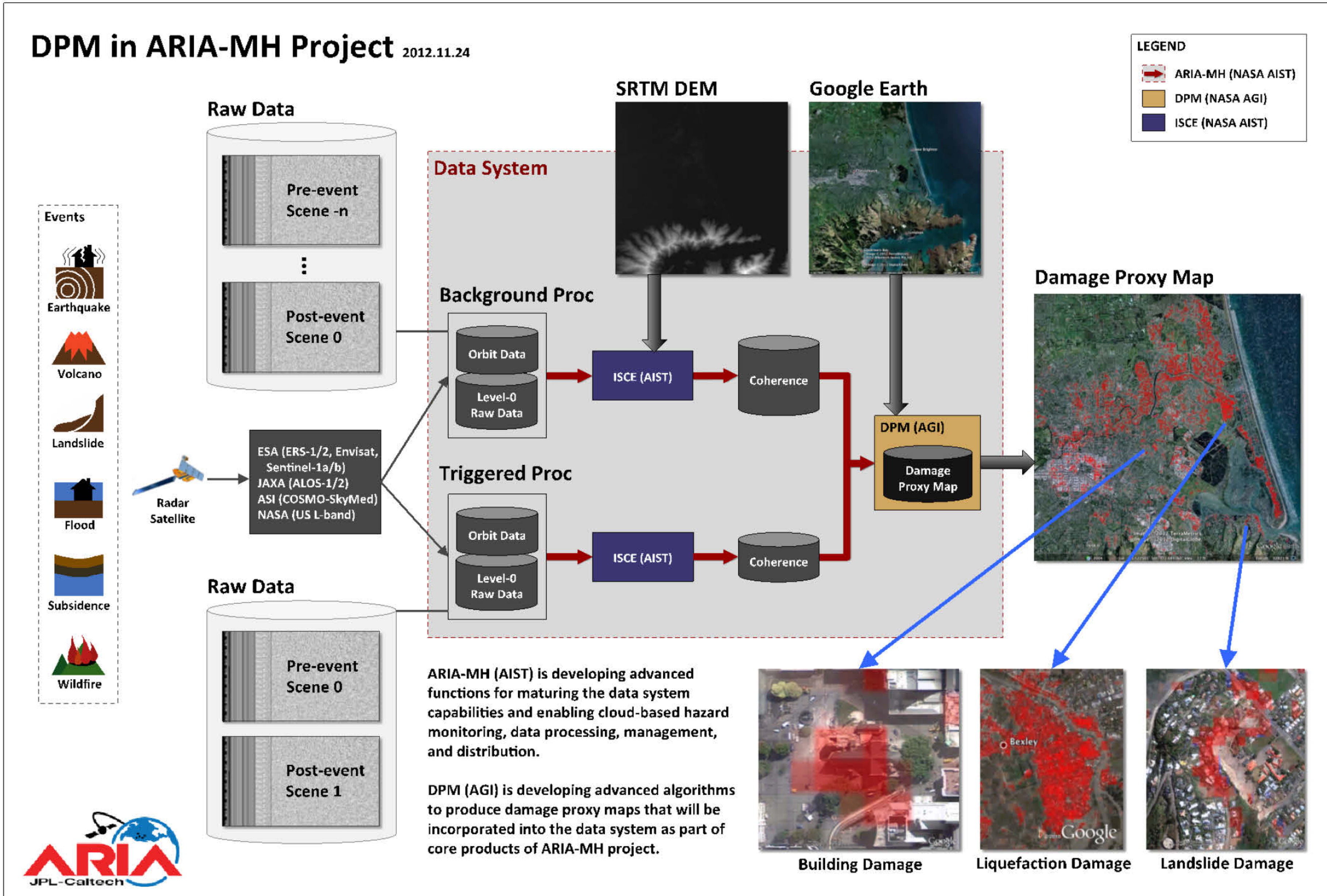
High-Level Architecture



Hybrid Considerations

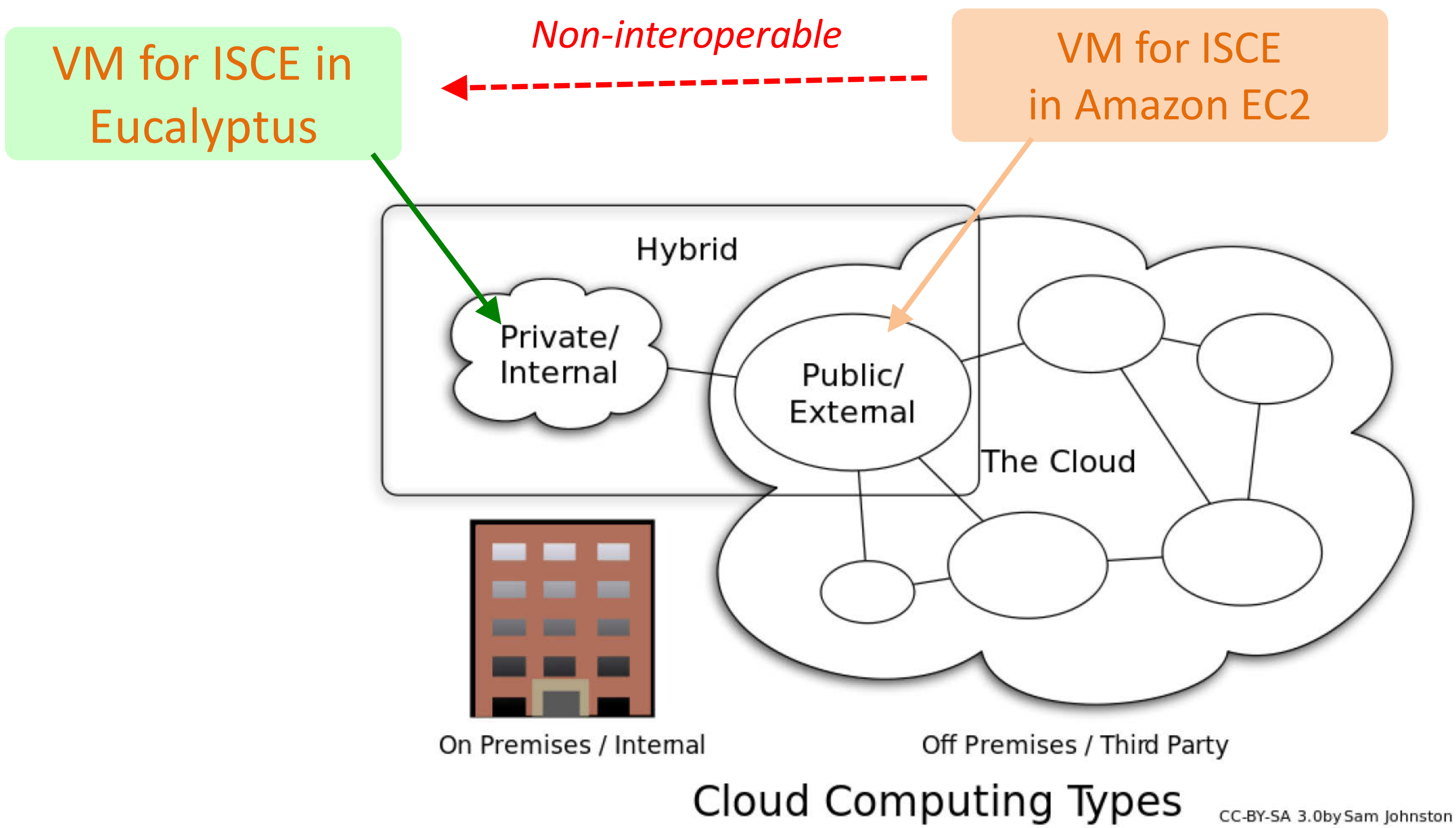
- Heterogeneity of platforms supported
  1. Amazon AWS
  2. Eucalyptus
  3. Ligh*ter* weight VMware
- Eucalyptus setup
  - Upfront effort to setup on-premise cloud stack
  - Institutional constraints on IP address pools
  - Service topography
- Incompatibilities across on-premise and off-site virtualization environments
  - Hypervisors and formats (e.g. Open Virtualization Format)
  - Setup of VMs
    - EBS mount and Vmware shared mounts
- Lowest common denominator of services
  - Compute, Block storage, Bucket Storage, VM, and Identity

