AIST Technology Perspectives for ESO Mission Processing and Open-sourced Science

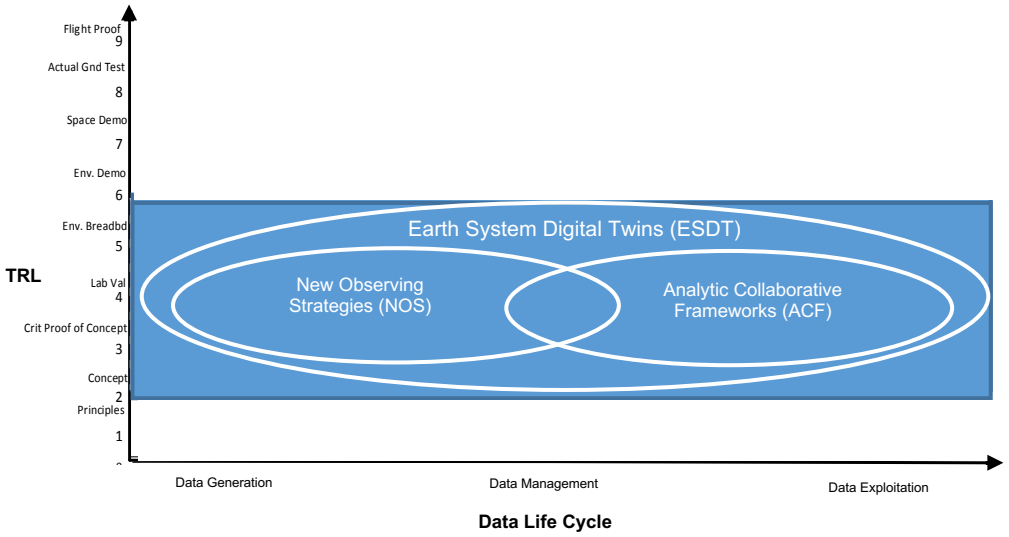
ESO Mission Data Processing Workshop. October 19, 2021

Jacqueline Le Moigne | *AIST Program Manager*
Ben Smith | *AIST Associate*
NASA Earth Science Technology Office

Advanced Information Systems Technology Program (AIST)

AIST Program Objectives

AIST identifies, develops, and supports adoption of software and information systems, as well as novel computer science technologies expected to be needed by the Earth Science Division in the 5-10-year timeframe.



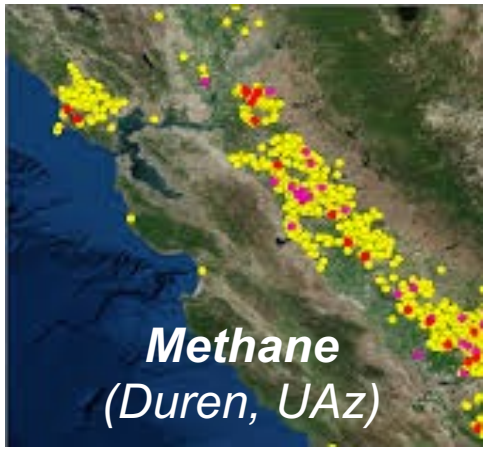
Three Thrust Areas

New Observing Strategies (NOS) dynamically coordinate and collaborate observations across multiple platforms (space, air, ground) to acquire a more complete picture of Earth Science phenomena.

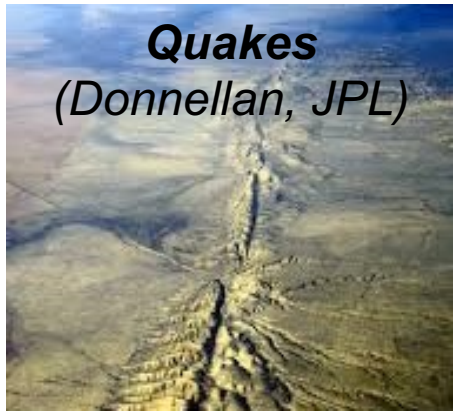
Analytic Collaborative Frameworks (ACF) address the challenges associated with an increased variety and volume of data over various geographical scales, latencies, and frequencies, and which facilitate access, integration, and understanding of large amounts of disparate datasets.

Earth System Digital Twins (ESDT) are *interactive* and *integrated* multidomain, multiscale, digital replica of the state and temporal evolution of Earth Systems.

