

Open Source Science for ESO Mission Processing Study

Identify a system architecture that meets the ESO mission processing objectives, supports open science, enables system efficiencies, and promotes earth-system science.

Workshop #1
October 19-20, 2021

NISAR Program Perspective
Gerald Bawden

NISAR Science Objectives

Key Scientific Objectives

- Understand the response of ice sheets and glaciers to climate change and the interaction of sea ice and climate
- Understand the dynamics of carbon storage and uptake in wooded, agricultural, wetland, and permafrost systems
- Improve knowledge for forecasts of earthquakes, volcanic eruptions, and landslides

Key Applications Objectives

- Understand societal impacts of dynamics of water, hydrocarbon, and sequestered CO₂ reservoirs
- Enhance agricultural monitoring capability in support of food security objectives
- Apply NISAR's unique data sets to explore the potentials for urgent response and hazard mitigation





Program Scope for NISAR (1 of 2)

Surface
Deformation
and Change



- NISAR will collect data over all land and all ice on every orbit, including sea ice and coastal regions.
- NISAR supports research and applications that span nearly every Focus Area and Application Science Program Area. See <https://nisar.jpl.nasa.gov/>
- The volume of data that NISAR will collect/produce is staggering. In 2017, the entire ESDIS archive had 22 PB (all missions including 45 year of Landsat) of data, NISAR will generate about 140 PT (petabytes) in 3 years.... >6-fold increase in ESDIS 2017 holdings.
- This presentation is focused on NISAR, but concepts apply to SDC with NISAR as the pathfinder.

Scope of Responsibility and Program Objectives *in context of OSS & MDPS*

- **Ensure that NISAR data are used to their fullest potential...** *“No photon unused,”* Craig Dobson.
- **Lower the bar for data usability for all NISAR data and products...** *“Take the SAR out of SAR,”* First Applications Workshop.
 - SAR data is complex with real (backscatter) and imaginary (phase) components ... *most people do not know how to work with complex numbers.*
 - NISAR will produce RTC (Radiometric Terrain Corrected) backscatter images and interferograms (phase for deformation change)
- **No other SAR mission (open access data policy) currently generates L2 data:** *Only raw SAR data or SLC. Therefore, a user must know how to process SAR/InSAR data for their remote sensing needs.*
NISAR will be the 1st SAR mission that will systematically release L2 products + L3 SNWG soil moisture.
- **Expand/grow the global SAR usership across the sciences and applications.**
 - Application workshops (10 so far and 2 more in the works for 2022).
 - OSS will help develop the tools that grow communities (especially non-technical) through the creation and sharing of new tools and capabilities that target the diverse reach of the community... including users that do not know that they would benefit from SAR/InSAR data products.

Vision for Open Science (1 of 3)

How does open-science fit within your program objectives?

- The pathway forward with NISAR is with L3-L4 products, cloud computing, and **OSS**.
- NISAR's systematic global collection of L-band SAR data will be a paradigm shift:
 - Local/Regional Continental Global Science
 - Data characterization Data Assimilation/Modeling (AI, ML, Ensemble) Geophysical Processes
- Fully taking advantage of the volume of NISAR data will require new approaches that work with dense time-series data... Open Science could help lead the way.

What investments are you making towards open-science and what are the expected outcomes?

- NISAR ATBDs - Jupyter Notebooks....
- ASF – OpenSARLab
- GRFN – Getting Ready for NISAR

Vision for Open Science (2 of 3)

What barriers exist across NASA that inhibit participation to open-science?

- Easy pathways/incentives to bring researchers (new and established) into the cloud. ... *Old dog new tricks*
 - PIs that have existing algorithms, raids/local clouds, computer clusters that meet their needs:
 - *Why migrate to the cloud? They need encouragement...*
 - Many universities have their own cloud platforms for researchers; PIs are encouraged to use these resources... therefore collaborations remain at the university.
 - *Opinion: Every mission (OK most) should have an Open Science/Cloud Computing Early Adopter Program that proactively brings the community into the Open Science platform from the beginning such that it is the established approach before the mission launches.*
 - A 'NISAR Cloud Computing Early Adopter Program' was proposed but was declined...something was in the works (a few years ago?).
- A successful OSS program will have a Cloud entry point that is:
 - Very compelling for a researcher to make the transition from local to the Cloud
 - Provides tools and capabilities that they may not have access to (further incentives)
 - Encourages the development of new OSS tools/algorithms
 - *Why is it in their best interest? What do they gain?*

Vision for Open Science (3 of 3)

What barriers exist across NASA that inhibit participation to open-science? (cont.)

