STARE for scalable unification of diverse data within Earth, Space, and Planetary Science

Michael Rilee[1], Kwo-Sen Kuo[2], James Gallagher[3], James Frew[4], Niklas Griessbaum[4], Kodi Neumiller[3], Robert Wolfe[5], Hongfeng Yu[6], and Pamela Clark[7]

[1]Rilee Systems Technologies LLC
[2]Bayesics, LLC
[3]OPeNDAP, Inc.
[4]University of California, Santa Barbara
[5]National Aeronautics and Space Administration
[6]University of Nebraska, Lincoln
[7]Jet Propulsion Laboratory

November 23, 2022

Abstract

The variety of missions and observational campaigns in Earth and Space Science has led to a vast number of files containing low-level datasets with native, incompatible arrays. While higher-level datasets re-interpret this low-level data onto common, comparable arrays, this standardization moves scientists farther away from the observations, constraining the analysis and science that can be performed. SpatioTemporal Adaptive Resolution Encoding (STARE) provides a parallelizable, scalable index for co-aligning data with different native spatiotemporal formats efficiently across distributed computing resources. STARE is particularly useful for opening low-level datasets to intercomparison and integrative analysis by providing “array” indexes that carry spatiotemporal semantics, unifying datasets with previously incomparable native array indexing. We are developing STARE as a software library with both C++ and Python APIs and are integrating STARE indexing with existing data transfer tools (OPeNDAP). By organizing data in a hierarchical format and taking advantage of the Hierarchical Data Format’s (HDF) virtualization features, STARE may provide an end user with familiar HDF usability with STARE-enhanced performance and data unification on the back end. Furthermore, STARE’s spatial encoding can be used to index and integrate datasets associated with other planetary bodies, bringing scalability and unification of diverse data for planetary and space science as well.

AGU Fall Meeting, 9-14 December 2019, San Francisco, CA, USA IN31B-0791

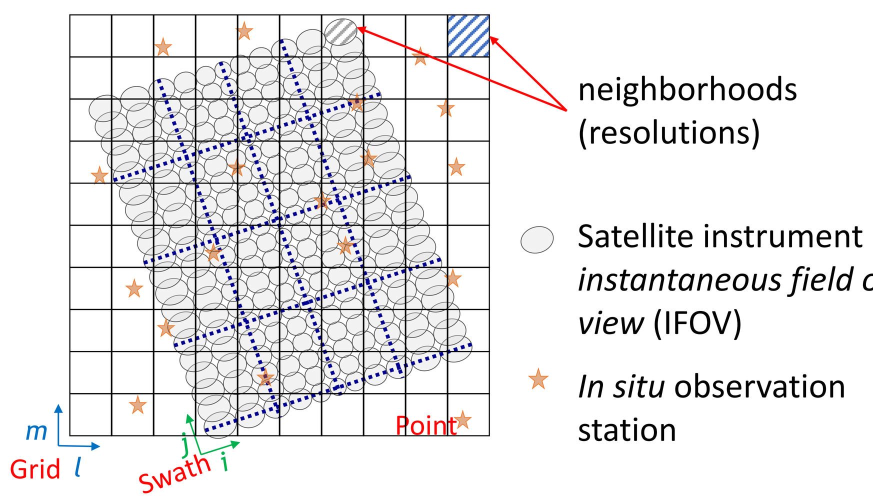
STARE for scalable unification of diverse data within Earth, Space, and Planetary Science

M L Rilee[1] K-S Kuo[2] J H R Gallagher[3] J Frew[4] N Griessbaum[4] K Neumiller[3] R E Wolfe[5] Hongfeng Yu[6] P E Clark[7]
[1]Rilee Systems Technologies, LLC, Derwood MD; [2]Bayesics LLC, Bowie, MD; [3]OPeNDAP, Inc., Narragansett, RI; [4]University of California, Santa Barbara; [5]NASA Goddard Space Flight Center, Greenbelt, MD; [6]University of Nebraska, Lincoln; [7]Jet Propulsion Laboratory, Pasadena, CA

STARE provides an alternative to native-array indexing and longitude/latitude geolocation by combining spatiotemporal location and neighborhood semantics, opening diverse low-level datasets to customized integrative analysis using distributed/parallel computing or Cloud.

