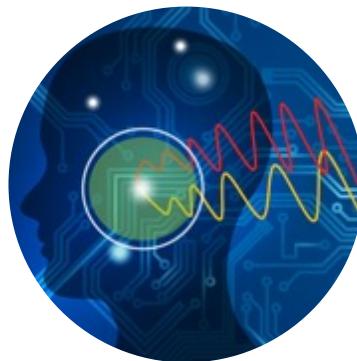




# NASA Earth Science Technology Office (ESTO) Intelligent Systems Technology (IST)



## NASA Earth System Digital Twins (ESDT) For a Sustainable Future

*Jacqueline Le Moigne*

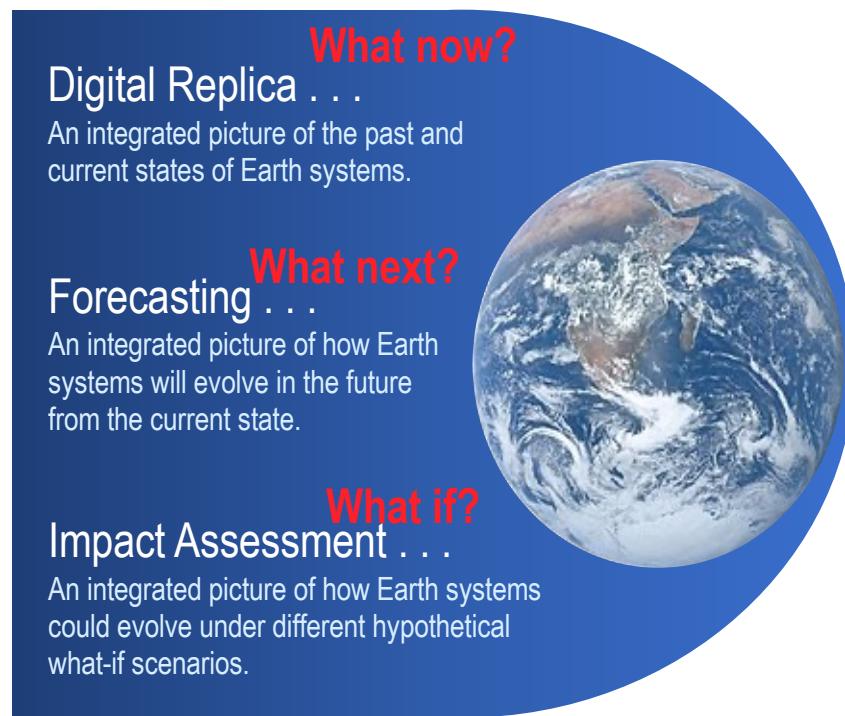
AMS'25, January 2025



# Earth System Digital Twins (ESDTs) Definition



Earth Systems Digital Twins (ESDTs) are information systems for understanding, forecasting, and conjecturing the complex interconnections among Earth systems, including anthropomorphic forcings and impacts to humanity.



An Earth System Digital Twin or ESDT is a dynamic and interactive information system that first provides a digital replica of the past and current states of the Earth or Earth system, as accurately and timely as possible, second allows for computing forecasts of future states under nominal assumptions and based on the current replica, and third offers the capability to investigate many hypothetical scenarios under varying impact assumptions.

## => What Now? What Next? What If?

An ESDT includes:

- Continuous observations of interacting Earth & human systems
- From many disparate sources
- Driving inter-connected models
- At many physical and temporal scales
- With fast, powerful and integrated prediction, analysis & 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



# 2 IST Programs: AIST and AMT



## Intelligent Systems Technologies (IST) Focus Area

### Advanced Information Systems Technology (AIST) Program

Includes:

- Novel Observing Strategies (NOS) development and demonstrations
- Advanced Computational Capabilities, e.g., Quantum and Neuromorphic
- Early-Stage Technology (EST), i.e., promising ground-breaking computer Science & Information Systems technology of interest to Earth Science in 5. to 10 years, e.g., AI for edge- and on-the-ground autonomy

### Advanced Modeling Technology (AMT) Program

Includes:

- Earth System Digital Twins (ESDT) development and prototypes
- Advanced Modeling Capabilities, e.g., Surrogate Models and Emulators, Models Performance Improvement, etc.
- Development, Integration and Validation of State-of-the-Art Mathematical Techniques of interest to Modeling and Digital Twins, e.g., Machine Learning, Uncertainty Quantification and Provenance



# Software and Information Systems Technology for ESDT



**ESDT = Earth System Digital Twins**

## **Interrogate, Simulate, Trade and Visualize**

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

## **Intervene and Assess**

Running “what-if” scenarios to assess the impact of natural and human activities on the planet.