AIST ACF Investments support Earth Science Disciplines across ESO



SBG

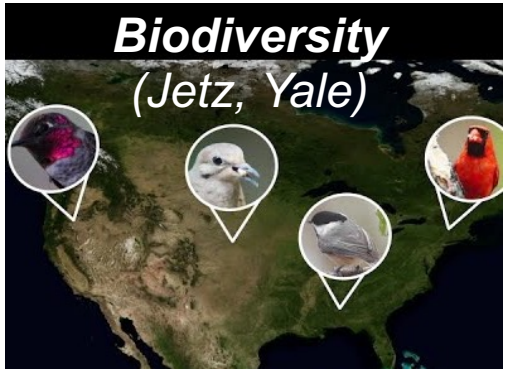


SDC, STV

ACCP

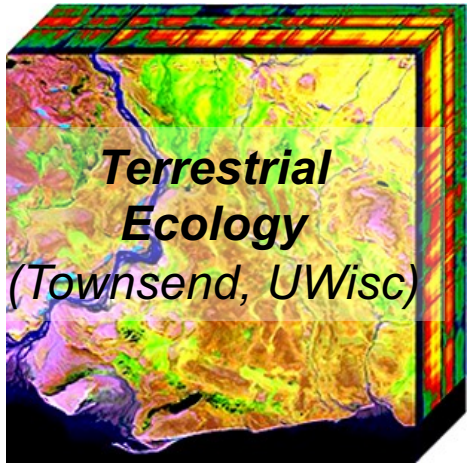
ACCP

ACCP



SBG

ACCP



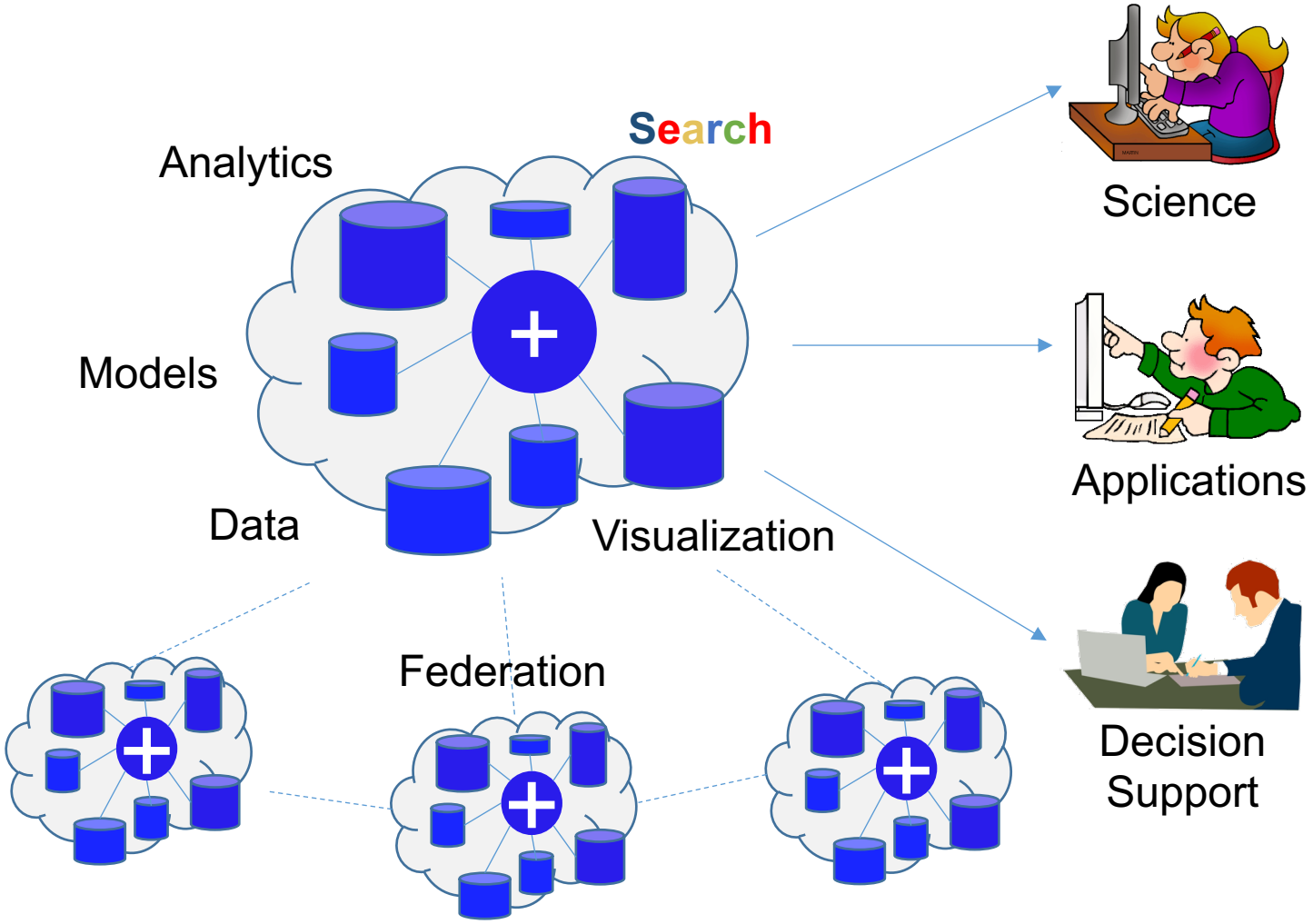
SBG

Analytic Collaborative Frameworks

Integrated data, analytics, and tools tailored for a science discipline, typically as a hosted service or portal.

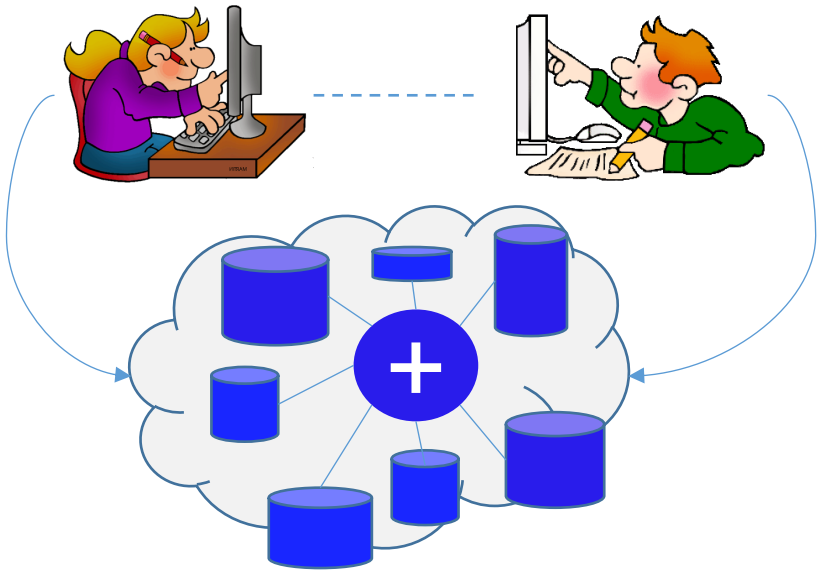
Facilitates collaborative science across missions and data sets

ACFs can be federated into larger collaborative networks.