Damage Proxy Map

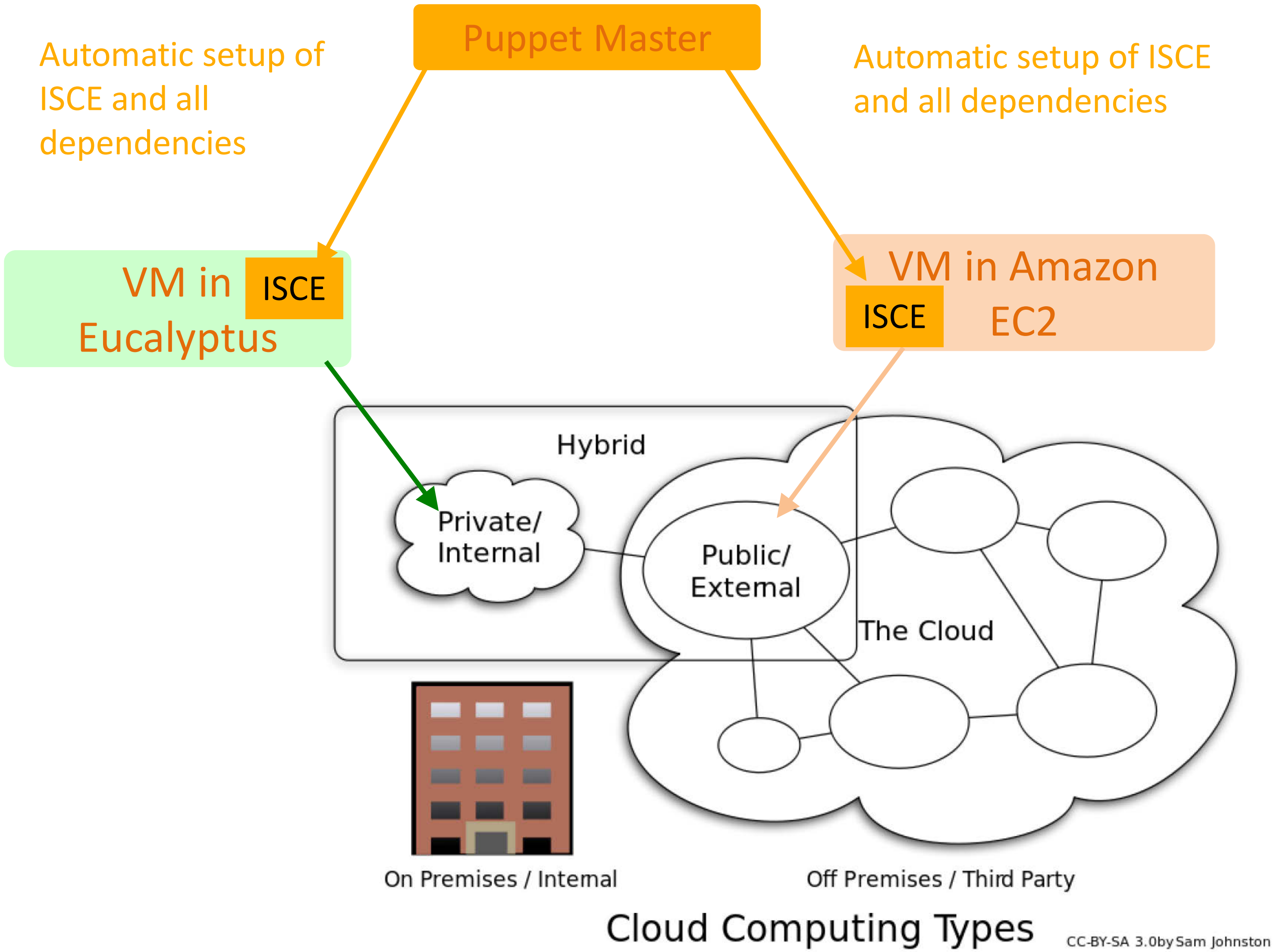


Migrating Geodetic Processing into Virtual Machines

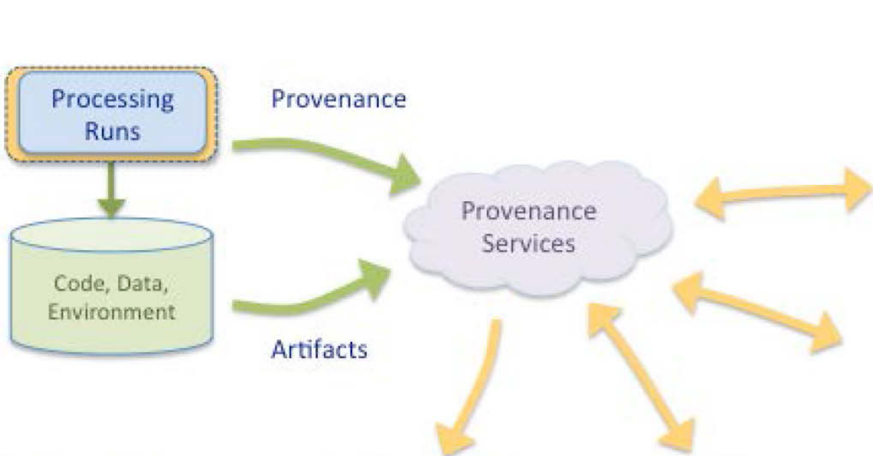
- Setup of geodetic processing environment
- Variability in local processing disk storage
  - Amazon EBS mounts vs Vmware mounts vs internal VM storage