**NOS = Novel Observing Strategies**

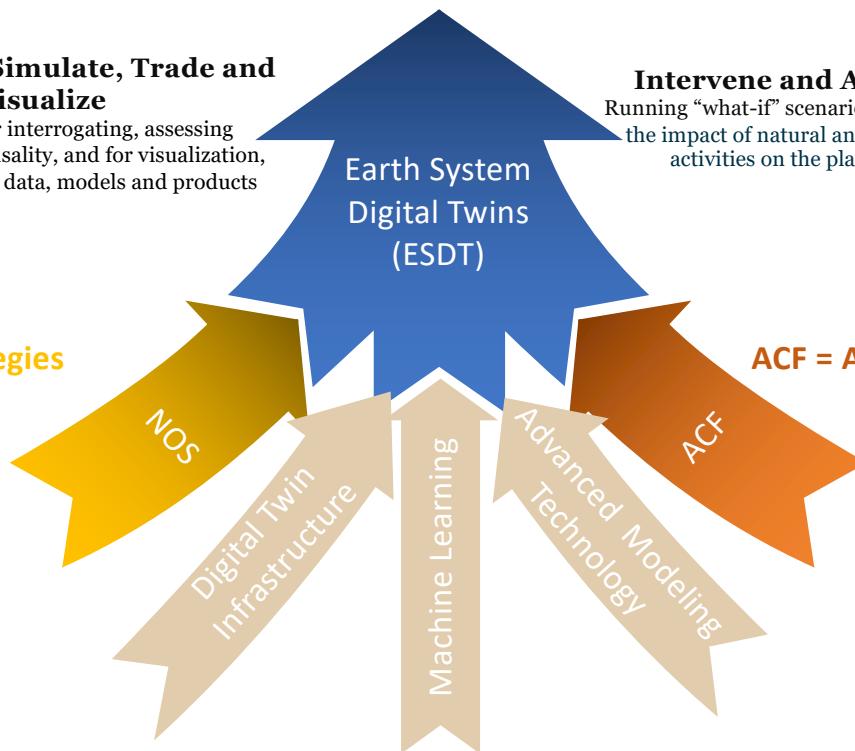
## **Observe, Target and Coordinate**

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

**ACF = Analytic Collaborative Frameworks**

## **Fuse, Analyze, Share and Collaborate**

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





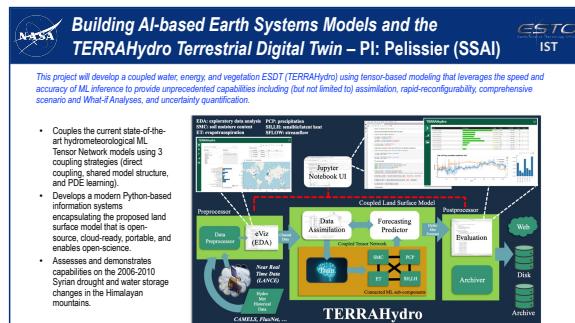
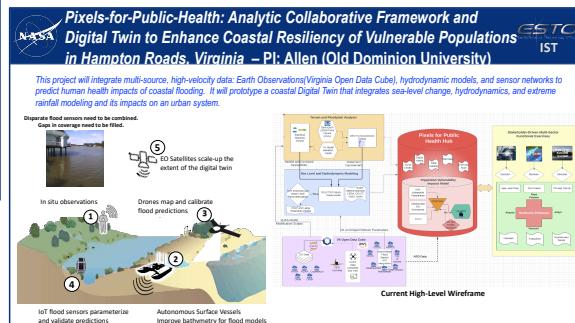
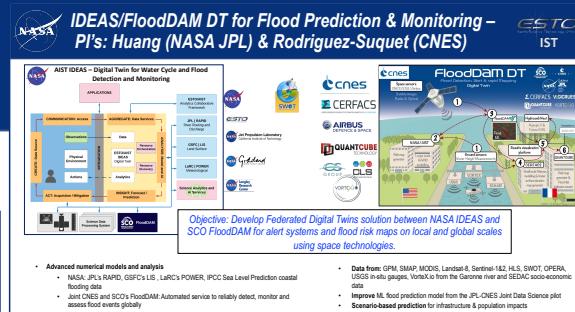
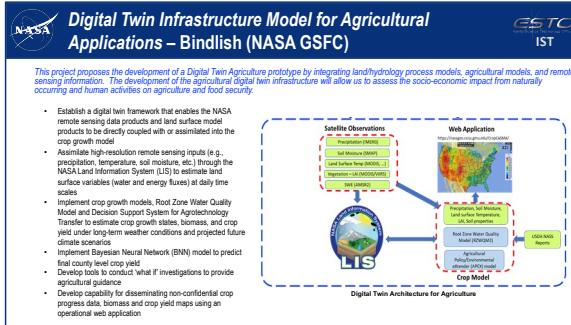
# IST Portfolio (AIST-21 & AIST-23) (64 Active or Newly Selected Projects)



- **15 NOS:**
  - Observing Systems/Missions Design, including OSSEs with spaceborne, airborne and ground-based assets
  - Assets and Constellation Coordination, Planning and Targeting
  - Event and Feature Detection and Tracking
  - Autonomy, Smart Sensors, and Onboard Intelligence
  - Node Security, e.g., Blockchains
  - End-to End NOS Concepts and Demonstrations for:
    - Dynamic Targeting
    - STV concepts
    - Detecting and tracking Deep Convective Systems
    - Hardware-in-the-loop validation of autonomous, context-aware NOS system
- **2 Advanced Computational Capabilities**
  - Physics-aware quantum-classical hybrid NNs to solve complex partial differential equations
  - Onboard quantum sensor denoising
- **9 Early-Stage Technology:**
  - Explainable, physics-constrained, hybrid ML and Foundation Models/Transformer architectures for various applications (soil moisture, fuel moisture content, fires, hydrology, harmful algae blooms, etc.)
  - Novel ML such as Neural Radiance Fields
- **15 ESDT:**
  - Underlying Analytic Capabilities to Build Digital Replicas
  - Novel ESDT Infrastructure Technologies
  - Surrogate Modeling and ML Emulators
  - ESDT prototypes for:
    - Coastal Zones
    - Carbon Corridors & Biodiversity
    - Hydrology
    - Agriculture
    - Wildfires, Water Cycle, Urban Air Quality
- **19 Advanced Analytic and Modeling Capabilities**
  - Advanced Analytics (including Data Accessibility, Data Fusion, Data Mining, etc.)
  - AI Capabilities (Machine Learning, Deep Learning, Foundation Models, Uncertainty Quantification, etc.)
  - Computational Environments
  - Advanced Modeling Capabilities
    - Model improvement and coupling
    - Physics-informed ML for ESDT
    - Climate tipping-point simulation
- **4 State-of-the-art Mathematical Techniques**
  - Advanced AI Capabilities
  - Computational Environments
  - Uncertainty Quantification