The Problem

- ❖ Low-level datasets have immense value but are hard to use and integrate.
- ❖ High-level datasets reinterpret low-level data, e.g. to a common grid or resolution or to fill in gaps, easing integration, while washing out important details.
- ❖ Customized integration aligning diverse data (swath, grid, point) and exploiting high resolution data does not scale.



- ❖ Native array indexing disconnects data from its spatiotemporal arrangement impeding integration and coalignment.
- ❖ We need an index strategy supporting alignment of data in space and time and on distributed/parallel computing resources.

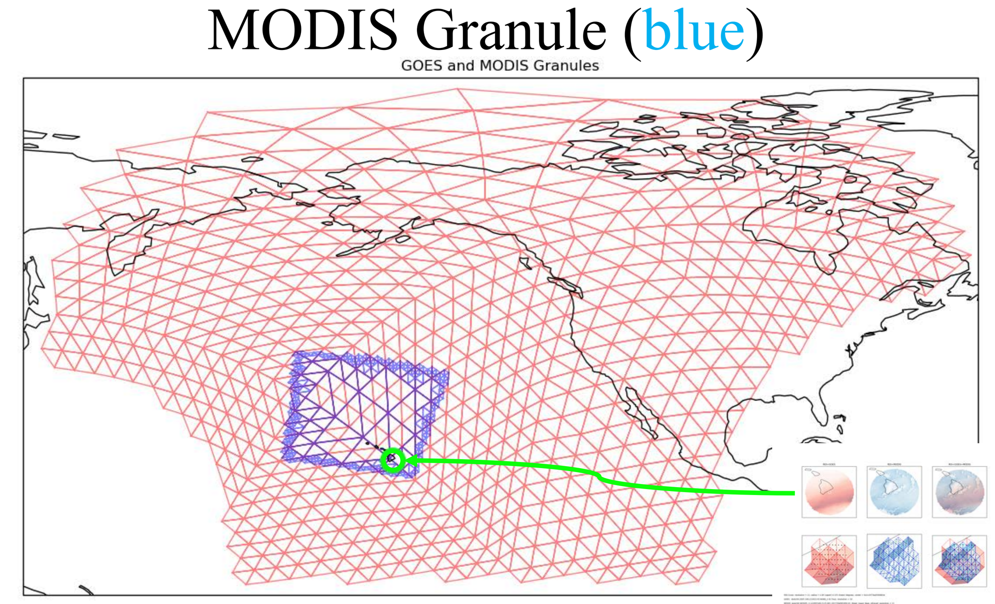
STARE

- ❖ **S**patio**T**emporal **A**daptive-**R**esolution **E**ncoding – A hierarchical geo-spatiotemporal reference using **integers**.
- ❖ STARE has two elements: The geo-spatial *Hierarchical Triangular Mesh* (HTM) and the temporal *Hierarchical Calendrical Partition* (HCP).
- ❖ HTM is a way to index the surface (actually, solid angle) of a sphere using a hierarchy of spherical triangles.
  1. Start with an inscribing octahedron of a sphere.
  2. Bisect each edge.
  3. Project the bisecting points from sphere center to the sphere surface forming 4 smaller spherical triangles.
  4. Repeat quadfurcation from 2.
- ❖ The index at the 27th quadfurcation level (Q-level) is used for indexing geolocation with better than 10 cm precision.
- ❖ *Neighborhood size* is given by Q-level, e.g Q-10~10km.
- ❖ HTM encodes both *location* and *neighborhood* (trixel).
- ❖ HCP encoding is similar to HTM, but branching at each "neighborhood" level depends on calendrical subdivision.
- ❖ Temporal indexing is built on international standards TAI and UTC using the International Astronomical Union's *Standards of Fundamental Astronomy* (via *ERFA).*

