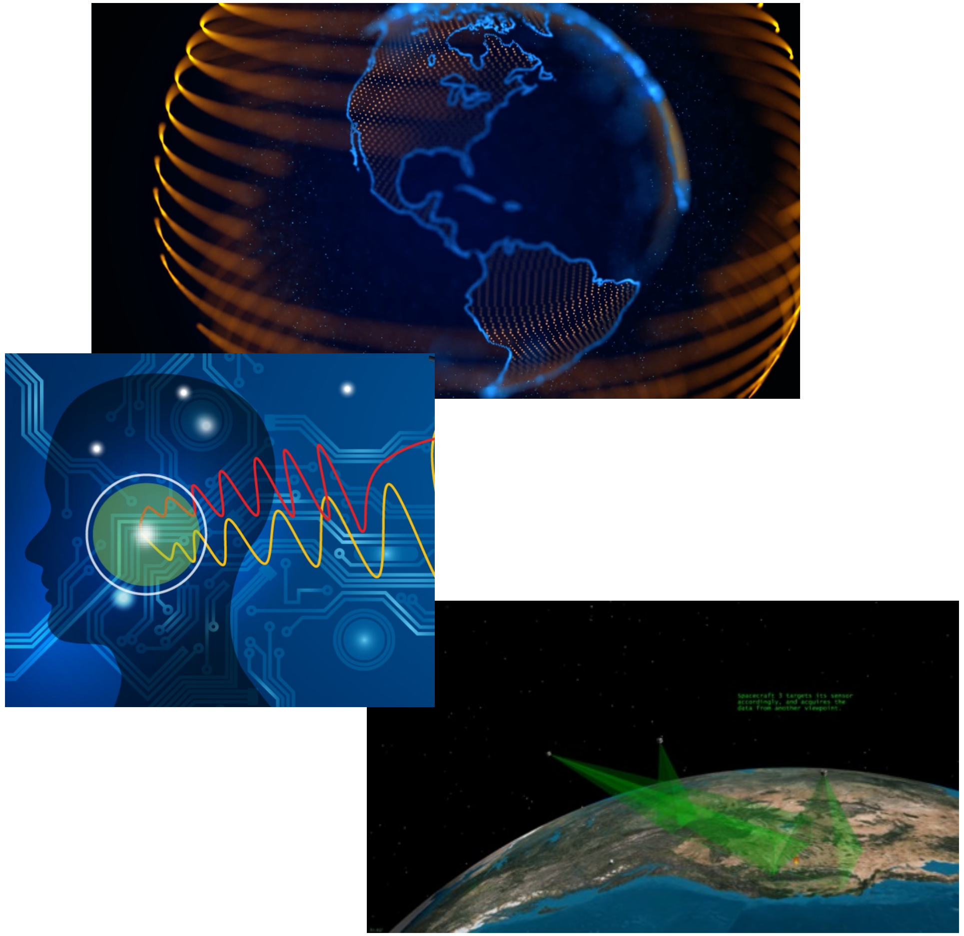


Artificial Intelligence for Future Earth Science Information Systems and Earth System Digital Twins

Jacqueline Le Moigne
NASA Earth Science Technology Office

March 21, 2023

AIST Objectives

Innovate in software and information systems technology that enable:

O1. New observation measurements and new observing systems design and operations through intelligent, timely, dynamic, and coordinated distributed sensing
=> Novel Observing Strategies (NOS)

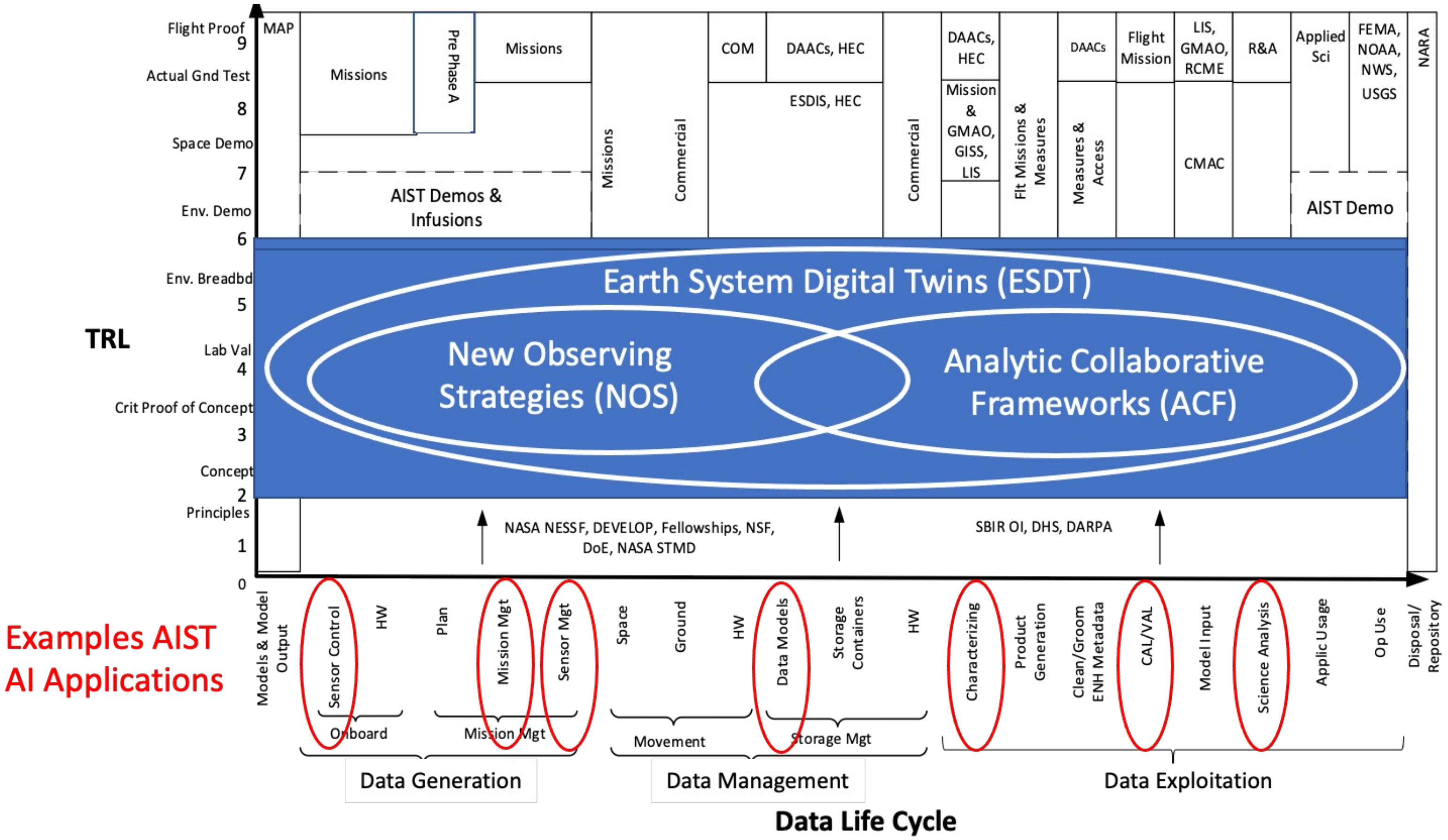
O2. Agile science investigations that fully utilize the large amount of diverse observations using advanced analytic tools, visualizations, and computing environments, and that interact seamlessly with relevant observing systems
=> Analytic Collaborative Frameworks (ACF)

O3. Developing integrated Earth Science frameworks that mirror the Earth with state-of-the-art models (Earth system models and others), timely and relevant observations, and analytic tools. This thrust will provide technology for enabling near- and long-term science* and policy decisions
=> Earth System Digital Twins (ESDT)

* "Science decisions" including planning for the acquisition of new measurements; the development of new models or science analysis; the integration of Earth observations in novel ways; applications to inform choices, support decisions, and guide actions for societal benefit; etc.