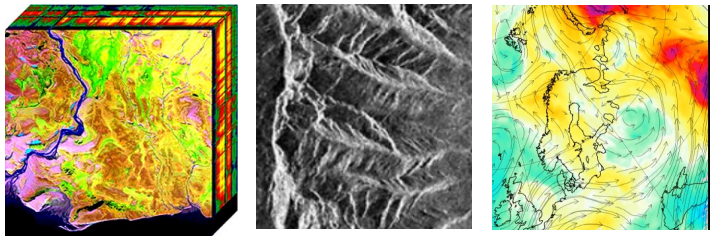
ACFs: A Vision for Mission Data Processing and Open-Sourced Science

Open-sourced Science

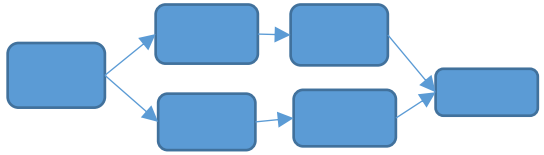


- Collaborative analysis environment
- Shared data, tools, visualization.
- Sharable workflows and products.
- Workflow provenance

Low-latency Data Processing

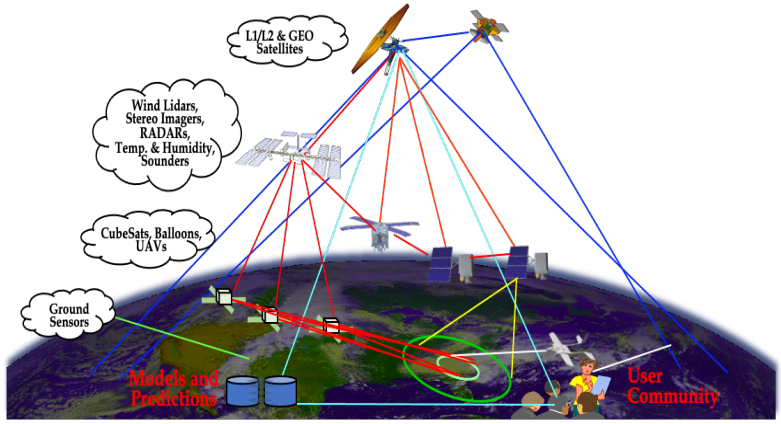


Hyperspectral SAR Models



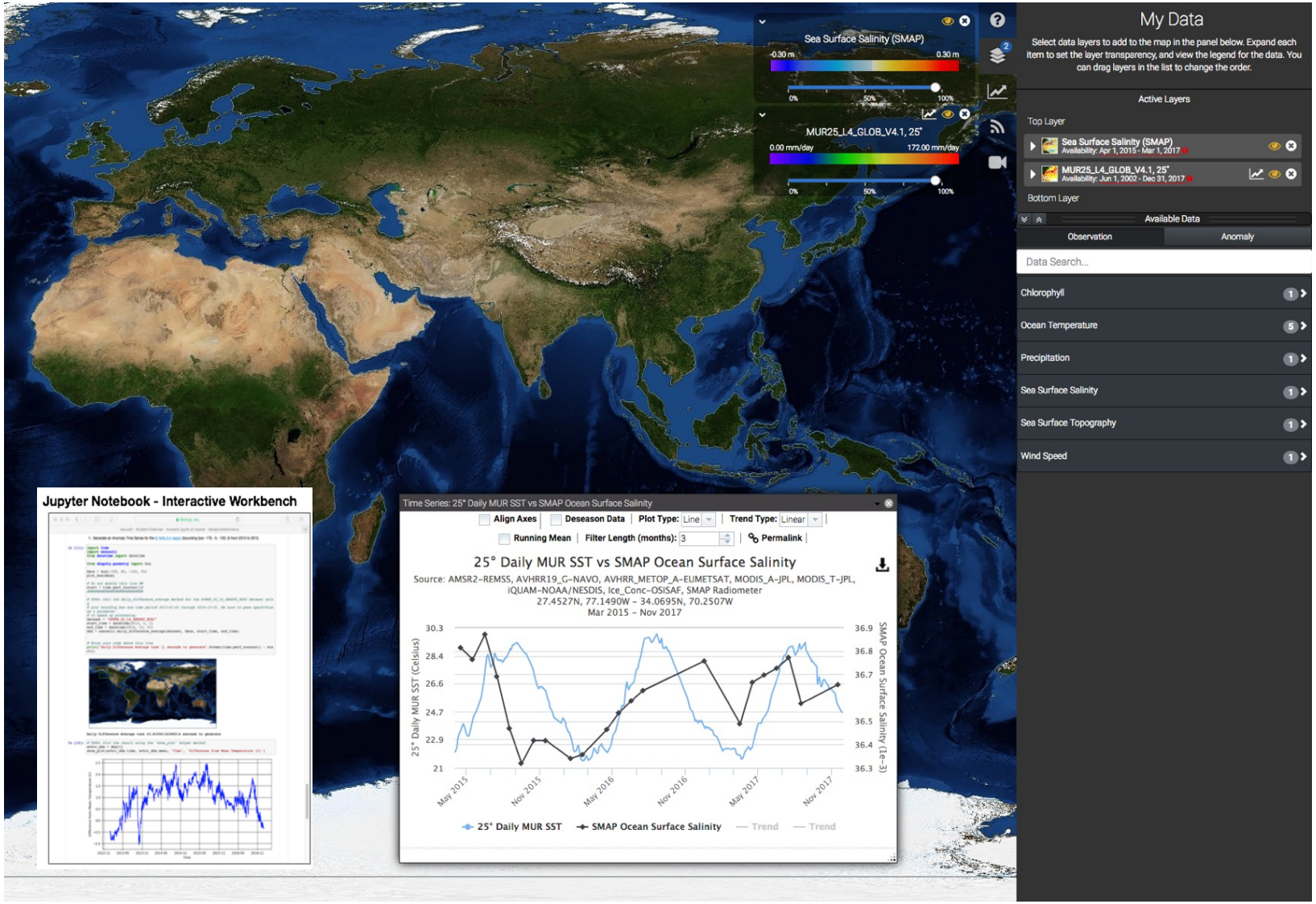
- High-performance, scalable data processing workflows
- Hybrid cloud architectures
- Model acceleration
- AI/ML analytics