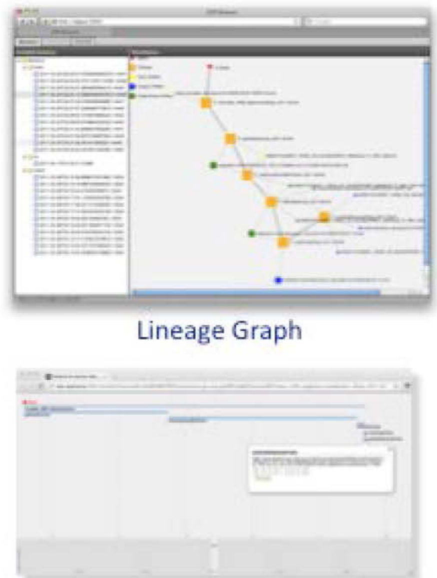


"Puppetizing" the Services



Provenance as a Service



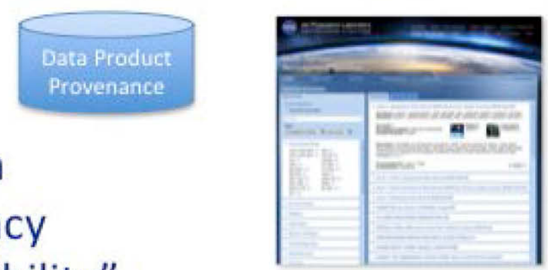


Lineage Graph

Processing Timeline

Enable:

1. Capture
2. Exploration
3. Transparency
4. "Reproducibility"

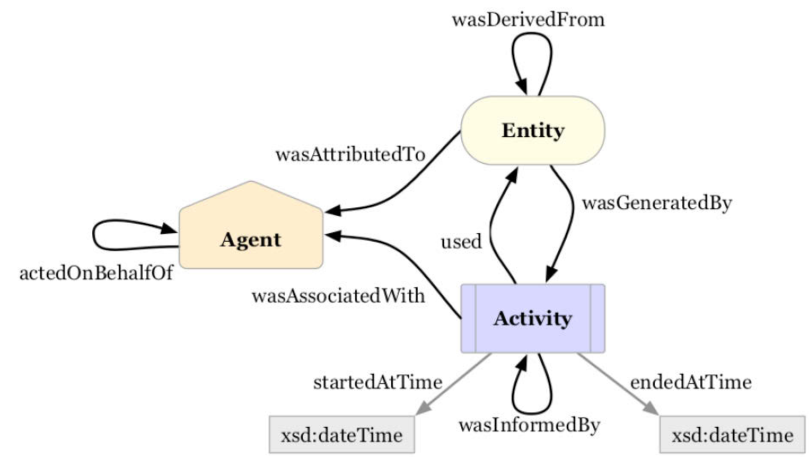


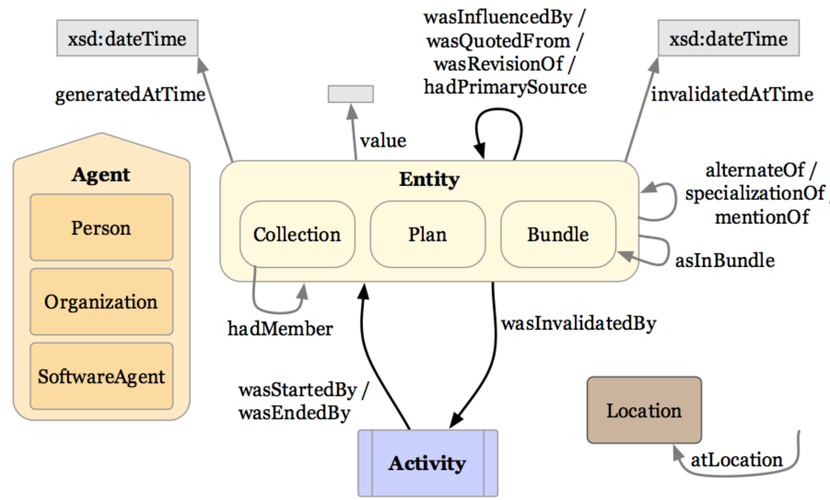
Faceted Search

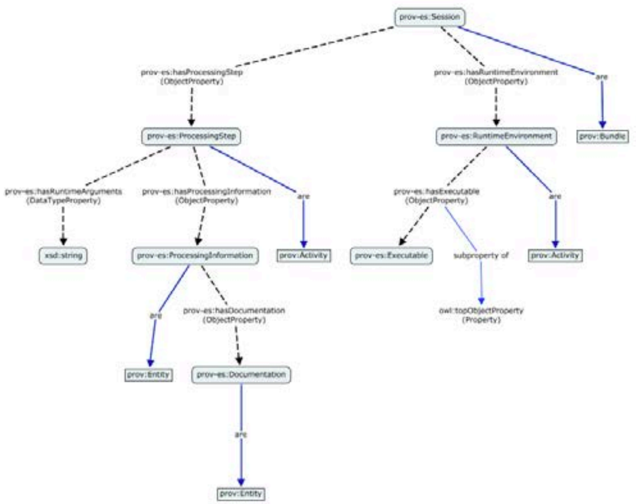
- Provenance capture of metadata and artifacts
- Semantic rules
  - Infer facets for discovery
  - Summarize for presentation
- Lineage graph
  - levels of detail
  - session chaining
- Faceted Browsing
  - Drill down into multidimensional provenance
  - Constrain by time, user, session, data type, version, etc…

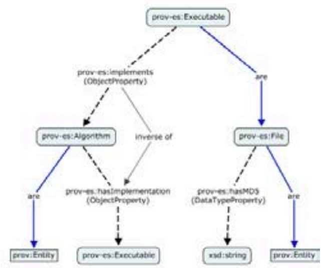
Provenance for Earth Science

- *Enable more transparency and trust—especially for actionable data products*
- Leverage W3C international standards
- Encodes PROV Data Model (PROV-DM) in Web Ontology Language (OWL2).
- Good decoupling of provenance models and encodings
- Supports Qualified terms enabling more custom metadata
- Extend W3C PROV-Ontology for modeling Earth Science Data System Processing (PROV-ES)









Conclusions

- The global coverage offered by satellite-based SAR missions, and rapidly expanding GPS networks can provide orders of magnitude more observations and improve hazard response
  - …if we have a data system that can efficiently monitor and analyze the voluminous data, and provide users the tools to access data products.
- Hybrid cloud may be more effective for these needs
  - Leverage existing on-premise investments
  - Off-site elasticity (bursting)
- Interoperability issues of hybrid cloud
  - Heterogeneity: more difficult to deploy than traditional cloud approaches

Other ARIA-related Activities at AGU FM2012

- **Tonight: Ignite @ AGU “Radar for Life”, Sang-Ho Yun**
- G52A-05 “The ARIA project: Advanced Rapid Imaging and Analysis for Natural Hazard Monitoring and Response”, Susan E. Owen; Frank Webb; Mark Simons; Paul A. Rosen; Jennifer Cruz; Sang-Ho Yun; Eric J. Fielding; Angelyn W. Moore; Hook Hua; Piyush Agram; Paul Lundgren
- T12C-07 “Imaging transient slip events in southwest Japan using reanalyzed Japanese GEONET GPS time series”, Zhen Liu; Angelyn W. Moore; Susan E. Owen
- IN31C-1508 “InSAR Scientific Computing Environment on the Cloud”, Paul A. Rosen; Khawaja S. Shams; Eric M. Gurrola; Brett A. George; David S. Knight
- IN32A-06 "Rapid and Robust Damage Detection using Radar Remote Sensing” S. Yun; E.J. Fielding; M. Simons; F. Webb; P.A. Rosen; S.E. Owen
- IN33C-1539 “Development of a System to Generate Near Real Time Tropospheric Delay and Precipitable Water Vapor in situ at Geodetic GPS Stations, to Improve Forecasting of Severe Weather Events”, Angelyn W. Moore; Yehuda Bock; Jianghui Geng; Seth I. Gutman; Jayme L. Laber; Todd Morris; D. G. Offield; Ivory Small; Melinda B. SquibbG41B-07 9:30 AM "Combined InSAR, Pixel Tracking, GPS, and Seismic Waveform Analysis for Fault Slip Evolution Model of the 2011 M7.1 Van Earthquake in Turkey and InSAR Time-series Analysis for Postseismic Deformation”, Eric J. Fielding; Paul Lundgren; Jascha Polet; Tuncay Taymaz; Susan E. Owen; Mark Simons; Mahdi Motagh; Sang-Ho Yun
- G43A-0897 “GIAnT - Generic InSAR Analysis Toolbox”, Piyush Agram; Romain Jolivet; Bryan V. Riel; Mark Simons; Marie-Pierre Doin; Cecile Lasserre; Eric A. Hetland
- G43A-0916 "Deformation precursory to the March 2011 Kamoamoa fissure eruption, Kilauea Volcano, Hawai'i”, Paul Lundgren; Michael P. Poland; Asta Miklius