AIST Program Scope

NASA's Advanced Information Systems Technology (AIST) Program
identifies, develops, and supports adoption of advanced 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



AI in AIST for Earth Science Applications

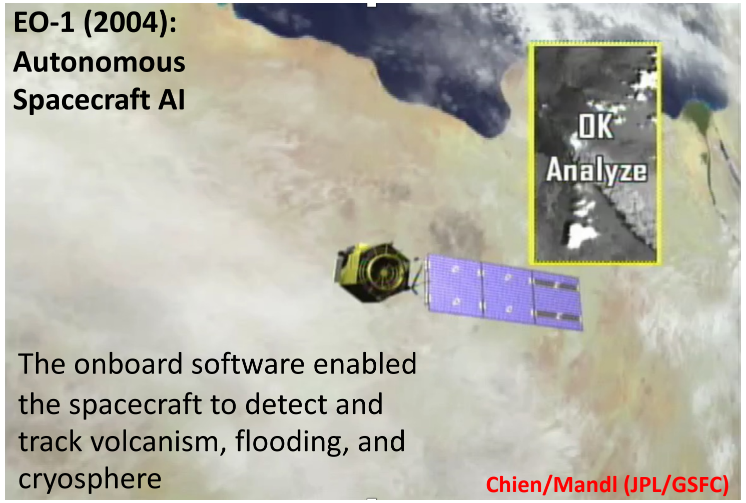
Two Main Areas

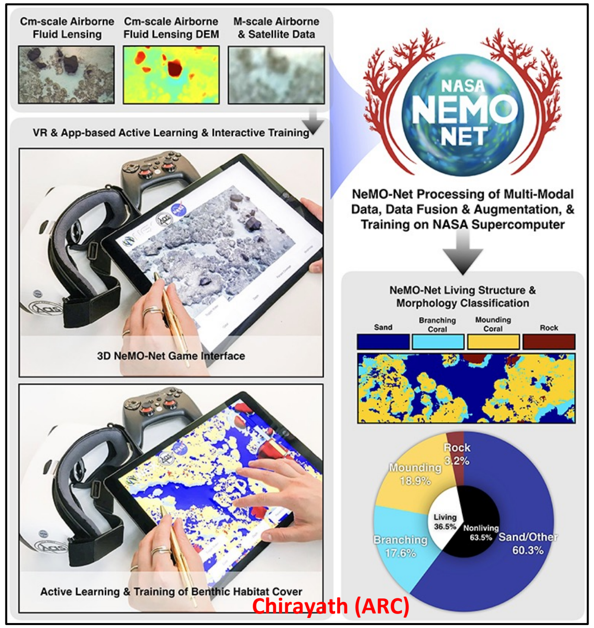
- **Improved Agile Observation Coordination and Mission Operations (Onboard or on the Ground) – *NOS***
  - At the edge intelligent data analysis
  - Semi-autonomy and autonomy for Distributed Spacecraft Mission (DSM) coordination and decision making
  - Anomaly and fault detection
  - Engineering Support for large constellations
  - Advanced Interoperability between sensor and modeling nodes

Technologies: Smart Sensors, Planning & Scheduling, Intelligent Agents, Cognitive and Knowledge-Based Systems, Reasoning, …

- **Science Discovery and Advancement – *ACF and ESDT***
  - Multi-source data discovery and integration (also related to NOS)
  - Big data analytics: discover correlations in large amounts of data
  - Improvements and support to forecasting, simulation, science modeling and data assimilation
  - Advanced Visualization

Technologies: Machine Learning/Deep Learning, Intelligent Search, Computer Vision, Data Fusion, Surrogate Modeling, Interactive Visualization & Analytics, Natural Language, …



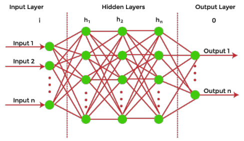


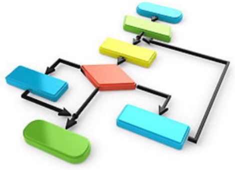
Artificial Intelligence and Machine Learning for Novel Observing Strategies (NOS)

Novel Observing Strategies (NOS)

... or How to create an Internet-of-EarthThings (IoET)

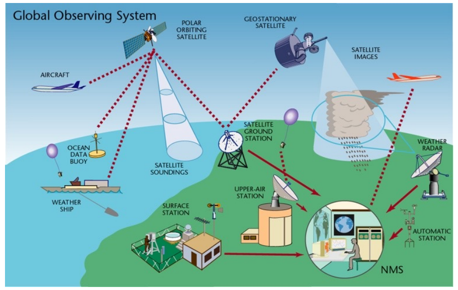
Resources





Data | Sensors & Platforms | Models | Algorithms | IoT & Social Media

An Observation System is a configuration of resources to answer science questions in order to advance knowledge and benefit society.



Observation System

Novel Observing Strategies (NOS)

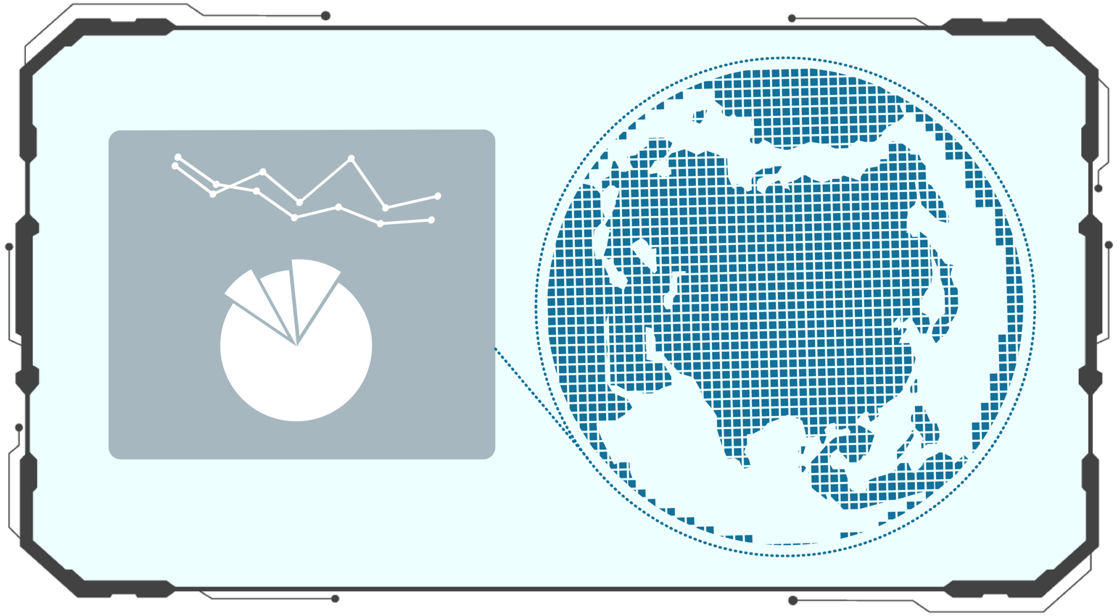
... or How to create an Internet-of-EarthThings (IoET)

OBJECTIVES:

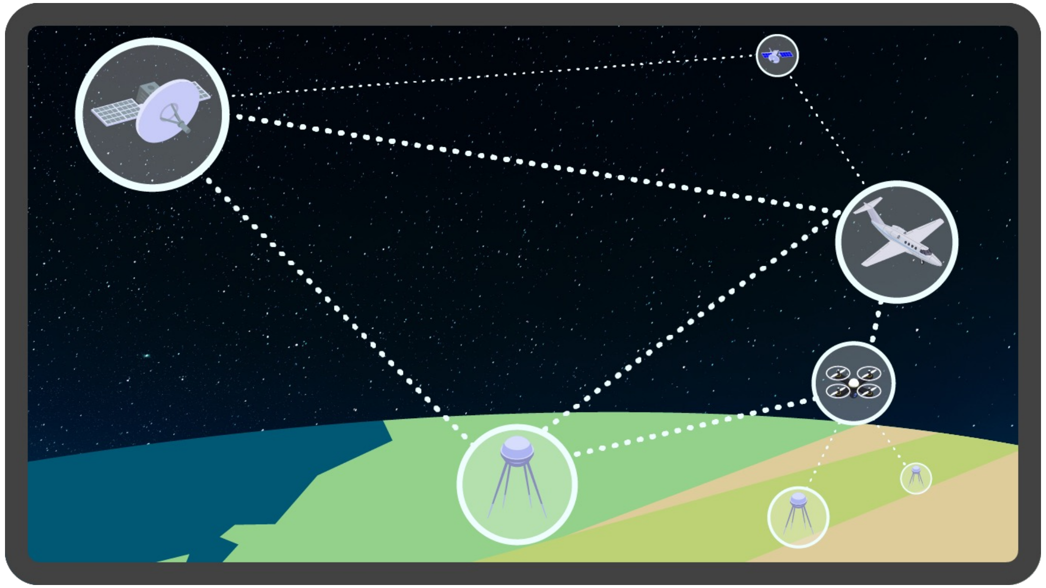
1. **Design and develop New Observing Concepts:**
   - From Decadal Survey or Model; **Various size spacecraft;** Systems of systems **(*Internet-of-Space*); Various organizations**
   - **Perform trades** on sensor number/type, spacecraft, orbits; resolutions; onboard vs. on-the-ground computing; inter-sensor communications, etc.
   - System being **designed in advance** as a mission or observing system **or incrementally and dynamically over time**
2. **Respond to various science and applied science events of interest:** Various overall observation timeframes; Various area coverages; Dynamic/Timely; Scheduling, re-targeting/re-pointing assets, as possible

System-of-Systems NOS-Testbed for technologies & concepts validation, demonstration, comparison and socialization

Optimize current & future Mission Portfolio and Mission Design



Drive Coordinated, Event-Driven Observations



Novel Observing Strategies (NOS)

... or How to create an Internet-of-EarthThings (IoET)

Earth Observatory Optimization

"Hello NOS, I am the 2027 Earth Science Decadal Survey: I have received many science wishes for new science measurements. With a limited budget, how could I optimize the number and cost of satellites we will need to launch and yet maximize the number of measurements that will satisfy these science requirements?"

Coordinated Observations

"Hello NOS, I am the LIS Model: there was a fire in San Bernardino, CA, last month, and with all the rain lately, I am almost sure that there is going to be a flood in that area: can you ask Capella to retask their sensor so I can improve my prediction of which areas will be flooded next week?"