Analyze Multiple Data Sources



- Analyze data across instruments and observing platforms.
- Data fusion over temporal and spatial scales.
- Federate with related projects.

Example: AIST OceanWorks
An Analytic Center Framework for Ocean Science



Dozens of Ocean Data Sets
In the cloud, integrated, ready to go

Cloud-based analytics. Analyze years of data over multiple data sets in seconds… without downloading data.

Custom analytics Scientists can also run custom tools and algorithms

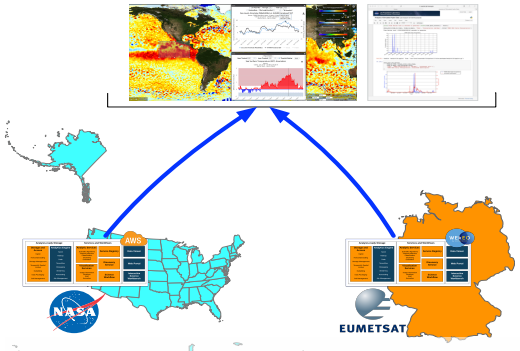
Integrated data. Match up in-situ and remote sensing data, despite differences in scale and resolution.

Search. Find relevant data sets

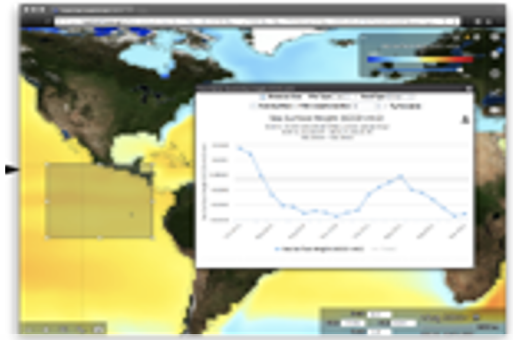
Visualization. Subsets, layers, animations. Integrates with ArcGIS and Jupyter Notebooks.

https://oceanworks.jpl.nasa.gov

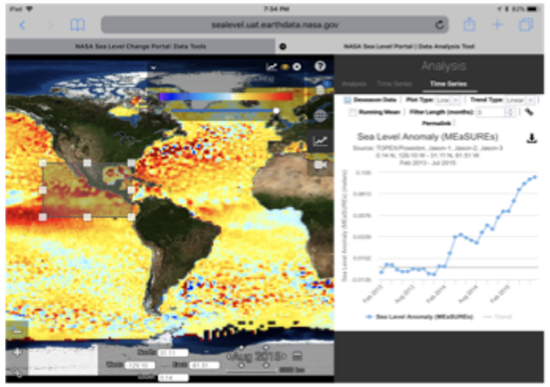
Open-source facilitates infusion and adoption



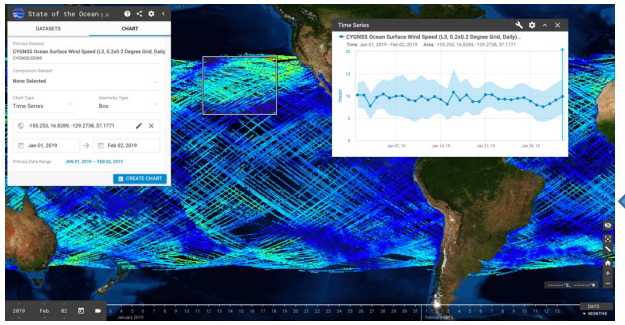
CEOS COVERAGE
Distributed Analytic Architecture