# ESDT Prototypes



## Prototypes to:

- Identify & Develop ESDT “Building Blocks”
- Address Overarching Questions

- How will various data, models, ESDTs be accessed/interoperate/be federated?
- Which basic interfaces, standards, and protocols will be required? At syntactic, semantic, legal and organizational levels?
- How to facilitate quick, automated and seamless access to data and data products?
- Which computational resources will be required? Cloud, GPU's, Quantum, Neuromorphic, edge computing, etc.?
- How will continuous data be integrated? How often should digital replica be refreshed?
- Which user interfaces will be needed? UI/UX, impact assessment, decision support, visualization?
- How will ESDTs be validated (e.g., using historical data, etc.)? How will ESDTs' performance be measured? To which level of fidelity do they represent the current, future and hypothetical states of Earth systems? How to quantify uncertainty?
- Which sustainable digital twin governance model should be adopted to address software configuration changes, security and full life cycle management?



# Coastal Zone Digital Twin (CZDT)

PI: Walter/NASA – Collaboration with CNES and NOAA



An Earth System Digital Twin of local and regional coastal zones considering both natural and human systems to understand changes in coastal flooding severity, land and marine morphology, nutrients and water quality, ecological makeup, sea level, and short and long-term risks to climate change adaptation, sustainable development, disaster management, tourism and recreation, quality of life, ecosystem management, and coastal infrastructure management.

What is the effect/impact of changing climate on coastal environment under various sea level and storminess scenarios?

What would be the economic health changes if flood risks were lowered?  
Increased?

What are the shifts in phytoplankton types under different natural/human forcings with improved HAB forecasting?



What would be the changes in ecological makeup if cities reacted to increased flood risk?

What would be the flood risk changes if global temperature goals were met? Not met ?

What would be the water quality changes under different water management structures/policies?

What are the impacts of management on blue carbon ecosystems to support climate mitigation and adaptation and improve resiliency to climate impacts?

What would be the economic outlook be if biodiversity changed as a result of city or industry change?

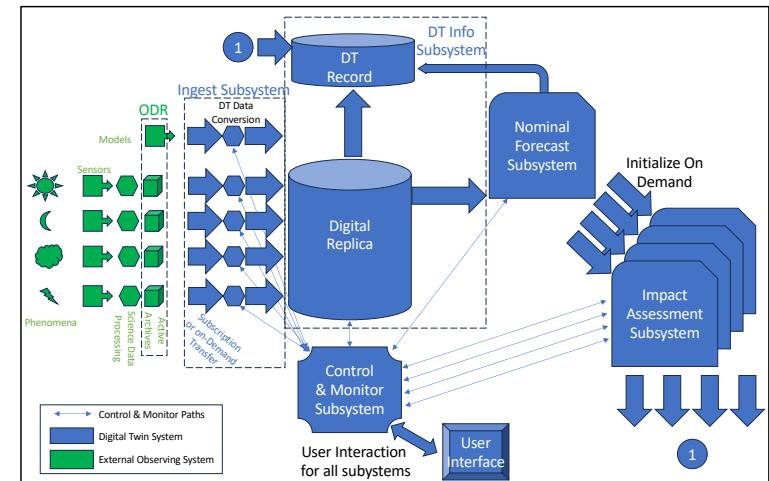
How can we support cities to mitigation if flood risk increased?



# ESDT Architecture Framework Considerations



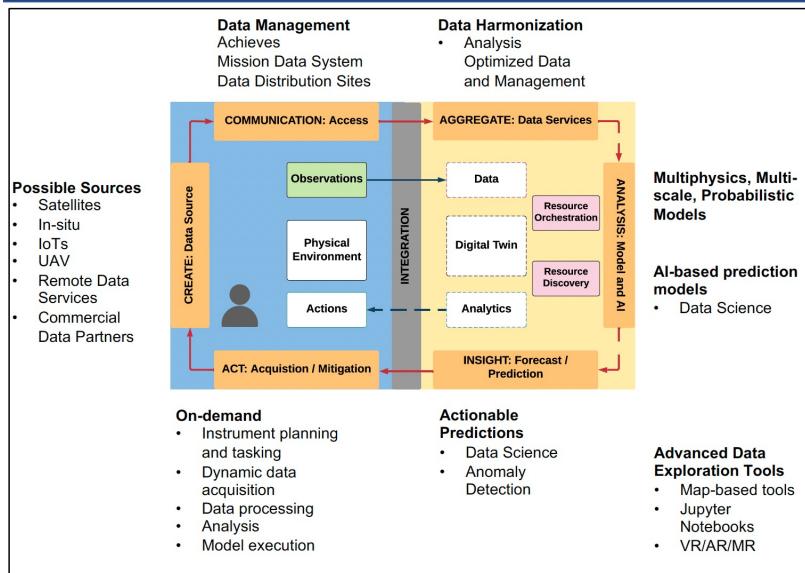
- Consider architecture principles
  - Modularity
  - Process automation and error checking
  - Comply with Open-Source Science principles from SPD-41A
  - Permit evolution of concepts and uses and reasonable addition of new components
  - Provide the Glue to stitch together all ESD capabilities
  - Open-standard interfaces enabling opportunities for broader use
  - Interfaces for federation with other ESDTs
  - User interfaces for a range of skill levels and interests (i.e., "from farmer to scientist")