AI in AIST NOS Projects

AI for Mission Design and Observation Simulation Synthesis Experiments (OSSEs)

- A Mission Planning Tool for Next Generation Remote Sensing of Snow (Forman/AIST-16&18)
- TAT-C: Trade-space Analysis Tool for Constellations Using Machine Learning (LeMoigne-Grogan/AIST-16;QRS-20)
- SOS: A New Snow Observing Strategy in Support of Hydrological Science and Applications (Vuyovich/AIST-21)

AI for Coordinated Observations

- SPCTOR: Sensing Policy Controller and OptimizeR (Moghaddam/AIST-18)
- Edge Intelligence for Hyperspectral Applications In Earth Science for New Observing Systems (Carr/AIST-21)
- D-SHIELD: Distributed Spacecraft with Heuristic Intelligence to Enable Logistical Decisions (Nag/AIST-18)
- 3D-CHESS: Decentralized, Distributed, Dynamic and Context-aware Heterogeneous Sensor Systems (Selva/AIST-21)
- ILEOS: Intelligent Long Endurance Observing System (Chandarana/AIST-21)
- An Intelligent Systems Approach to Measuring Surface Flow Velocities in River Channels (Legleiter/AIST-21)

AI for Onboard Processing and Smart Sensors

- ISS Spaceborne-2 Experiments (Chien/QRS-21)
- Innovative Geometric Deep Learning (GDL) Models for Onboard Detection of Anomalous Events (Gel/AIST-21)
- VIDEO: Sensor-in-the-Loop Testbed to Enable Versatile/Intelligent/Dynamic Earth Observation (Blackwell/AIST-21)
- AI for CASALS and Multi-Path Fusion Machine Learning for New Observing System Design and Operations (MacKinnon/AIST-21)

AI in NOS-Testbed (NOS-T) Demonstrations

- NOS-T Hydrology Demonstrations (2020)
- NOS-L (NOS-Live): Autonomous, Model-Driven Tasking of Flood Events (2022)

SpaceBorne-2 Experiments/Chien (NASA JPL) –
ISS Onboard Processing Experiment

This project conducts both technology validation and TRL improvement experiments as well as it will demonstrate enhanced and enabling capabilities.

New Technology Demonstrations: *Validating New Hardware and Software Technologies*

- **Re-Tasking Demonstration :** using onboard data analysis to create alerts
    - Use onboard data analysis to generate alerts, NOS- or SensorWeb-like, e.g,: task other assets
- **Live-Instrument Data Feed :** run experiments with data generated on ISS, e.g., ECOSTRESS, EMIT, OCO-3
    - Using both pre-uploaded and potentially live data from onboard ISS instruments
- **Co-Processors Experiments:**
    - **Intel Movidius/Myriad Neuromorphic Processor**
        - Currently on ESA's Phi-Sat and terrestrial drones
    - **Qualcomm Snapdragon Processor**
        - Flying on Mars Helicopter
    - Gain tremendous in-space processing experience with 2 processors that are well on the path to mission use

Data Processing and Machine Learning Experiments

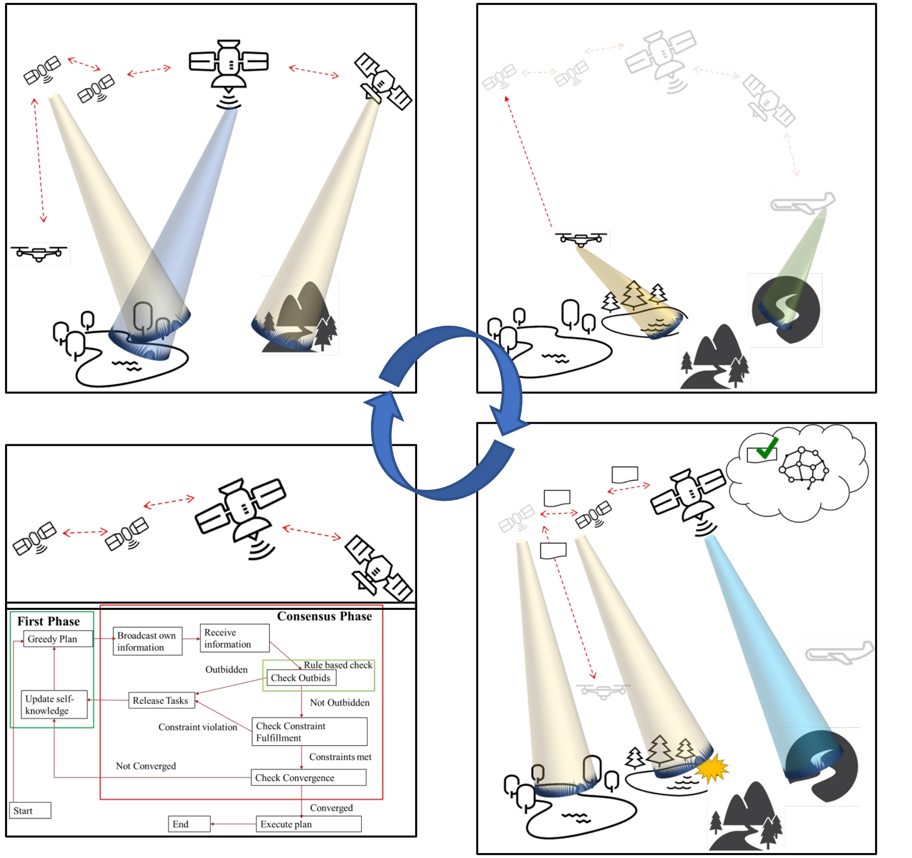
- **Radar Processing: l**everage NISAR and UAVSAR radar pipelines => data reduction, low-latency downlink
- **Thermal Infrared Processing :** experiment with TIR data from ECOSTRESS
    - Onboard pipeline: radiometric calibration, geolocation, land surface temperature, etc.
    - Applied Science Value: orders of magnitude data reduction, low-latency downlink
- **VSWIR Processing :** experiment with VSWIR data from EMIT
    - Heritage technology from HysPIRI Intelligent Payload Module (IPM)
    - Applied Science Value: data reduction, low-latency downlink, alerts for tasking other sensors
- **Machine Learning Demonstration** : perform ML/imagery classification techniques like HiRISNet, MSLnet, and Hirise

AIST-21/D. Selva (Texas A&M University) –

3D-CHESS: Decentralized, Distributed, Dynamic, and Context-aware Heterogeneous Sensor Systems

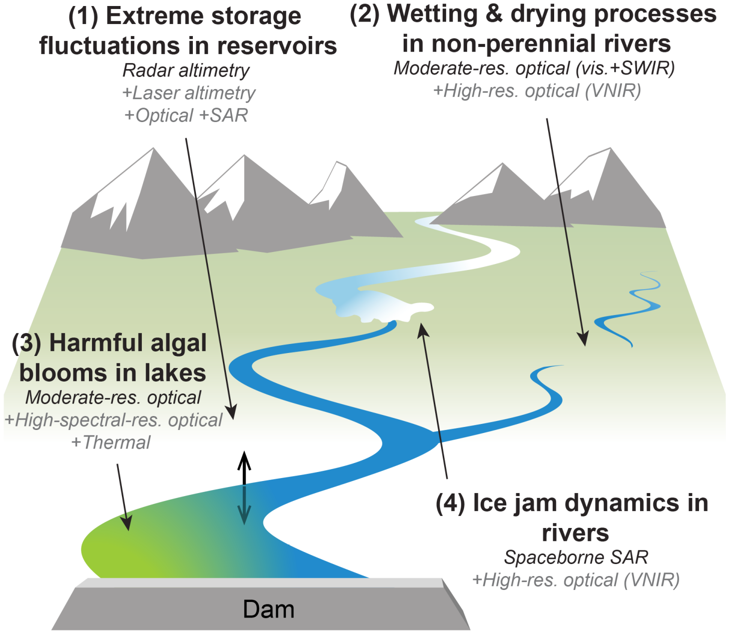
***Proof of concept** for a context-aware Earth observing **sensor web** of interconnected space, air and ground nodes.*

#1 – Default Missions **#2 – Event Detected**



#4 – Decentralized Planning **#3 – Instrument Capability Reasoning**

1. **Default Missions – Data Processing**
   Each node conducts its nominal mission, following a default plan.
2. **Event Detected – Task Request Sent**
   One node detects (or predicts) an event of interest at a certain location. It sends a Task Request to the network.
3. **Instrument Capability Reasoning**
   Nodes receive the task request and use contextual information from their knowledge base and the reasoning algorithm to autonomously determine if they can perform this task.
4. **Decentralized Planning**
   Nodes enter a decentralized planning phase to coordinate The tasks. Nodes use a science-driven utility function to determine if they should perform the task at the cost of temporarily abandoning their current mission. Engineering costs (e.g., energy) are also considered. Once converged, the nodes update their plans.
5. **Updated Plans – Iterate**



Technology feasibility and value will be demonstrated in a multi-sensor in-land hydrologic and ecological monitoring system with 4 inter-dependent objectives.