ECCO Ocean and Ice State
NASA ACCESS Program

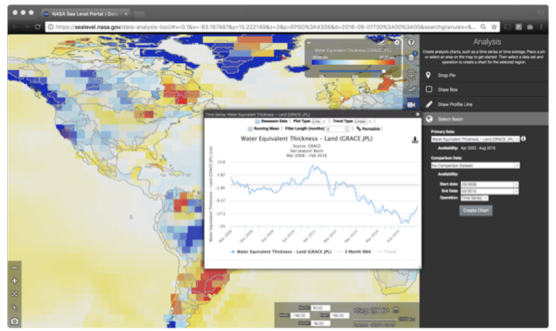


Sea Level Change Portal
https://sealevel.nasa.gov/data-analysis-tool



PO.DAAC State of the Ocean

Science Data Analytic Platform
An open-source ACF platform



Grace Follow-on
https://grace.jpl.nasa.gov/data-analysis-tool

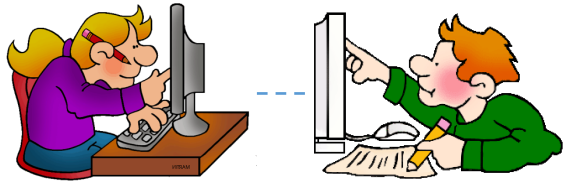
Federating Analytic Frameworks

ACFs can be federated to share data, tools, and models among related disciplines and missions.

Facilitates cross-discipline analyses with existing ACFs.

Common APIs and data representations enable federation.

Wildfires



Precipitation

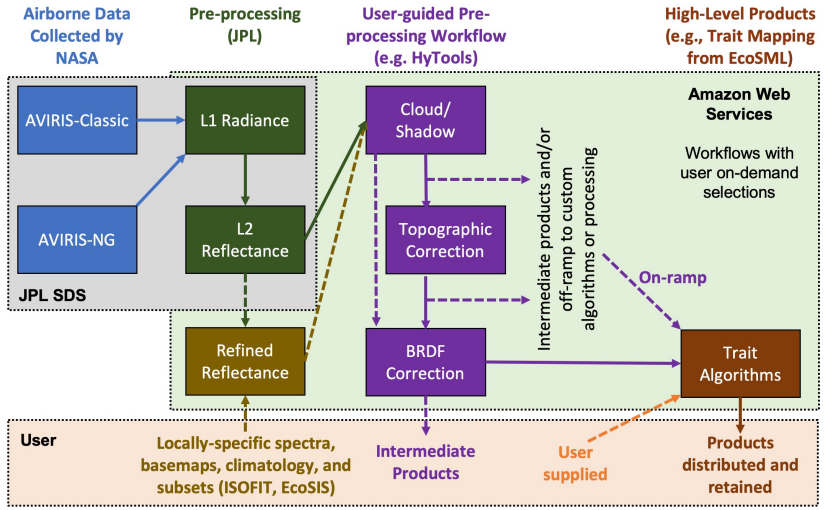
Air Quality

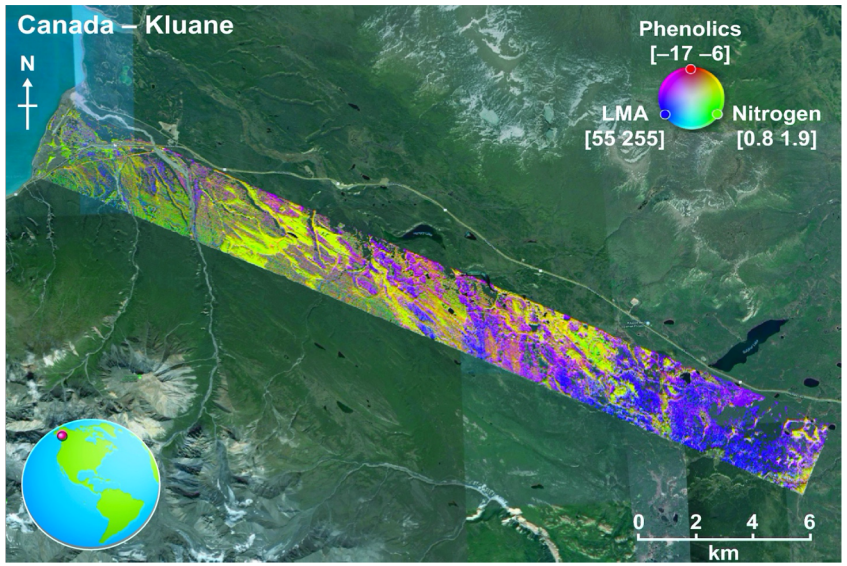
ImgSPEC

- Prototype a science data system to satisfy unprecedented data volume and user/algorithm diversity in NASA's Surface Biology and Geology (SBG) mission.
- Produce **End-to-end on-demand cloud framework for imaging spectroscopy** Level 1 calibrated radiance through Level 3+ products including atmospheric and surface reflectance retrieval, bidirectional reflectance distribution function (BRDF) correction, topographic correction, L3 retrieval algorithms, mosaicking and analytics.
- Basis for SISTER science data processing prototype

Scaling ACFs to massive hyperspectral datasets and diversity of analytic workflows.