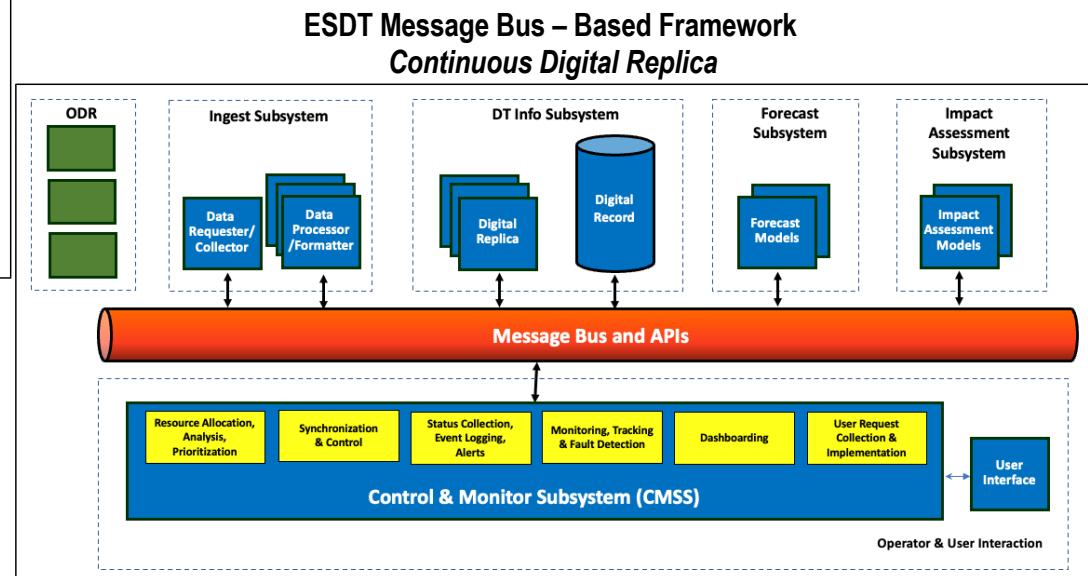
- Enable component technology developers to consider their place in the overall architecture
  - Re-use beyond a single architecture
  - Identify technology gaps and what is required to fill them