Main AIST Technologies Needed for NOS

e.g., AI for Autonomy

- Situational Awareness
- Planning / Scheduling
- Node Identification & Discovery
- Intelligent & Dynamic Targeting and/or Tasking
- Communications Standards & Protocols
- Information & Data Fusion

Some Examples of Capabilities Needed Onboard:

- Recognizing science events of interest
- Analyzing data for optimal science return
- Exchanging data inter-spacecraft
- Reconfiguring the spacecraft based on coordinated observations

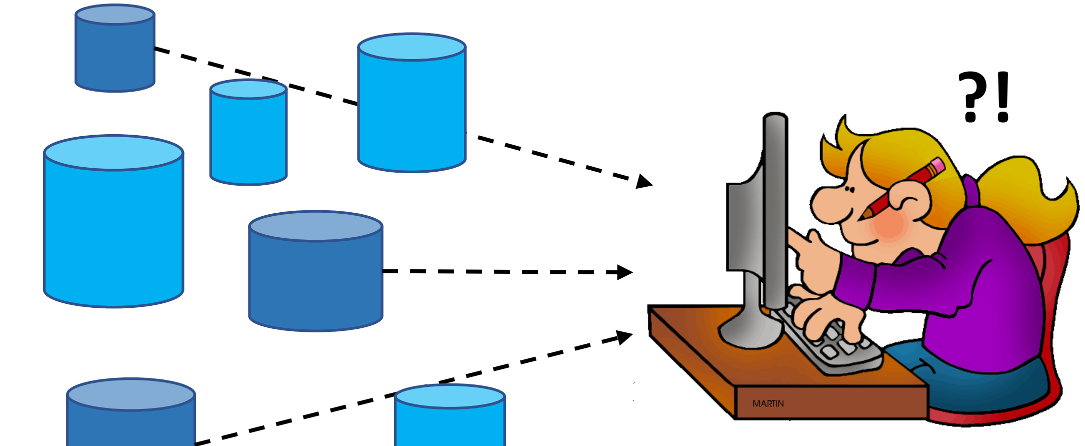
Machine Learning for Analytic Collaborative Frameworks (ACF)

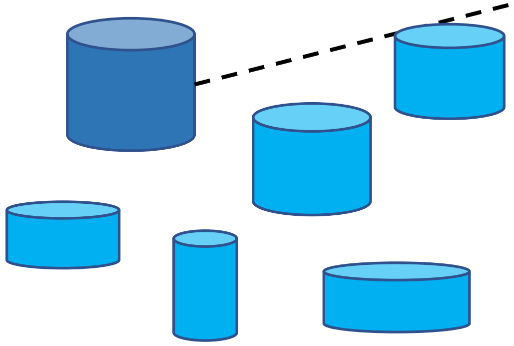
From Archives to Analytic Frameworks
Focus on the Science User

Data Archives

Focus on data capture, storage, and management

Each user has to find, download, integrate, and analyze





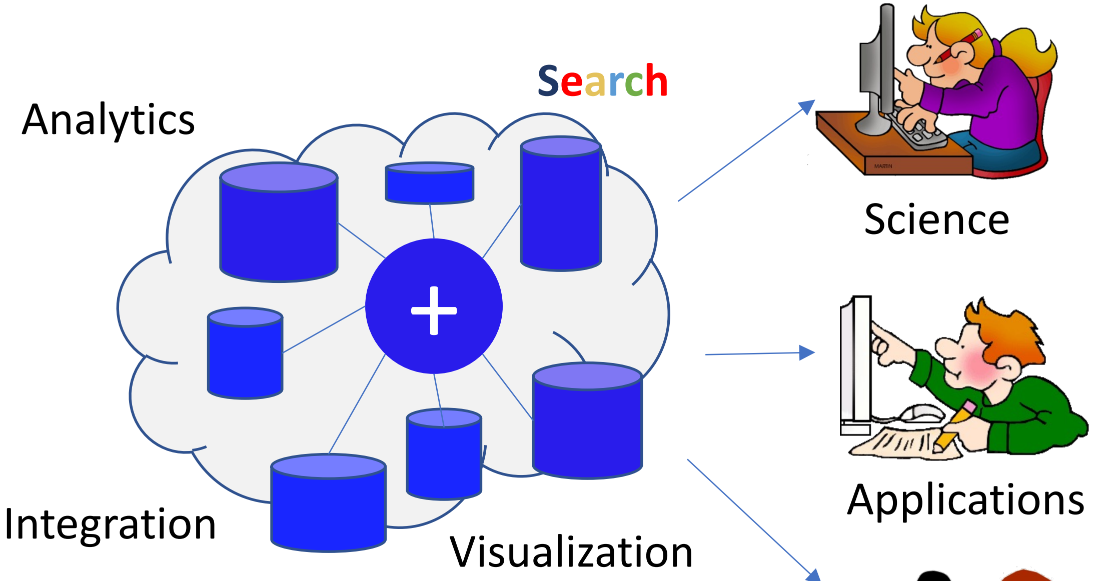
80% **Prepare**

20% Analyze

Analytic Centers

Focus on the science user

Integrated data analytics & tools tailored for a science discipline



20% Prepare

80% **Analyze**

Facilitates collaborative science across multiple missions and data sets

Analytic Collaborative Frameworks (ACF)

Focus is on the Science User

NASA
ESTO
Earth Science Technology Office

Allow flexibility/tailor configurations for Science investigators to choose among a large variety of datasets & tools

User

- Project Definition
- Plan for Investigation

Data

- Catalog
- NASA DAAC
- Other US Govt
- Non-US
- Local or non-public

Project Work Environment

Tools

- Discovery & Catalog
- Work Management
- Data Interfaces
- Analytic Tools
- Modeling
- Collaboration
- Visualization
- Sharing/ Publication
- Local/custom

Reduce repetitive work in data access and pre-processing, e.g., develop reusable components

Computational Infrastructure

- Computing
  - Capacity
  - Capability
- Storage
- Communications

Storage

- Data Containers
- Thematic model
- Metadata/Ontology
- Resulting Products
- Published data
- Provenance

Computing

- Local systems
- High End Computing
- Cloud Computing Capability
- Quantum Computing
- Neuromorphic Computing

AI in AIST ACF Projects

AI for Time Series and for Science Models

- Advanced Phenology Information System (APIS) (Morisette/AIST-16)
- NASA Evolutionary Programming Analytic Center (NEPAC) (Moisan/AIST-18)
- Mining Chained Modules in Analytics Center Frameworks (Zhang/AIST-18)
- Surrogate modeling for atmospheric chemistry and data assimilation (Henze/AIST-18)
- Predicting What We Breathe: Using Machine Learning to Understand Urban Air Quality (Holm/AIST-18)
- Detection of Artifacts and Transients in Earth Science Observing Systems with Machine Learning and Visualization (Wilson/AIST-21)
- Coupled Statistics-Physics Guided Learning to Harness Heterogeneous Earth Data at Large Scales (Xie/AIST-21)

AI for Pattern and Information Extraction

- Computer-Aided Discovery and Algorithmic Synthesis for Spatio-Temporal Phenomena in InSAR (Pankratius/AIST-16)
- Supporting Shellfish Aquaculture in the Chesapeake Bay using AI for Water Quality (Schollaert-Uz/AIST-18&21)
- Knowledge Transfer for Robust GeoAI Across Space, Sensors and Time via Active Deep Learning (Prasad/AIST-21)
- Detection of Artifacts and Transients in Earth Science Observing Systems with Machine Learning and Visualization (Bock/AIST-21)

AI for Image Processing and for Data Fusion

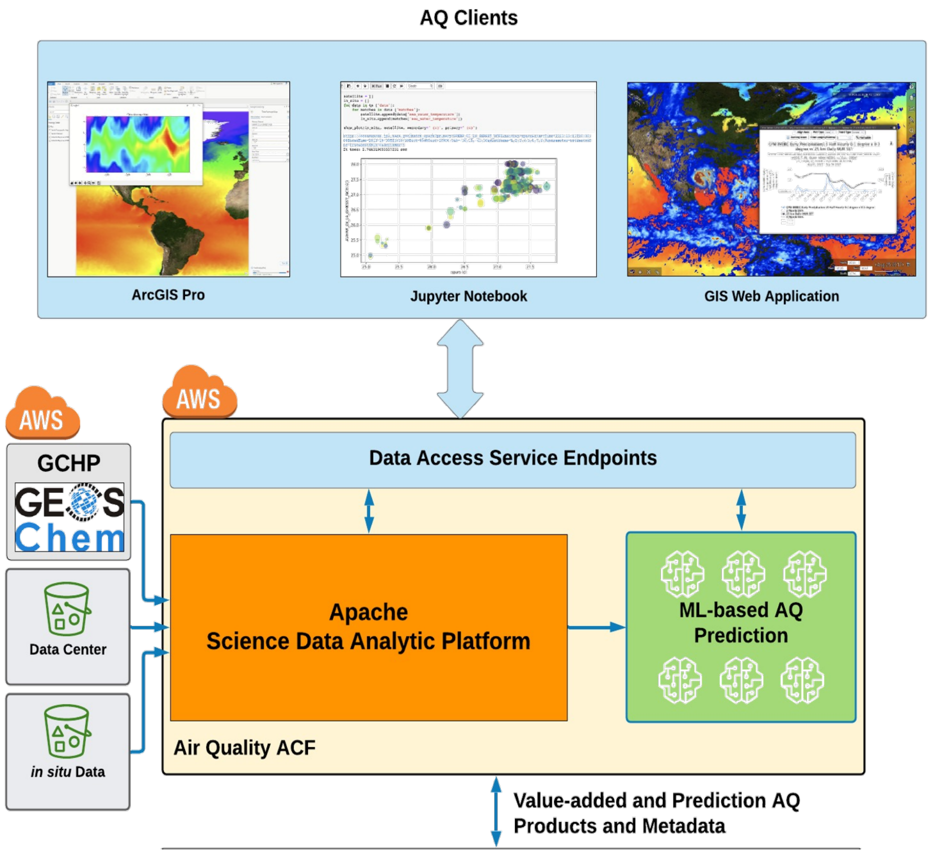
- Software Workflows and Tools for Integrating Remote Sensing and Organismal Occurrence Data Streams to Assess and Monitor Biodiversity Change (Jetz/AIST-16)
- NeMO-Net - The Neural Multi-Modal Observation & Training Network for Global Coral Reef Assessment (Chirayath/AIST-16)
- CAPRi: Cloud-based Analytic framework for Precipitation Research (Beck/AIST-18)