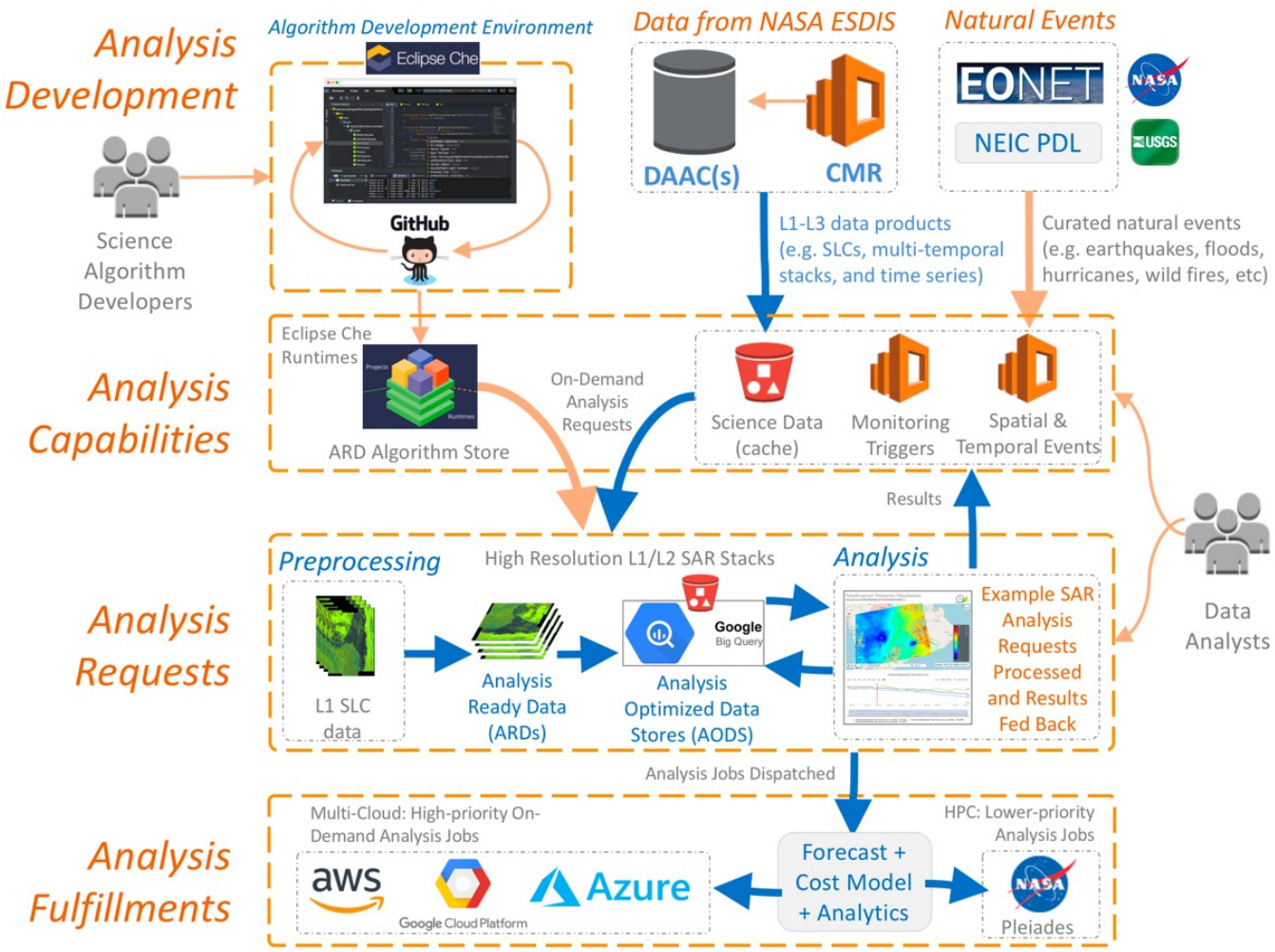
PI: P. Townsend, *Univ Wisconsin*





Schimel, Schneider et al. 2019, *New Phytologist*

Multi-Cloud Processing of SAR ARDS

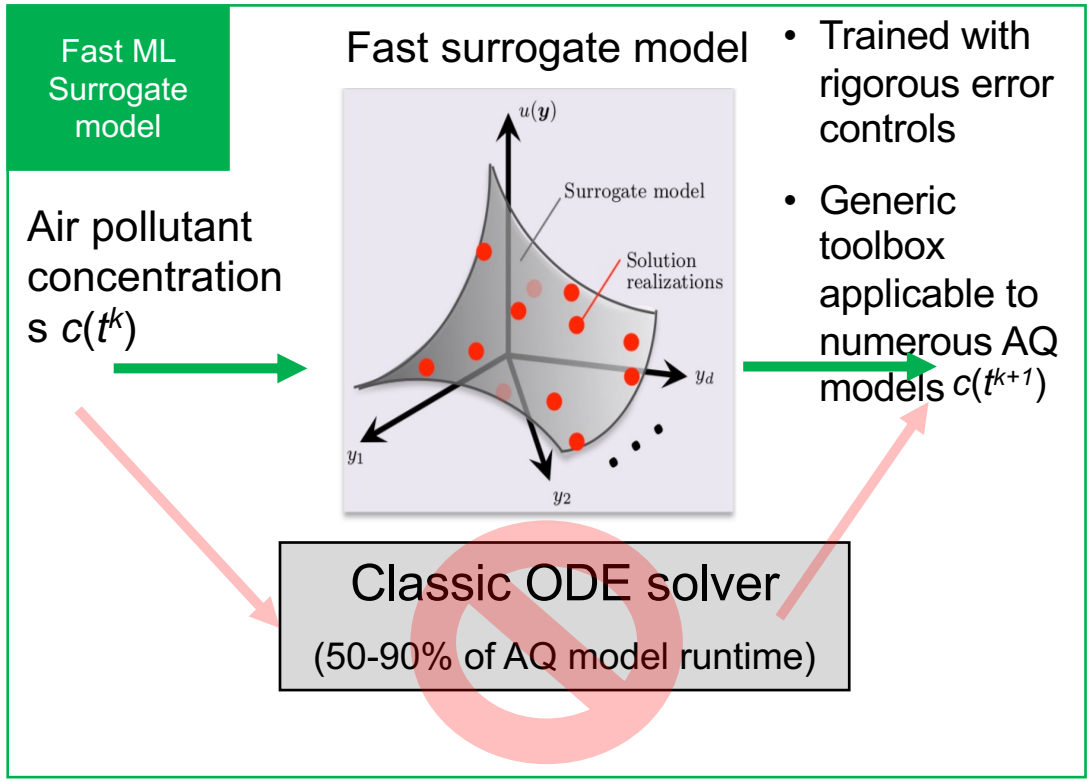


- Parallel Cloud processing framework for SAR data.
- Minimizes costs by distributing jobs over multiple cloud providers, on-premesis, and NASA HPC.
- Enable **algorithm development** and deployment **at scale**.