STARE Unified Indexing

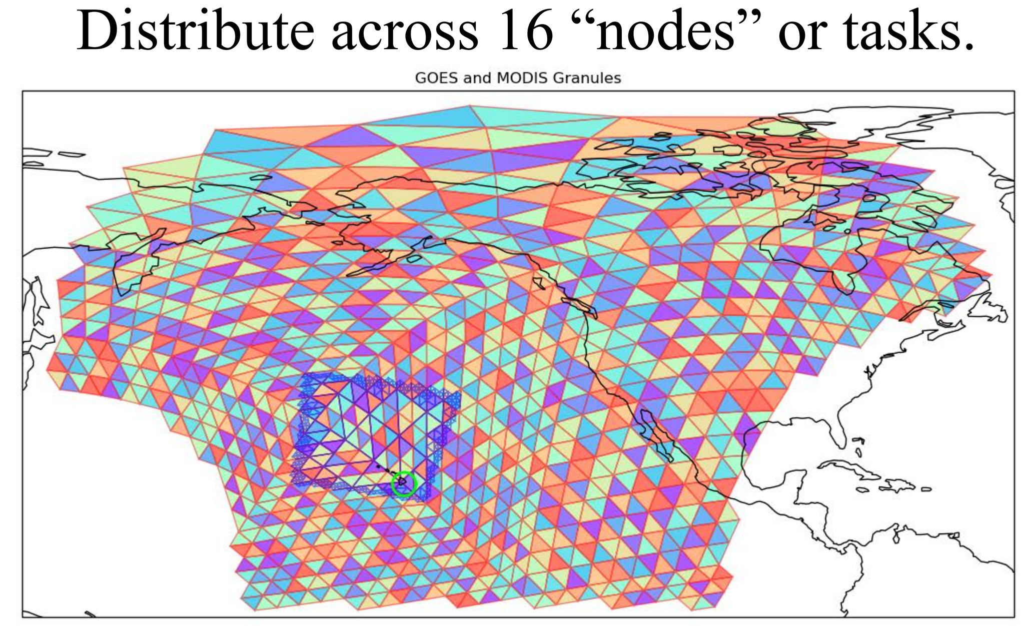
GOES footprint (red)

MODIS Granule (blue)