# Prototypes to Explore ESDT Architectures



**Integrated Digital Earth Analysis System (IDEAS) Framework**  
**On-demand Data Access**

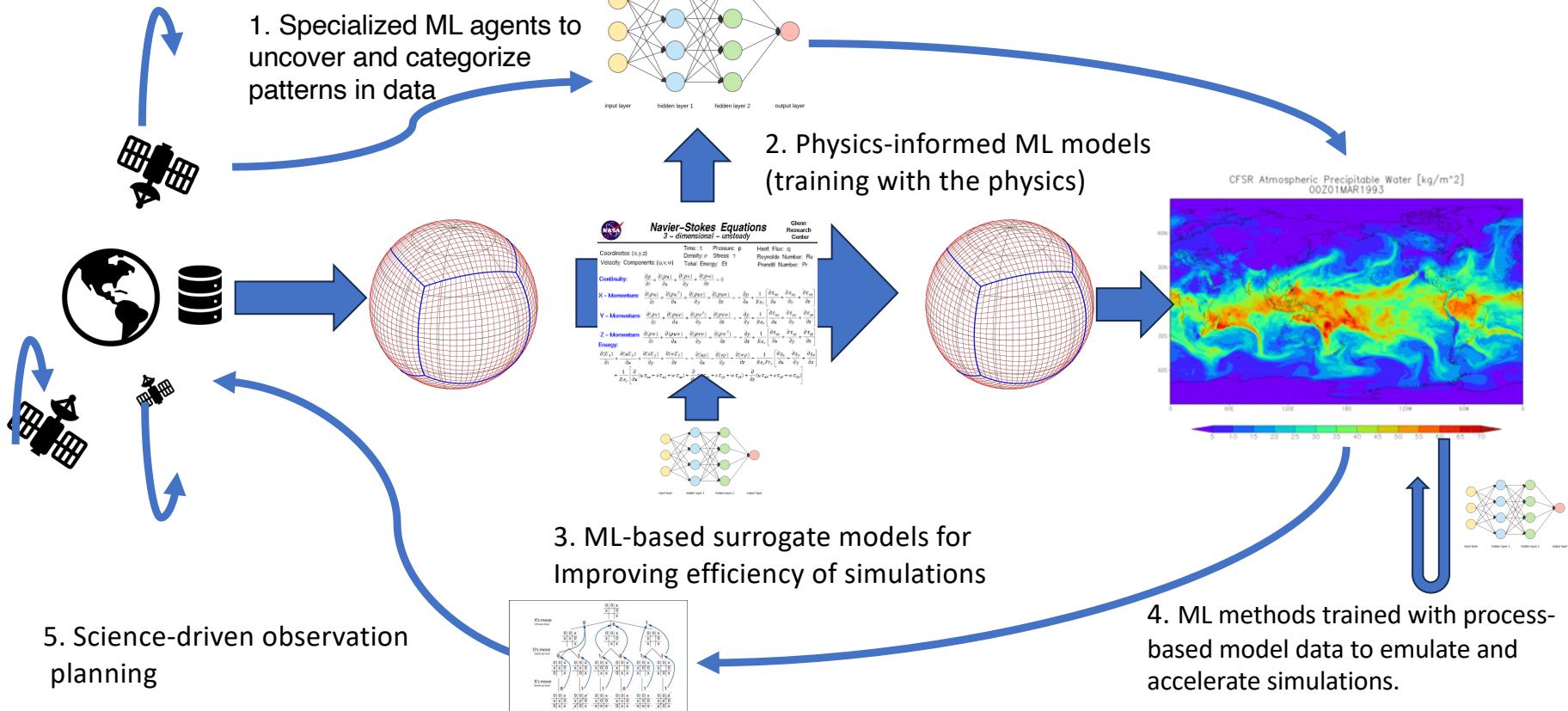




# Prototypes to Explore the Use of AI for ESDT



## 6. On-board autonomy





# Prototypes to Explore ESDT Computational Challenges



Function	SubSystem	A100 GPU-node-hours - annual	CPU-node hours-annual
Pre-train Forecast Models-Hydrology	FMSS	25 runs* 64 Nodes x48h from 800 TB data	
Pre-training updates (Hydrology)	FMSS	1 runs/mo* 64 Nodes x48h from 800 TB data	
Periodic forecast update	NFSS	2 Nodes*5min *6 runs/day*365d	
Monitor	CMSS		24h*7d*52w*3nodes
Ingest and feed DR	Ingest		24h*7d*52w*4 nodes*1time/hour
DR file service	Info SS		24h*7d*52w*2 nodes
Impact Assessment (Planner)	IASS	1 Node*2 min * 12/h * 10h (working cycle) * 5d/wk*52 wk/year	1 Node*30 min * 12 requests/h * 10h (working cycle) * 5d/wk*52 wk/year. (peak)
Visualization Support	UISS	1 Node * 15min* 6 requests/h *12h/d* 5d/w*52w/y	1 node *45min/req 6 requests/h *12h/d *5d/w*52w/y
DT Record file service	Info SS		24h*7d*52w*2 nodes
Initial Data conditioning	Ingest		16 nodes*16h/d*10d
Re-delivery Data conditioning	Ingest		4nodes*24h/d*365d/y
SysAdmin	CMSS		1 nodes*4h/d*365d/y
Archive	CMSS		4 nodes*24h/d*365d/y

**CZDT Example for Chesapeake Bay: 1st YEAR TOTAL NODE HOURS: 81,857 GPU or 270,180 CPU or about 10 GPU nodes for a year**

*Note: Duty cycle not regular; some evolutions require a large number of GPU nodes for a few days, but only happen a few times a year. Others require a single node for a few seconds but occur frequently and on demand in a more interactive way. Load distribution among GPUs not always even, especially for training.*

Function	Assumptions
Pre-training (Initial)	24 runs for experiments + 1 final run w/ 800TB data (INTERACTIVE) AI models are constructed using a lot of experiments to try the model settings to get the best results.
Pre-training Updates	Repeat 1 pre-training run and no experiments once per month (INTERACTIVE)
Periodic forecast update	Inference run on model updated with prompts every 4 hours (BATCH)
Monitor	Monitor System Status, run message bus, direct traffic for an IASS request and monitor for faults, failed deliveries, etc. (BATCH)
Archive	Moving data from DR to DTRecord on schedule (BATCH)
Ingest and feed DR	Ongoing acquisition and conditioning observational data and putting into the Digital Replica (BATCH)
DT Record file service	continuous availability
DR file service	continuous availability for ongoing archiving and requests for historical data
Initial Data conditioning	Workflow to produce initial load of data for constructing AI model (assume 800 TB) (BATCH)
Re-delivery data conditioning	Out of cycle fixes to missed deliveries (BATCH)
Impact Assessment (Planner)	Peak is 12 requests/hour. Day work cycle is 12 hours for peak. (INTERACTIVE)
Visualization Support	6 requests/hour peak at 12 hour days (INTERACTIVE)
SysAdmin	24x7 availability to inspect system (INTERACTIVE)



# NASA Earth Science to Action (ES2A) Strategy



NASA's Earth Science to Action strategy aligns our assets to provide actionable information for a wide range of actors and decision-makers, and to do so for a variety of impactful areas identified for their strategic importance to national and international priorities.