- Execute analysis *runtimes* as on-demand processes distributed across **multi-cloud** (AWS, Google Cloud Platform, and Microsoft Azure) and **NASA HPC** (Pleiades) environments.

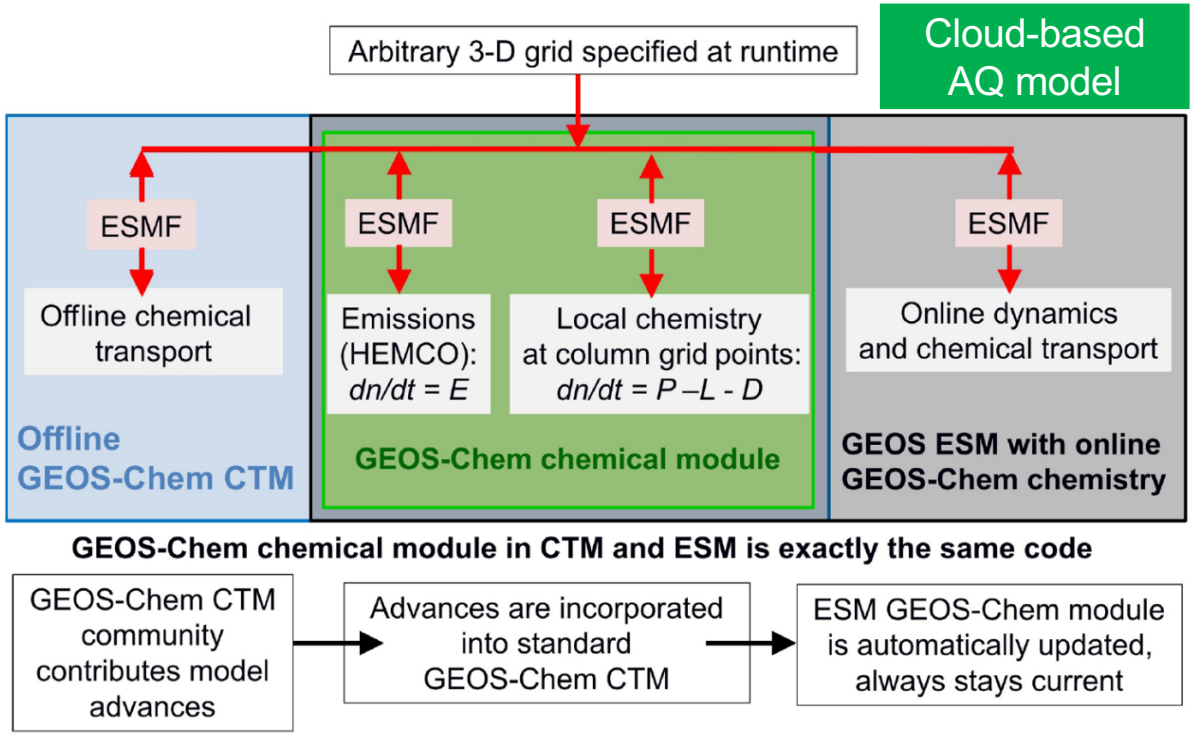
Multi-cloud architectures for ACFs

PI: H. Hua, *Jet Propulsion Laboratory*

ML-Accelerated and Cloud-based Models (GeosCHEM)



ML surrogate learns what the chemical-transport model would output but runs much faster once trained. Speeds up a core element of GeosCHEM.