AI for Quantum Computing

- Mining and Analysis of Petabyte-size Time-series on the NASA Earth Exchange (NEX) (Michaelis & Nemani/AIST-16)
- Hybrid Quantum Annealing Approaches for Inferring and Assimilating Satellite Surface Flux Data into Global Land Surface Models (Halem/AIST-16)
- Quantum Computing and Classical Machine Learning for Image-to-Image Translation of Earth Images (Rieffel/QRS-21)
- Stochastic Parameterization of an Atmospheric Model assisted by Quantum Annealing (Guillaume/AIST-21)
- A Path Towards Quantum-Computing-Assisted Earth Science Data Acquisition Tasking and Processing (Grabbe/AIST-21)

QRS-20/Huang (JPL, City of La, WUSTL, U. CO, GMU) – *AQACF: Air Quality Analytic Collaborative Framework*

This project develops an Analytic Collaborative Framework (ACF) for Air Quality that can perform analyses and projections of air quality across multiple synergistic data sets and models. It will demonstrate an analysis application focusing on air pollution in large cities (e.g., Los Angeles).



AQACF platform integrates data, analytics, and models for air quality prediction and analysis.

- Build on the Apache Science Data Analytics Platform (SDAP) architecture to harmonize air quality data sets and analytics.
- Formalize inputs to prediction models, GEOS-Chem, and GCH, as well as models to be developed through this AQACF effort
- Integrate with cloud-provisioned ML services
- Demonstrate AQACF data access and analytic services through integration with Jupyter Notebooks
- Provide a built-in toolbox of algorithms and workflows for on-demand analytics
- Develop interfaces to enable predictive models of air quality to be driven by harmonized data from multiple instruments
- Develop web-based application for data access, interactive visualization and analytics

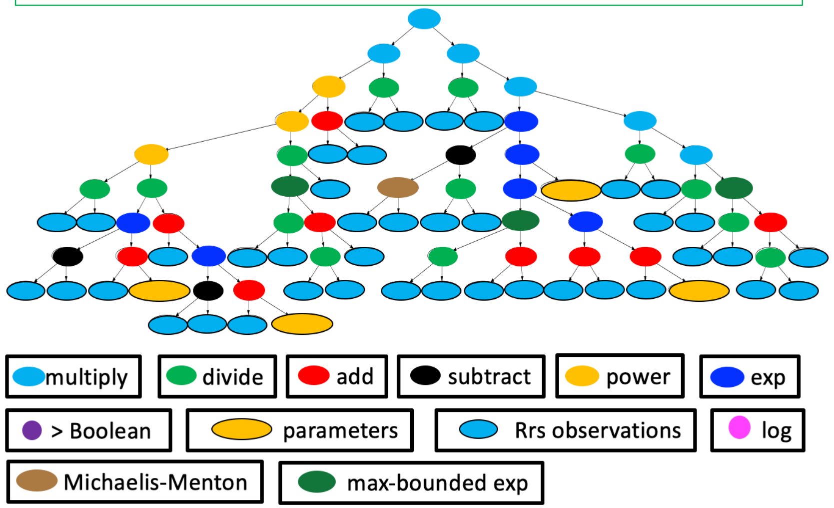
AIST-18/Moisan (GSFC) and Zhang (SMU) –
Using AI for Facilitating the Development of Novel Science Data Processing Workflows and Algorithms

Moisan (GSFC)/"NASA Evolutionary Programming Analytic Center (NEPAC)"

Develops a framework that uses genetic programming to generate new chlorophyll-a (chl-a) algorithms with reduced uncertainties and anneal chl-a estimates across multiple ocean (OC) satellite data sets.

User Requirements/variables (e.g., Sea Surface Temperature, Salinity) & Training Datasets
=> Generate Tree of parameters => Optimal Equation(s)

A GPCODE-generated Chl-a Algorithm (10 levels)

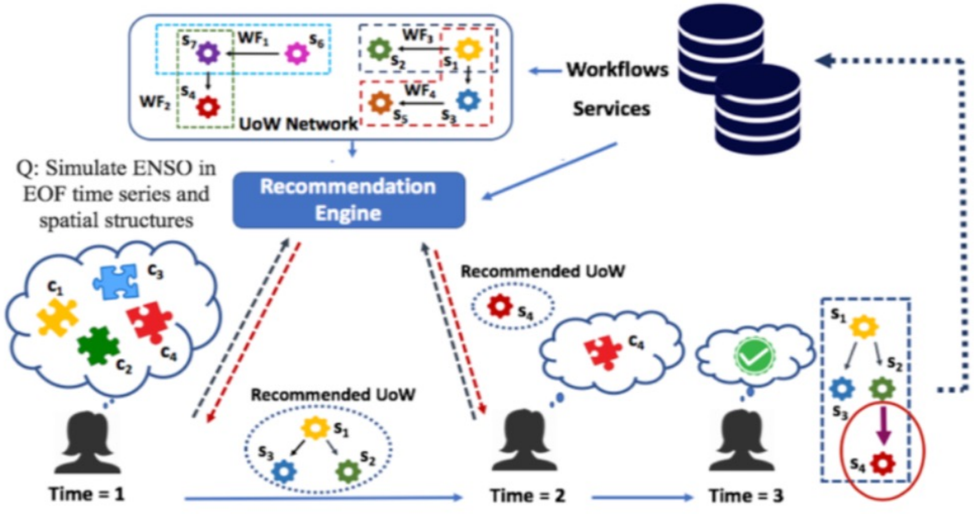


Improved Ocean Color Algorithms and Climate Data Records Using Machine Learning/Genetic Programming

Zhang (CMU)/"Mining Chained Modules in Analytic Center Framework"

Builds a workflow tool as a building block for Analytic Center Frameworks that is capable of recommending chained software modules.

Mine software usage history => construct a knowledge network => explore reusable software module chains => Intelligent service for personalized workflow design recommendations



Workflow Recommendation Framework

Main AIST Technologies Needed for ACF

e.g., ML for Data Accessibility, Analytics, Fusion and Modeling

Develop ACFs for new science domains:
- Manage large data volumes
- Manage wide variety of data types
- Manage frequent data updates (high data velocity)

AI CAPABILITIES:
- Machine Learning
- Deep Learning
- Data Services Discovery
- Uncertainty Quantification Methods

ADVANCED ANALYTICS:
- Data Accessibility
- Data Fusion
- Big Data Analytics
- Data Mining
- On-Demand Product Generation
- Data Operations Workflows
- Data Incorporation of Metadata, Provenance, Semantics, etc.

User
- Project Definition
- Plan for Investigation

Data
- Catalog
- NASA DAAC
- Other US Govt
- Non-US
- Local or non-public

Project Work Environment

Tools
- Discovery & Catalog
- Work Management
- Data Interfaces
- Analytic Tools
- Modeling
- Collaboration
- Visualization
- Sharing/ Publication
- Local/custom

Computational Infrastructure
- Computing
  - Capacity
  - Capability
- Storage
- Communications

Storage
- Data Containers
- Thematic model
- Metadata/Ontology
- Resulting Products
- Published data
- Provenance

IMPROVED MODELING CAPABILITIES:
- Science Data Model Validation
- Software Architecture Frameworks
- Science Code Development & Reuse
- Modeling Systems
- Model Data Inter-Comparisons
- Custom Tools
- Forecasting/Prediction

Computing
- Local systems
- High End Computing
- Cloud Computing Capability
- Quantum Computing
- Neuromorphic Computing

Machine Learning for Earth System Digital Twins (ESDT)

AIST Earth System Digital Twin(s)

AIST defines an Earth System Digital Twin (ESDT) as an interactive and integrated multidomain, multiscale, digital replica of the state and temporal evolution of Earth systems that dynamically integrates:

- Relevant Earth system models and simulations
- Other relevant models (e.g., related to the world's infrastructure)
- Continuous and timely (including near real time and direct readout) observations (e.g., space, air, ground, over/underwater, Internet of Things (IoT), socioeconomic)
- Long-time records
- Analytics and artificial intelligence tools.

Effective ESDTs enable users to run hypothetical scenarios to improve the understanding, prediction of and mitigation/response to Earth system processes, natural phenomena and human activities as well as their many interactions.
An ESDT is a type of integrated information system that, for example, enables continuous assessment of impact from naturally occurring and/or human activities on physical and natural environments.