<https://science.nasa.gov/earth-science/earth-science-to-action/>

## Objectives:

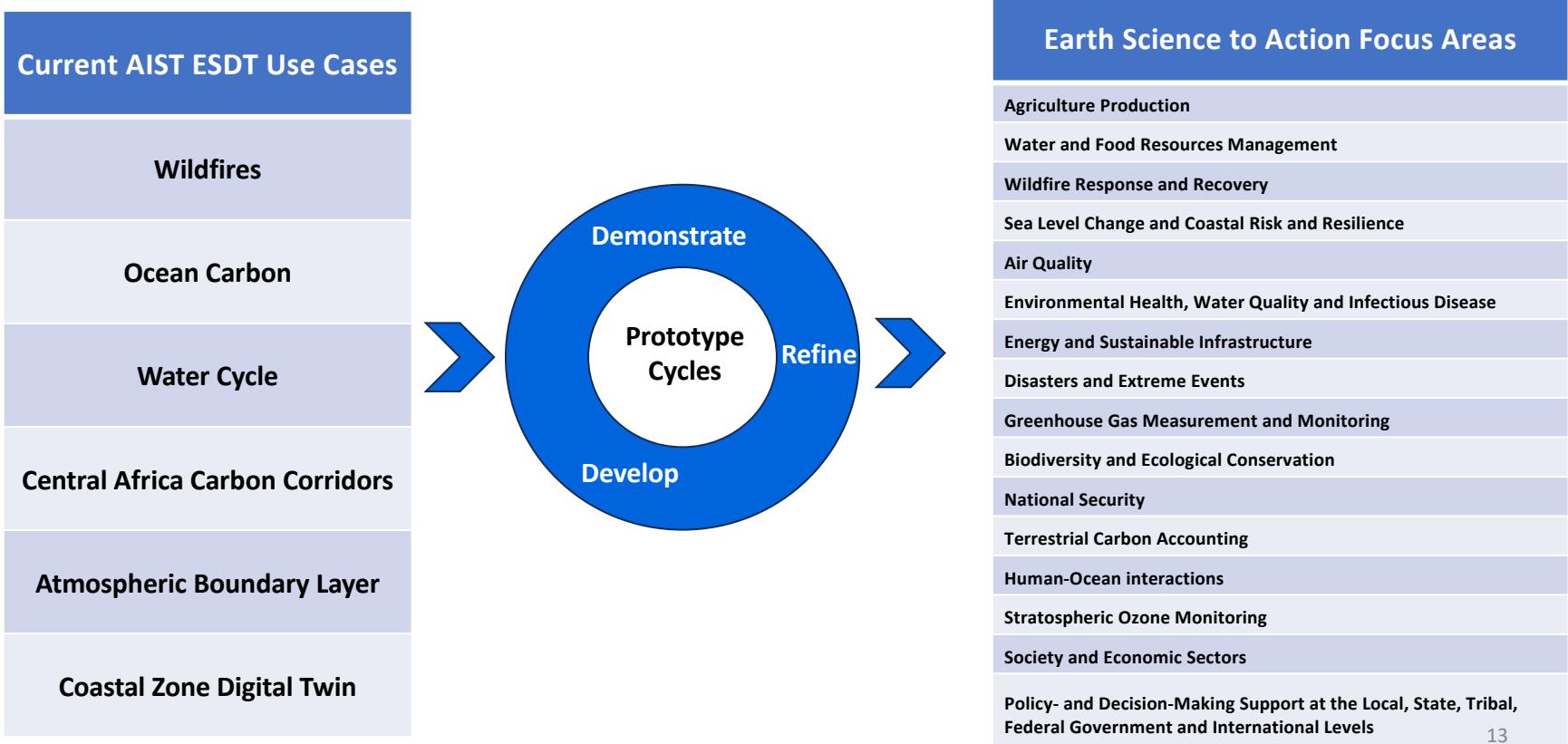
**Objective 1:** Holistically Observe, Monitor, and Understand the Earth System

**Objective 2:** Deliver Trusted Information to Drive Earth Resilience Activities



# ESDT Benefits to NASA Earth Science

## Mapping ESDT Use Cases to ES2A Focus Areas





# AIST-23 ESDT Awards



## • ESDT Infrastructure Capabilities

PI's Name	Organization	Title	Synopsis
Randall Martin	Washington University	Connecting a Broad Community to Earth System Digital Twin Technologies at the Interface of Atmospheric Composition with the Earth System	This work proposes to enhance the high-performance cloud version of GEOS-Chem, an atmospheric chemistry and transport model, GCHP, to several other models - namely, the Modeling, Analysis and Prediction Layer (MAPL), the integrated Methane Inversion model (IMI) and NU-WRF for regional two-way coupling with clouds, precipitation, and land processes. This responds to the ESDT need of "inter-connected modeling" described in the solicitation and provides a new capability for coupled modeling of atmospheric chemistry, transport, and methane that runs in a scalable high-performance cloud environment.
Matthias Katzfuss	University Of Wisconsin, Madison	A software tool for probabilistic calibration and data assimilation for geophysical models	This project will develop a new software toolbox for digital capabilities such as probabilistic model-data fusion, and merging noisy and heterogeneous data with complex, expensive, and imperfect models. It will take into account uncertainty and non-Gaussian properties in the forecast distribution and will provide tools to assimilate real-time observations and contribute to the integration of uncertainty quantification in important components of ESDTs. In their AIST-21 project, the team successfully developed a state-of-the-art probabilistic emulator based on Bayesian transport maps that will be used for offline and online model-data fusion. The toolbox will be demonstrated in a modeling and assimilation system for monitoring river discharge using data from SWOT and NISAR.