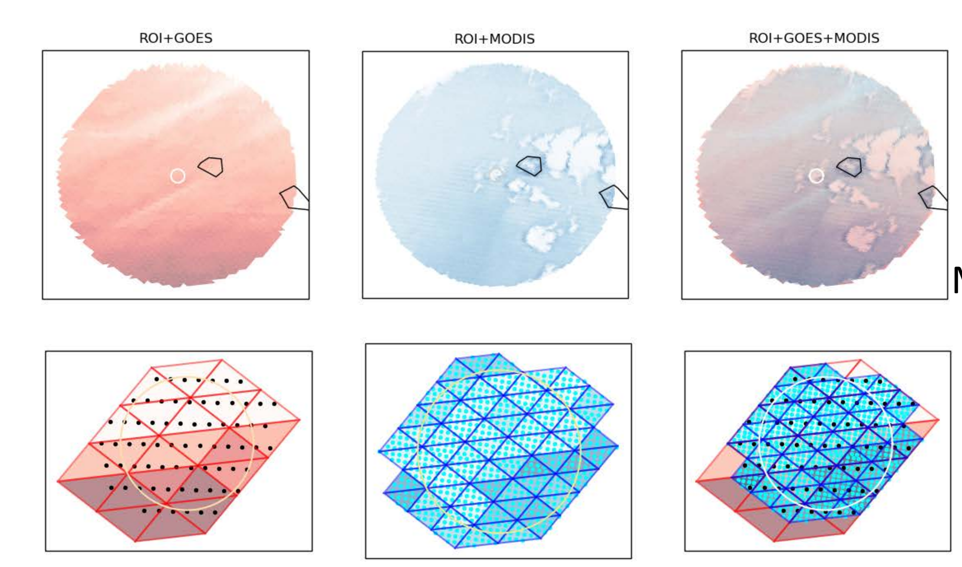


Distributed Coalignment

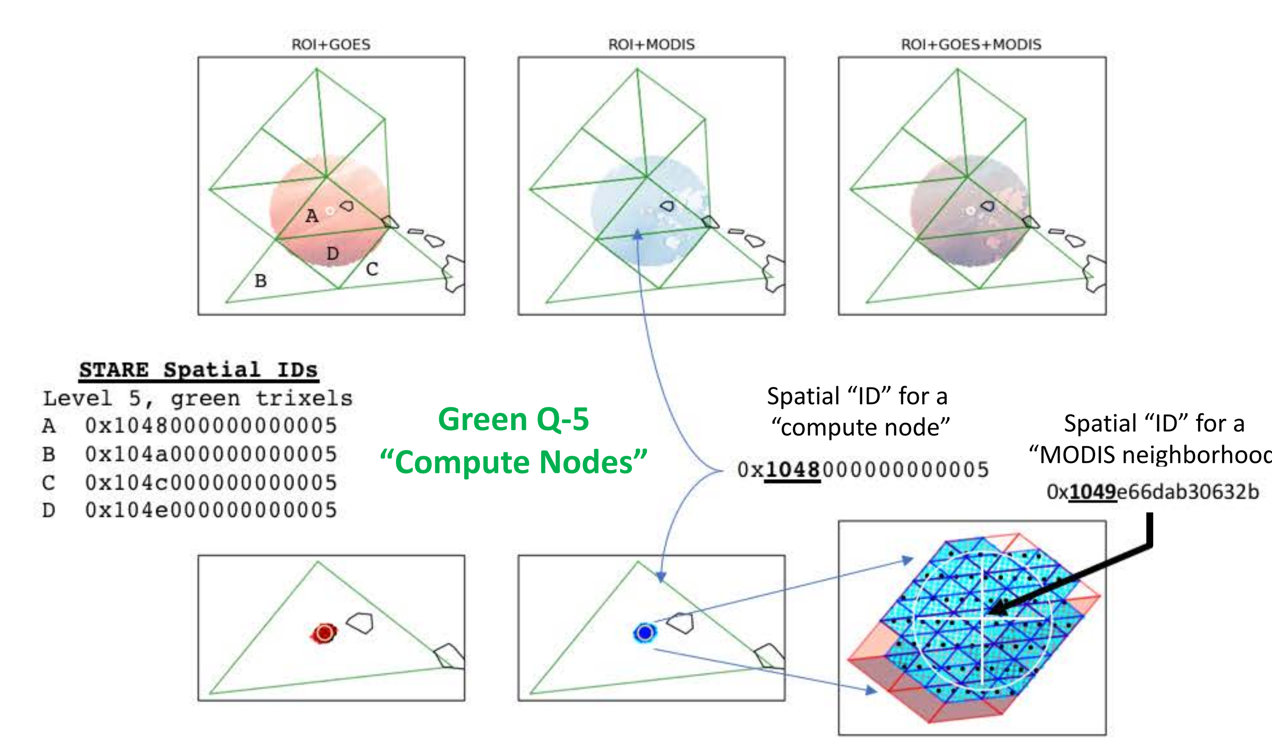
Distribute across 16 "nodes" or tasks.



Easy Low-Level Data Integration



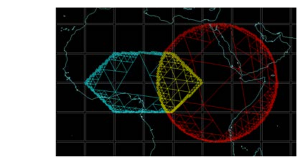
Spatial and Cyber Coalignment

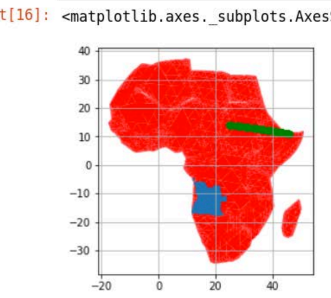


https://github.com/michaelleerilee/STARE

STARE-based Tooling

- ❖ C/C++ STARE Library
- ❖ PySTARE
- ❖ budding GeoPandas/Shapely support





Disaster Recovery in Africa (excerpt)
A STARE-based Jupyter Notebook showing STARE/HTM trixels for nations (red), Angola (blue), and a Tropical Rain Mapping Mission (TRMM) granule (green). UCSB.

STARE Use Cases Under Study

- ❖ OPeNDAP
- ❖ UCSB - Multi-sensor snow cover MODIS/VIIRS/LANDSAT/GOES
- ❖ LAADSWEB/MODAP – MODIS, VIIRS
- ❖ PMM IMERG – Microwave radiometry
- ❖ Tropical land surface – MODIS/VIIRS & LANDSAT TM colocation
- ❖ ICESAT/MODIS – Colocation over poles
- ❖ CLARREO – Cross calibration

STARE Future/Take-away

- ❖ STARE is maturing, becoming usable for combining diverse data by early adopters.
- ❖ A growing variety of functions in convenient APIs are being developed.
- ❖ STARE provides a foundation for searching, processing, and packaging data on distributed/parallel systems and Cloud.
- ❖ Improvements including radial indexing, alternative root polyhedron, and non-spherical bodies are in study.

Acknowledgement.
NASA ACCESS program provided partial support to this effort, award ID 80NSSC18M0118.