AIST ESDT strategic goals are to:

- Develop information system frameworks to provide continuous and accurate representations of systems as they change over time;
- Mirror various Earth Science systems and utilize the combination of Data Analytics, Artificial Intelligence, Digital Thread*, and state-of-the-art models to help predict the Earth's response to various phenomena;
- Provide the tools to conduct "what if" investigations that can result in actionable predictions.

* The digital thread designates the communication framework that links all digital twin data flow throughout its lifecycle.

Earth System Digital Twins Components

Digital Replica . . . What now?

An integrated picture of the past and current states of Earth systems.

Forecasting . . . What next?

An integrated picture of how Earth systems will evolve in the future from the current state.

Impact Assessment . . . What if?

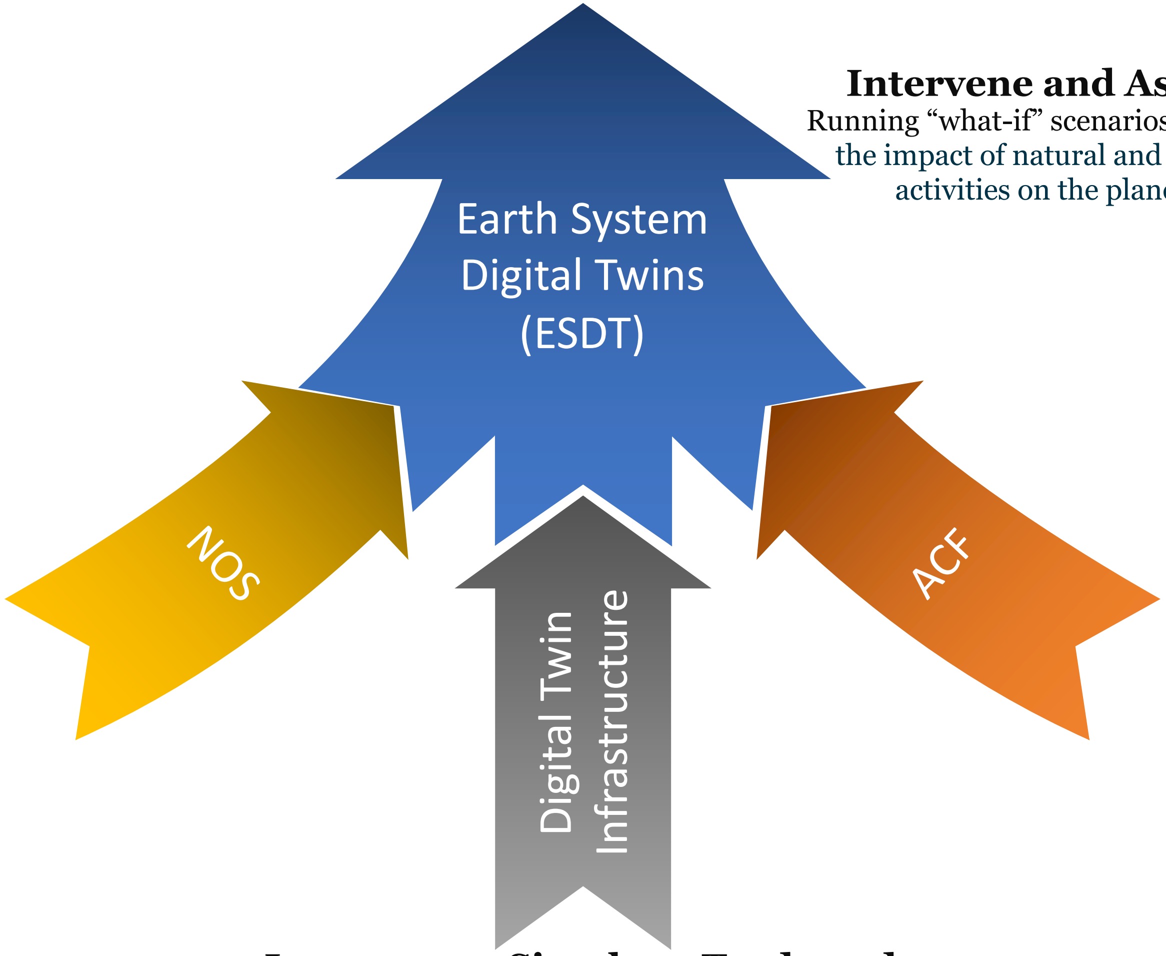
An integrated picture of how Earth systems could evolve under different hypothetical what-if scenarios.

. . .

- **Continuous observations** of interacting Earth systems and human systems
- From many **disparate sources**
- Driving **inter-connected models**
- At many **physical and temporal scales**
- With fast, powerful and integrated **prediction, analysis** and **visualization** capabilities
- Using **Machine Learning, causality and uncertainty quantification**
- Running at **scale** in order to improve our **science** understanding of those systems, their **interactions and** their **applications**

ESDT one of 3 AIST Thrusts

Intervene and Assess
Running "what-if" scenarios to assess the impact of natural and human activities on the planet.



Observe, Target and Coordinate
Edge and on-the-ground intelligent planning, evaluating, coordinating and operating collections of diverse and distributed observing assets

Fuse, Analyze, Share and Collaborate
Simplify access to diverse and large amounts of data, analytics & modeling tools and advanced computational resources for collaborative science

Interrogate, Simulate, Trade and Visualize
Robust tools for interrogating, assessing uncertainties & causality, and for visualization, leveraging diverse data, models and products

NOS = Novel Observing Strategies

ACF = Analytic Collaborative Frameworks

AI in AIST ESDT Projects

AI for Data Fusion, Assimilation and Surrogate Modeling

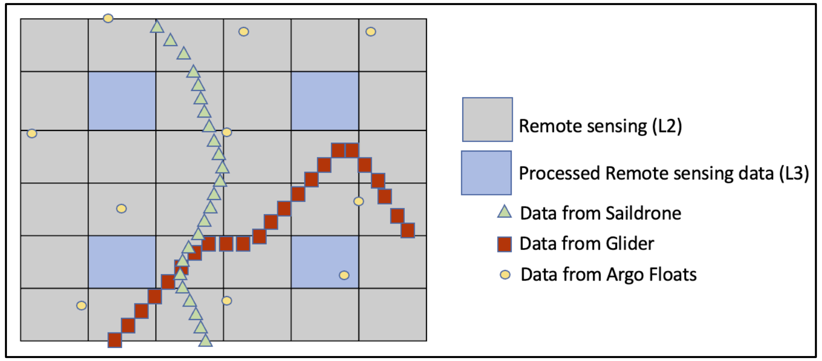
- IDEAS: Integrated Digital Earth Analysis System (Huang/QRS-21)
- A prototype Digital Twin of Air-Sea Interactions (Gray/AIST-21)
- Development of a next-generation ensemble prediction system for atmospheric composition (Keller/AIST-21)
- Terrestrial Environmental Rapid-Replicating Assimilation Hydrometeorology (TERRAHydro) System: A machine-learning coupled water, energy, and vegetation terrestrial Earth System Digital Twin (Pelissier/AIST-21)
- Fire Alarm: Science Data Platform for Wildfire and Air Quality (Huang/AIST-21)

AI for Data Fusion, Assimilation and Surrogate Modeling

- IDEAS: Integrated Digital Earth Analysis System (Huang/QRS-21)
- A prototype Digital Twin of Air-Sea Interactions (Gray/AIST-21)
- Development of a next-generation ensemble prediction system for atmospheric composition (Keller/AIST-21)
- Terrestrial Environmental Rapid-Replicating Assimilation Hydrometeorology (TERRAHydro) System: A machine-learning coupled water, energy, and vegetation terrestrial Earth System Digital Twin (Pelissier/AIST-21)
- Fire Alarm: Science Data Platform for Wildfire and Air Quality (Huang/AIST-21)