# AIST-23 ESDT Awards



## • ML-Based Modeling Capabilities for ESDT

PI's Name	Organization	Title	Synopsis
Romit Maulik	Pennsylvania State University	Machine-learning to improve cycling and forecasts with GEOS and expedite the evaluation of assimilating observations from new instruments	The project addresses the need for developing Machine Learning (ML) surrogate models that will be used to enhance the GEOS Atmospheric Data Assimilation System (ADAS). It will facilitate the integration of new instruments and will improve forecasts of sea-surface temperature (SST) and sea ice by using AI predictions in ocean boundary conditions. The project will use the Stormer Vision Transformer-based (ViT) neural network model that will be initially trained using ERA5 data and refined using MERRA-2 data. Uncertainty Quantification will be computed, and the proposed work will create datasets for training and validation. It will provide a prototype for agile science analysis using advanced ML tools.
Jennifer Sleeman	Johns Hopkins University	AI Climate Tipping-Point Simulator (ACTS)	The project addresses "multi-discipline analytic concepts, e.g., enabling analysis of cascading impacts.". It will develop an AI-based climate tipping point system for climate tipping point and cascade discovery; the outcome of the proposed work will be an end-to-end pipeline for investigating the key inputs for tipping points in climate phenomena, using Deep Learning and Foundation Models configured to investigate a wide range of potential use cases. The system will be able to be interrogated with natural language and will provide causal assessments.
Tianle Yuan	University of Maryland Baltimore County	Building Earth System Digital Twins with Deep Generative Models for Improved Simulation of Clouds and Associated Feedbacks and Impacts	The objectives of this proposal are to utilize Deep Generative AI to model instantaneous cloud fields and long-term cloud feedbacks, with potential use in future ESDTs since it could generate synthetic cloud data. The technique is using latent diffusion trained on Global Climate Models, observations and reanalysis products. It will investigate equilibrium climate sensitivity and cloud-climate feedbacks and to which degree AI can improve representations of clouds in reanalysis data. It is focused on determining how much warming there will be given a doubling in CO <sub>2</sub> (2, 3 or 4 degrees?), a very important question that has remained uncertain for 40 years: clouds are the largest source of uncertainty in model predictions, and the difference between 2 degrees of warming vs 4 degrees is extreme.



# AIST-23 ESDT Awards



## • ESDT Prototype Components & Early Prototypes

PI's Name	Organization	Title	Synopsis
Yiqun Xie	University of Maryland	A Digital Twin Integrating Physics and AI for Understanding Carbon and Biodiversity Corridors in Central Africa	This work is based on previous AIST-21 AI work and addresses several aspects in the development of an Earth System Digital Twin for one of the ESDT use cases, "Central Africa Biodiversity and Carbon Corridors." It will develop a 4D approach combining 2D space, 3D height and time observations to enable a high-fidelity reconstruction of the digital replica of forest structure and aboveground carbon stocks. It will also use high-accuracy AI-accelerated emulators for forecasting and for exploring future scenarios including alternate climate assumptions from CMIP6.
Sujay Kumar	NASA Goddard Space Flight Center	Mapping anthropogenic water cycle impacts in a future climate: A global digital twin for scenario-driven exploration	The proposed work will build deep-learning-based models (e.g., based on axial transformers to capture the relation between meteorological inputs and hydrologic state with relatively high accuracy) that will be used to develop a water cycle digital twin on a global scale for hydrological projections, with emphasis on the representation of anthropogenic influences in future climate. It will leverage land reanalysis, remote sensing, and climate projection datasets, using the NASA Land Information System (LIS). The proposed work will deliver the capability to look at various scenarios of terrestrial water cycle availability including explicit scenarios of human water use, and therefore will enable planning, adaptation, and mitigation strategies in the face of climate change impacts.
Nima Pahlevan	SSAI, Inc.	A forecasting scheme for accelerated harmful algal bloom monitoring	This project will complement the Coastal Zone Digital Twin (CZ-DT) ESDT development. It will develop and implement the FASTHAB framework, a machine learning-based system designed for forecasting harmful algal blooms (HABs) and water quality (WQ) conditions in estuarine environments. The project focuses on the Chesapeake Bay as a testbed ecosystem and aims to integrate NASA's historical ocean color (OC) products, coupled hydrodynamic-biogeochemical model outputs, and physical forcing data into a Convolutional Long Short-Term Memory (ConvLSTM) model. It will use ERA5 and SeaDAS data and will predict important WQ variables such as chlorophyll-a (Chla), Total Suspended Solids (TSS), Secchi disk depth (Zsd), temperature (T), and salinity (S) up to 30 days in advance. The project will leverage the CREST framework developed by AIST-21/Pelissier.



# AIST-23 ESDT Awards



## • ESDT Prototype Components & Early Prototypes

PI's Name	Organization	Title	Synopsis
Mohammad Pourhomayoun	California State University, Los Angeles	An Earth System Digital Twin for Wildfire: Predicting Wildfire Progression and Behavior, and Its Downstream Impacts on Air Quality	This project will leverage 2 previous AIST projects, the Integrated Digital Exploratory Analysis System (IDEAS) as well as the AIST-18 project that developed several ML-based Air Quality (AQ) prediction models, to build an ESDT for Wildfires and their impact on AQ. This work will extend the previous ML work by adding Visual Transformers and Self-Attention Memory (SAM) mechanisms and will add the capability to run scenario-based assessment and Augmented Reality (AR)-based visualization. Various satellite data will be utilized, including MODIS, VIIRS, and records from TROPOMI, CALIPSO, OMI, OMPS, AIRS, Landsat, as well as recent/forthcoming data such as TEMPO and MAIA. The plan is to potentially transition this work to the "Climate Emergency Mobilization Office" as an operational system.
Fritz Policelli	NASA Goddard Space Flight Center	A Terrestrial Water Budget Digital Twin	This project will develop an ESDT for terrestrial water budget analysis following the AIST ESDT Architecture Framework. It will utilize observational data from SMAP, GRACE and SWOT and CMIP6 climate model projections for precipitation and evapotranspiration with machine learning for projections of mass change and runoff. It will facilitate water budget analyses, broaden this research community, and extend its relevance to water basin managers and the public. A software system engineering approach will be adopted to develop the overall system and at the same time coordinate all their developments, architecture-wise with the AIST community and functionality-wise with the user community.