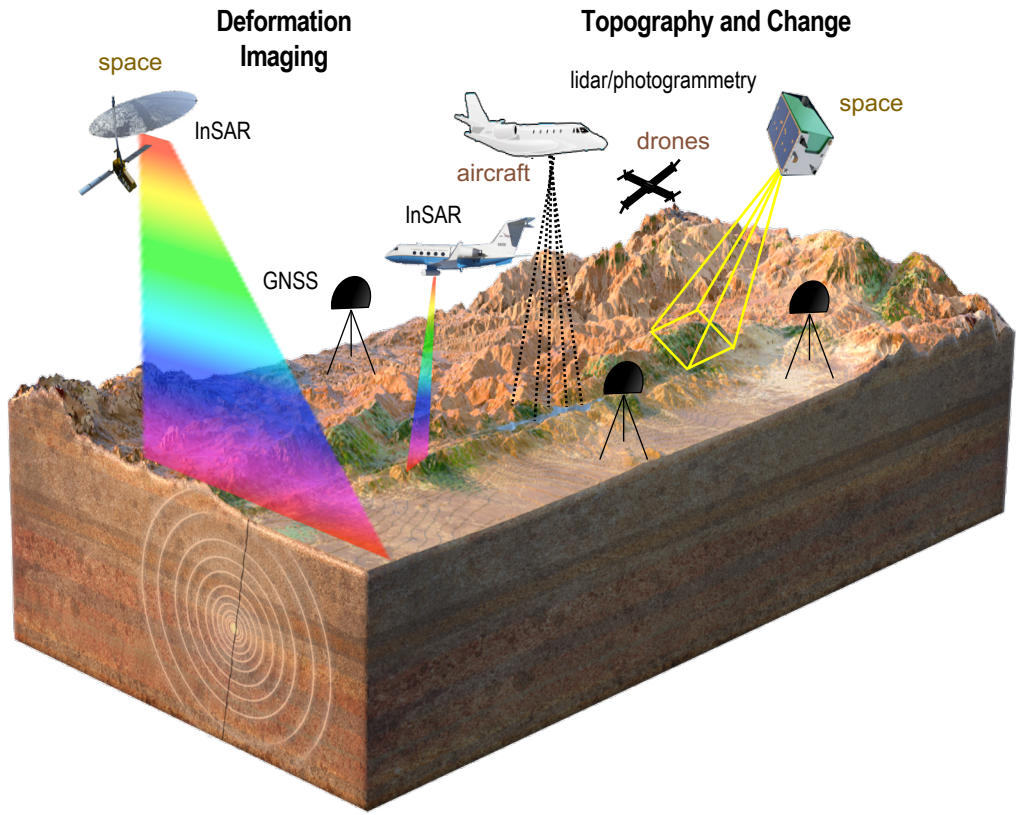
Cloud-based, optimized version of GeosCHEM. Allows model to run within cloud-processing ACF frameworks and exploit parallelism to accelerate workflows.

Models for ACFs

PI: D. Henze, *Univ. of Colorado Boulder*

PI: R. Martin, *Washington University in St. Louis*

Quakes-A:
An Analytic Framework for Quakes



- **Fuse** InSAR, topographic, and GNSS geodetic imaging data
- **Quantify uncertainties** for the reference model
- **Improve** earthquake forecast **models**
- **Improve understanding** of the physical processes leading to and following earthquakes

Analytic center will harmonize crustal deformation data from orbital, airborne, and in-situ sensors

Key product is a uniform crustal deformation reference model for the active plate margin of California

Fusion of space, air, and in situ data.

PI: A. Donnellan, *JPL*

Guidance and Recommendations

- **AIST projects offer potential for transition / infusion to ESO mission processing and open-sourced science**
- Analytic Center Frameworks can **facilitate open-sourced science**
  - Science users can **share workflows, tools, and data** tailored to needs of the ESO science community
  - **Collaboration**: visualization, ease of use, sharable results
  - **Open-source platforms** facilitate adoption and evolution
- ACF technologies **accelerate generation of information products** at scale
  - **Scalable parallelization** through cloud and HPC processing and hybrid cloud architectures
  - Consider **model acceleration** via surrogates and scaling to cloud architectures
  - Specific technologies for computationally intensive processing like SAR and hyperspectral
- Emerging technologies for **analysis across instruments and scales**
  - **Multi-scale and multi-instrument fusion** workflows in development for certain ESO disciplines
  - Cloud-based ACFs manage **differences in scales and representations**.
  - **Federation across projects** and science data processing systems

ESTO

Earth Science Technology Office

AIST-18: Current ACF Projects

PI's Name	Organization	Title
Uz	NASA GSFC	Supporting shellfish aquaculture in the Chesapeake bay using AI for water quality
Moisan	NASA GSFC	NASA Evolutionary Programming Analytic Center (NEPAC)
Jetz	Yale U.	Biodiversity - Environment Analytic Center
Townsend	U. Wisconsin, Madison	GeoSPEC: On-Demand Geospatial Spectroscopy Processing Environment on the Cloud
Swenson	Duke University	Canopy condition to continental scale biodiversity forecasts
Ives	U. Of WI, Madison	Valid time series analyses for satellite data
Beck	U. Of AL, Huntsville	CAPRi: Cloud-based Analytic Framework for Precipitation Research
Zhang	SMU	Mining Chained Modules in Analytics Center Framework

PI's Name	Organization	Title
Martin	Washington U.	Development of GCHP to enable broad community access to high-resolution atmospheric composition modeling
Duren	NASA JPL	Multi-scale Methane Analytic Framework
Henze	U. of CO, Boulder	Surrogate modeling for atmospheric chemistry and data assimilation
Holm	City of Los Angeles	Predicting What We Breathe: Using Machine Learning to Understand Urban Air Quality
Donnellan	NASA JPL	Quantifying Uncertainty and Kinematics of Earthquake Systems ACF (QUAKES-A)
Hua	NASA JPL	Smart On-Demand of SAR ARDs in Multi-Cloud & HPC
Huffer	Lingua Logica	AMP: An Automated Metadata Pipeline

https://esto.nasa.gov/project-selections-for-aist-18/