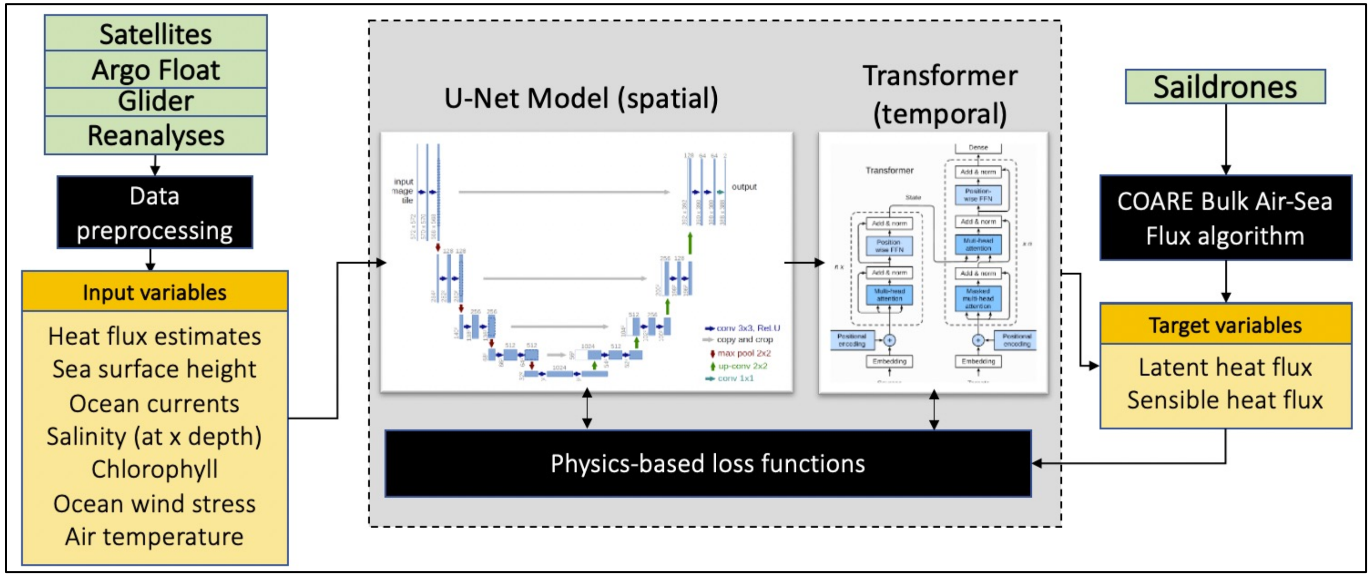
AIST-21/Gray (Univ. of Washington) –
A Prototype Digital Twin of Air-Sea Interactions

This project proposes to develop hybrid physics-informed AI model that ingests several existing flux estimates and observation data products and train against simultaneous ocean-atmosphere data from Saildrones. This will ascertain uncertainty of existing flux measurements and optimize combination of near-real-time existing flux data and observational data. This represents the first step towards a Digital Twin for the Planetary Boundary Layer.



The core technical challenge of this project is the development of the fully differentiable **hybrid model** to capture the boundary layer air-sea interactions. This model will be capable of realizing the non-linear interactions between atmospheric and oceanic surface variables such as sea surface temperature, wind speed and direction, and ocean currents, and the four different components of the heat fluxes. The model development task comprises two data-driven sub-components with physics-informed loss functions and constraints as described below.

- **Data-driven component:** A graph sampling method will be used to query multidimensional input-output data pairs. This data will be modeled using a U-net architecture that is specifically designed to handle spatial data.
- **Physics-informed constraints:** Constraining the model predictions during learning is a critical aspect of the model development to infuse domain knowledge. Constraints will be enforced in two ways:
    1. With ***soft enforceable physics constraints*** in the form of Lagrangian multiplier regularization losses. For example, changes in sea surface fluxes must be balanced by changes in atmospheric sensible heat and energy transport by the circulation.
    2. With ***hard constraints*** using periodic padding on the U-net architecture's direct neighbor layer. The data obtained from the in-situ observation platforms (Saildrones) will be used to validate and calibrate the data-driven model.

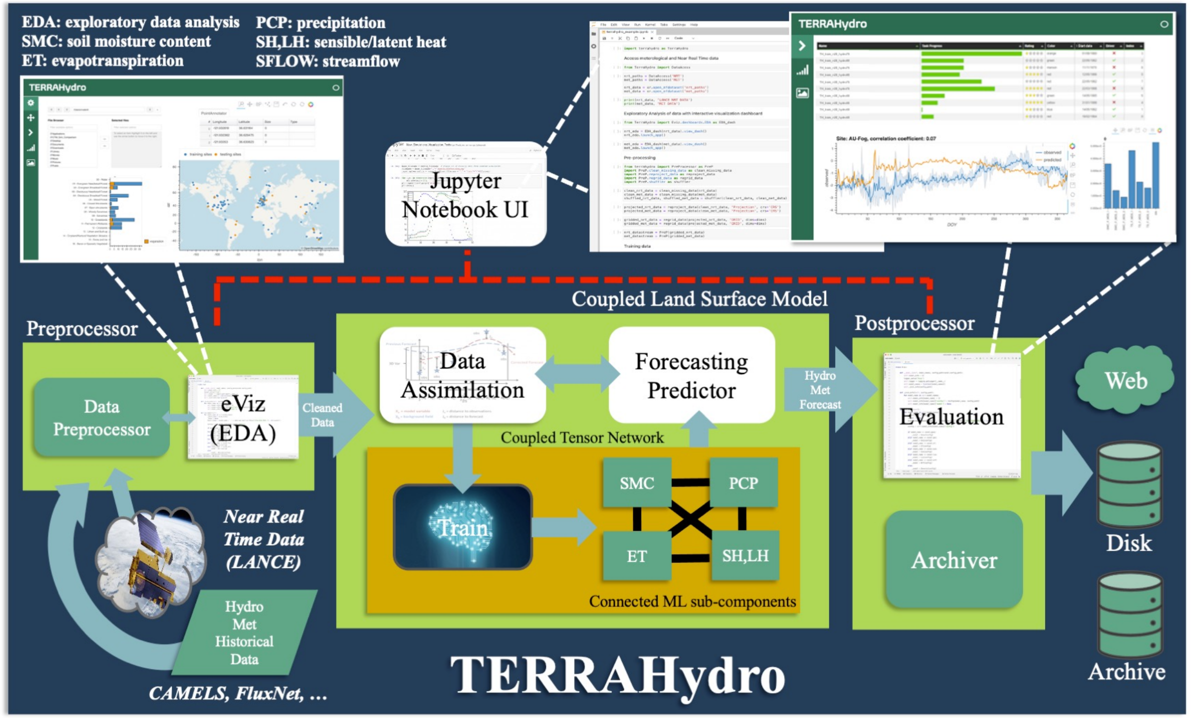


AIST-21/Pelissier (SSAI) – *Terrestrial Environmental Rapid-Replicating Assimilation Hydrometerology (TERRAHydro) System*

This project proposes to develop a Terrestrial Earth System Digital Twin (TESDT) that couples state-of-the-art ML with NASA (and other) EO data. It will combine the best ML hydrology models with capabilities for uncertainty quantification and data assimilation to provide ensemble & probabilistic forecasting, sensitivity analyses, and counterfactual "what if'' experiments.

The modeling framework will include several major components:

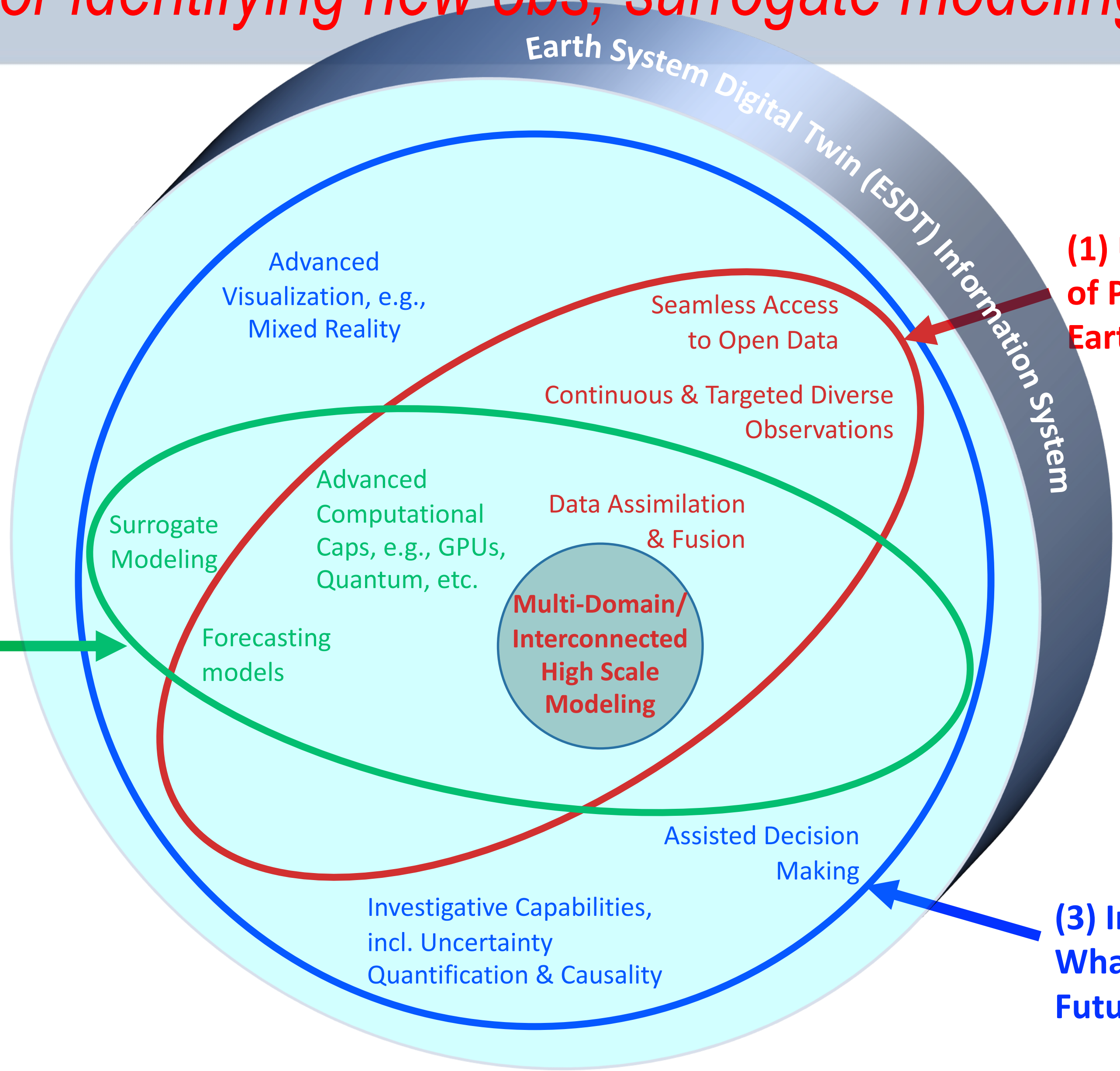
- Data ingestion and pre-processing pipelines that are model agnostic and allow for puzzle-piece integration with ML models.
- Individual ML models of specific land surface variables.
- Mechanisms for coupling individual ML models into a full terrestrial systems model.
- Support for integrating process-based modeling components.
- Support for real-time data assimilation.
- Tools for training, evaluation, scenario analysis, and benchmarking.