# AIST-23 ESDT Awards



## • ESDT End-to-End Prototypes

PI's Name	Organization	Title	Synopsis
Seungwon Lee	Jet Propulsion Laboratory	Earth System Digital Twin for Central Africa Carbon and Biodiversity Corridors	<p>The project will develop an end-to-end digital twin of the “Central Africa Carbon and Biodiversity Corridors” (CACBC), with the goal to support the Central African communities for informed decision-making towards conserving carbon, biodiversity and sustainable development goals, in harmony with the need for agriculture and infrastructure expansion. The work brings together decades of research in terrestrial ecology and animal movement and at the same time provides the technological innovations to advance the state of understanding through utilization and customization of the AIST ESDT framework. The proposed approach includes efficient workflows, running models, regular data integration, mapping and visualization using available software such as Azure, Jabbix, Reactjs, and Common Mapping Client.</p>
Craig Pelissier	SSAI, Inc.	Pilot Deployment of TERRAHydro: a framework, demonstration, and vision for Earth System Digital Twins	<p>This proposed work will continue the development of a Hydrometeorological Land Surface Digital Twin (TerraHydro) based on the development of various surrogate models (deep neural networks). It is based on a more general framework, the Coupled Reusable Earth System Tensor (CREST), using graph-based models that was developed during the AIST-21 project. CREST can be used for coupling ML-based models, for training ML models using process-based models, and then for running and operating those ML models. During AIST-23, CREST development will be completed and will be used for developing and demonstrating TerraHydro. This will provide a valuable resource for the future development of ESDTs.</p>
Thomas Huang	Jet Propulsion Laboratory	Rapid Streams ESDT: Federated Earth System Digital Twins for Global Flood Prediction	<p>This project is based on the previously developed Integrated Digital Earth Analysis System (IDEAS), an open-source framework, developed jointly between JPL, NASA Goddard, NASA Langley, in collaboration with the CNES-led Space for Climate Observatory (SCO)'s FloodDAM-DT effort for flood analysis and detection. This multi-center and cross-agency effort demonstrated a prototype solution of federated digital twins that combine observations and advanced numerical models for two small river basins: a portion of the Mississippi River and the Garonne River. This AIST-23 project will expand the previous ESDT solution for Global Flood Prediction: enabling models to immediately utilize the latest data from mission Science Data System (SDS) and/or the NASA DAACs without having to perform massive download, leveraging SWOT data and GSFC's LIS simulations that contain spatiotemporal seamless global river discharge data, leveraging probabilistic flood analysis and overlaying CoreLogic Inc's property and environmental risk data, and including augmented/virtual reality (AR/VR) system for immersive visualization and machine learning-based analysis of floods.</p>

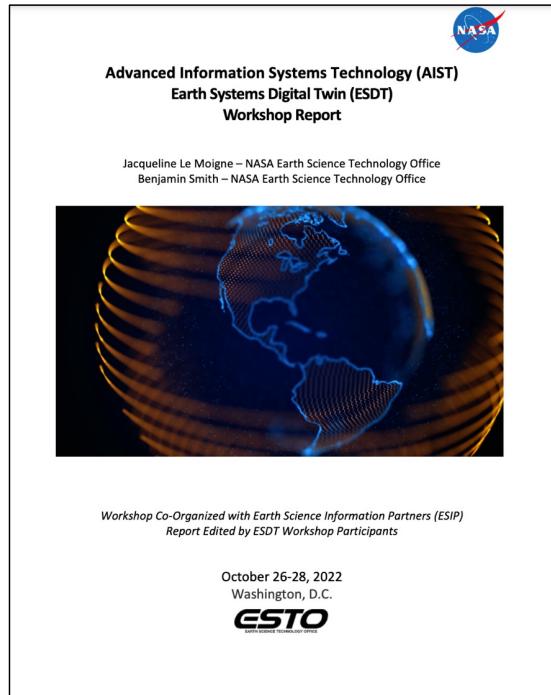


# For more information ...



**ESDT Workshop Report available  
on AIST Website:**

[https://esto.nasa.gov/files/ESDT\\_Workshop\\_Report.pdf](https://esto.nasa.gov/files/ESDT_Workshop_Report.pdf)



Advanced Information Systems Technology (AIST)  
Earth Systems Digital Twin (ESDT)  
Workshop Report

Jacqueline Le Moigne – NASA Earth Science Technology Office  
Benjamin Smith – NASA Earth Science Technology Office

Workshop Co-Organized with Earth Science Information Partners (ESIP)  
Report Edited by ESDT Workshop Participants

October 26-28, 2022  
Washington, D.C.  
**ESTO**  
Earth Science Technology Office

## ESDT Use Case

- Wildfires
- Ocean Carbon
- Water Cycle
- Central Africa Carbon Corridors
- Atmospheric Boundary Layer
- Coastal Zone Digital Twin

## Standards for Interoperable Digital Twins Workshop

September 18, 2023

- Presentations:  
<https://esto.nasa.gov/files/AIST/ESDT%20Standards%202023.pdf>
- Video:  
<https://www.youtube.com/watch?v=qdpLOUi-iqc>

## Project Selections for AIST-23:

<https://esto.nasa.gov/project-selections-for-aist23/>

**Document available  
on AIST Website:**

[https://esto.nasa.gov/files/AIST/ESDT\\_ArchitectureFramework.pdf](https://esto.nasa.gov/files/AIST/ESDT_ArchitectureFramework.pdf)



Advanced Information  
Systems Technology (AIST)

Earth System Digital Twin (ESDT)  
Architecture Framework

Jacqueline Le Moigne, Michael M. Little,  
Robert A. Morris, Nikunj C. Oza,  
K. Jon Ranson, Haris Riris,  
Laura J. Rogers, Benjamin D. Smith

October 2023

**ESTO**  
Earth Science Technology Office