- Provide a mechanism to track success – thereby help incentivize OSS participation
 - How do PIs/users get credit in the ‘publish or perish’ academic/research grade scientist landscape?
 - In academia and Gov. RGE (Research Grade Employees), “Publish” equates to journal articles... OSS is not part of their evaluation process.
 - *Why should a research share intellectual property that helps a competitor (perhaps without acknowledgement)?*
 - *This is needed for both academic and US Government researchers.*
- Access: Open-source science tools needs to well integrated with data access.
- Discoverability: Researchers need to be able to find the tools/solutions they seek.
- Quality control: There needs to be a way to assess the quality of the tools (i.e. reviews, stars, NASA evaluated, etc.) ... say Apple AppStore approach.
- Rewards system(?): Could a reward system to encourage OSS? ‘Carrot and Stick’ – although, I’m not necessarily fond of stick approaches. ROSES could be a starting point.

Vision for Mission Data Processing

What does it mean for mission data processing systems to support open science?

The algorithms that process NISAR's L1+ data and products would be open source. This includes all L2 products and algorithms for science cal/val: L3 products are only available for the defined cal/val locations for each science discipline.

The NISAR ATBDs are in Jupyter Notebooks such the science community can directly work with the same algorithms and apply the same algorithms anywhere in the world... say in areas that are not covered in the science cal/val.

What does it mean for mission data processing systems to enable earth system science?

- The L2 products from NISAR should be in a format that is widely accessible to their targeted communities.
 - Data needs to be directly accessible with other SAR missions: i.e. Sentinel-1
 - Data needs to be directly accessible with other satellite data: i.e. Landsat, Sentinel-2
- Analysis Ready Data – ARD – that is consistent (format, projection, etc.) with community accepted products (both SAR and optical)
- NASA's holding should be processed at a level that could be simply added as another band/wavelength combination for the processing of optical data.
- Easy access to cal/val data, all metadata data, calibration/ephemerides data, etc.
- *Committee on Earth Observation Satellites (CEOS) Working Group on Calibration and Validation (WGCV), SAR Subgroup: [link](#)*
 - *Chair: Dr. Bruce Chapman – JPL/NISAR's Cal/Val lead*



Vision for Mission Data Processing

What does it mean for mission data processing systems to be efficient?

- Based on the volume of data that NISAR will be generating, efficient algorithms will be necessary/required to prevent a data processing backlog.

What opportunities does NASA have for advancing mission data processing systems?

- The next two presentations on NISAR will review what we are doing.

Guidance for the Study

Do you have any recommendations for the study steering committee?

The vision for ESO is to study the Earth as an interconnected system through the co-collection of measurements from the DO constellation + NISAR, therefore, the DOs+NISAR measurements must be easy to combine (i.e. ARDs, data fusion, common formats, projections, etc.) and co-analyzed.

- Measurements Modeling (AI, ML, Data Assimilation, etc.) Scientific Advancement/Understanding. This all starts with the MDPS.

Need to understand the source data for each ESO missions and when to transform it to a product (data format, projections, L3, L4+) that can be combined/integrated with other data types (for example):

- SAR SLC RTC, which can be treated as another optical band

Need to understand the data/product needs among the ESO missions that enable integrated ESO science:

- NISAR/SDC Surface water extent/water storage Water availability Food security
- MC Water loss/gain Water use/availability Food security
- CCP Precipitation Immediate water availability Food security
- SBG Crop type/health Water needs Food security

Need to incentivize a culture for OSS and cloud computing. If it is too much work or no apparent benefit, then process will not be successful.

SNWG activities should be considered... they will provide detailed products supporting the overall objectives of ESO.

Start *now* and with NISAR (2023). The NISAR & SDC time-series will be the continuity thread and book ends for ESO. NISAR overlaps with similar current missions (i.e. MC & GRACE-FO... etc) and can be used to develop the OSS and MDPS approach as the other DOs missions are launched.