Various ML models have been developed and trained for many of the most critical terrestrial states and fluxes, e.g., for Soil Moisture, Streamflow, Surface Heat Fluxes, Net Ecosystem Exchange, and Snowcover. The goal of the project is to support multiple coupling approaches (e.g., direct coupling, shared model structure or PDE-based learning) and to provide the tools and interfaces that allow researchers to develop and test new coupling strategies. This is possible in an effectively general way because of the modular nature of tensor networks.

Main AIST Technologies Needed for ESDT

e.g., AI and ML for identifying new obs, surrogate modeling, fusion, etc.



(1) Up-to-Date *Digital Replica* of Past and Current States of Earth System

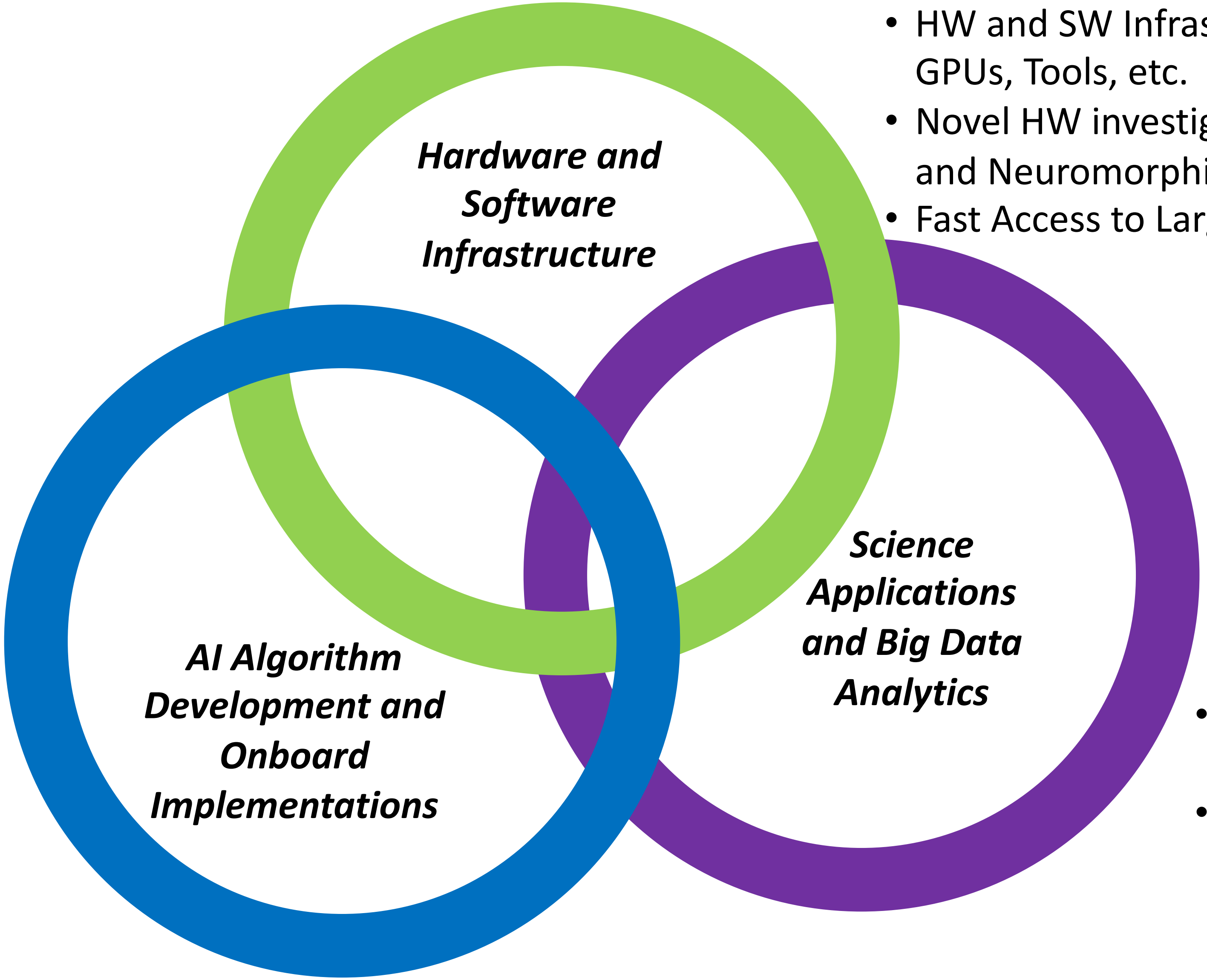
(2) Real-Time (RT) or Near-RT *Forecasting* of Future States of Earth System

(3) Impact Assessment by running What-If Scenarios for Hypothetical Future States of Earth System

What’s Next?

AI for NASA Applications
(Ref. 2018 1st NASA AI Workshop Introduction)

- HW and SW Infrastructure , including GPUs, Tools, etc.
- Novel HW investigation, e.g., Quantum and Neuromorphic Computing
- Fast Access to Large Amounts of Data



- AI Expertise
- Conceptual Software & Algorithm Development
- At the Edge/Onboard Implementations

- Science Applications and Data Analytics
- Algorithm Relevance and Validation

What's Next – AI and ML for AIST
Capability Gaps

- **Capability Gaps**
    - Semantic Representation/Ontologies (Bridge) of Disparate Observation Data and Sensors
    - Data Mining and Discovery
    - Data Analytics, Pattern/Feature Discovery
    - Fusion of Disparate including Unstructured Data Sources
    - Autonomy – Onboard Decision Making and On-the-Ground Semi-Autonomous Model Interactions
    - Causality Assessment between diverse Earth systems and with Human Activity
    - Validation and Uncertainty Quantification of AI Systems
    - Understanding/Evaluation of which AI Systems are Good at
    - Explainability of AI Systems

What’s Next – AI and ML for AIST
Some Promising AI Technologies

- **Some Promising AI and ML Technologies for Earth Science**
  - **Transformer Architectures and Foundation Models**
    - Tracking relationships in parallel in data series
    - Foundation models trained on general data sources in a self-supervised manner, then adapted to solve new problems
  - **Composite AI**
    - Combining multiple existing and new techniques => overcoming limitations of current single technique AI approaches
    - Go beyond data-driven ML and Deep Learning (DL). DL = Reflexive Reasoning vs. Deliberative/Deduction AI (e.g., heuristic search, probabilistic reasoning) vs. Hybrid AI (Reflexive + Deliberative) Combination of different AI or “multidisciplinary analytics”
    - Improve AI’s versatility and efficiency to solve a wider range of problems
    - Challenges: Choice of various techniques + How to combine multiple methods + Increased complexity
  - **Causal AI**
    - Go beyond statistical correlations
    - Discover and leverage cause-and-effect relationships
    - Towards true autonomy
    - Methods: causal rules, causal graphs (Bayesian), ML for causal analysis

Conclusion

In Summary

- **Machine Learning (and more generally AI) can be used in many steps of the Data Lifecycle**
  - For optimizing and coordinating the acquisition of new measurements and the design of future observing systems
  - For the fusion, mining and analysis and ever-increasing amounts of remote sensing data as well as diverse data (including non-traditional unstructured data such as from IoT)
  - For speeding up modeling and prediction capabilities

- **In particular, for Earth System Digital Twins, Machine Learning and other AI techniques are important tools for all 3 components:**
  - Digital Replica: for fusion, data assimilation and acquisition of relevant and timely observations and for model-coupling
  - Forecasting: for surrogate modeling, fast predictions
  - Impact Assessment: for visualization and what-if investigations

- **Challenges:**
  - *Questions [1]: Why Use AI for my application? Can I leverage scientific knowledge? Can I leverage explainable AI? Does my approach generalize? Are my results reproducible? Do I generate new scientific insights?*
  - Validation; Accuracy/Uncertainty Quantification; Explainability
  - Evaluation/Validation/Explainability as they relate to well-understood physical quantities

[1] Ebert-Uphoff, Samarasinghe and Barnes (2019), "Thoughtfully Using Artificial Intelligence in Earth Science": `

Questions?

AIST-21 Word Cloud of Selected Proposals Titles

ESTO

Earth Science Technology Office