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

ESTO
Earth Science Technology Office
IST

Technology for Earth System Digital Twins

Jacqueline Le Moigne, Ben Smith, Michael Little,
Robert Morris, Laura Rogers, Jon Ranson and Nikunj Oza

October 2024

Earth System Digital Twins (ESDTs) Definition

ESTO
Earth Science Technology Office
IST

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.

What now?

Digital Replica . . .

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

What next?

Forecasting . . .

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

What if?

Impact Assessment . . .

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

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

What is different about Digital Twins?

1. **Continuous integration** of timely data (real- or near-real-time for some applications, “timely for others)

2. **Interactivity** with users => “playing with the models and the data” for policy/decision making and conjecturing/planning

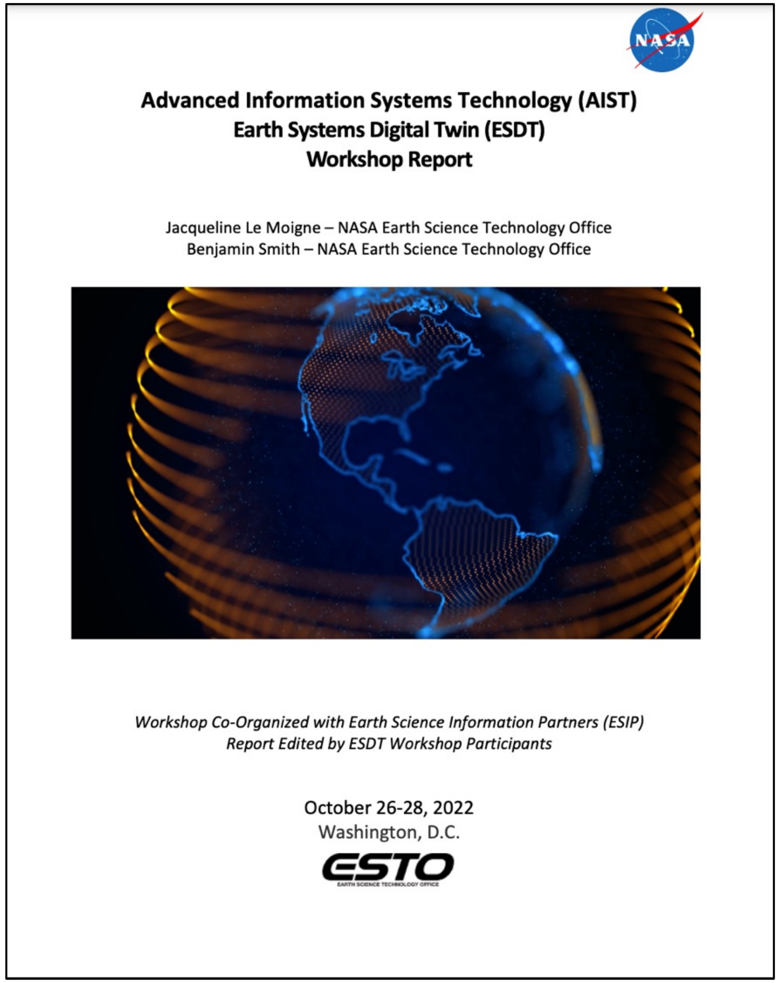
3. Integration of anthropomorphic forcing and **impact models**

4. Heavy use of **Machine Learning**
   - Data Fusion and Data Assimilation
   - Super-Resolution/Downscaling
   - Speeding up models => higher spatial and temporal resolution possible
   - Causal Reasoning

ESDT Background

IST

ESDT Workshop Report available
on AIST Website:
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

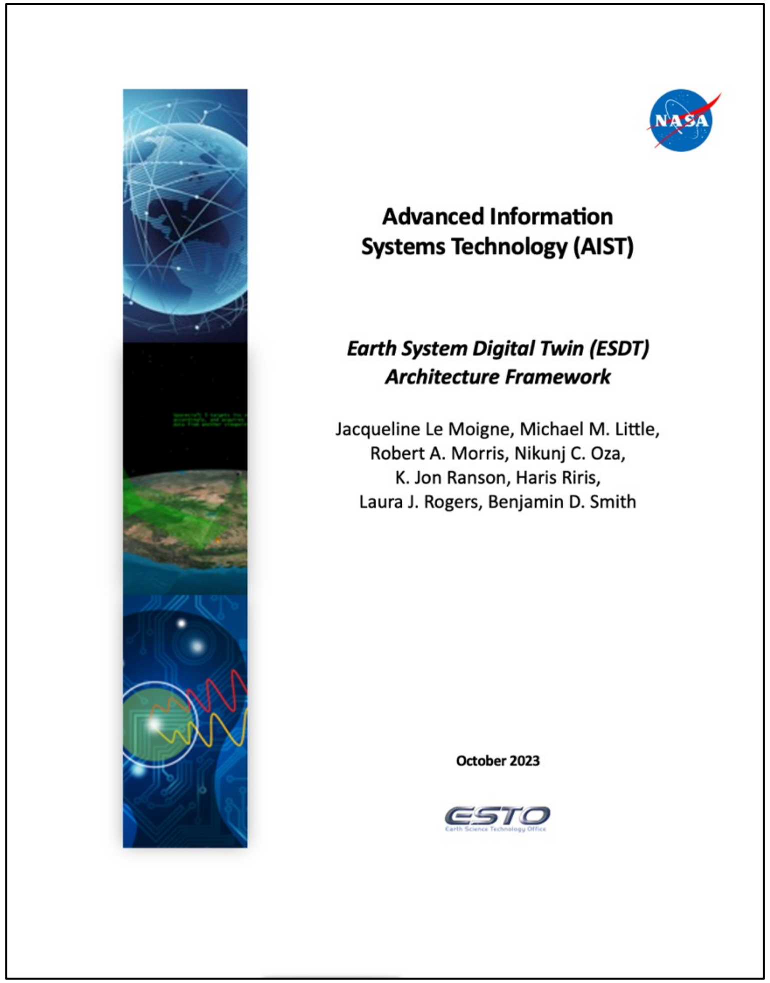
October 26-28, 2022
Washington, D.C.

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=qdpL0Ui-jqc

Document available
on AIST Website:
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

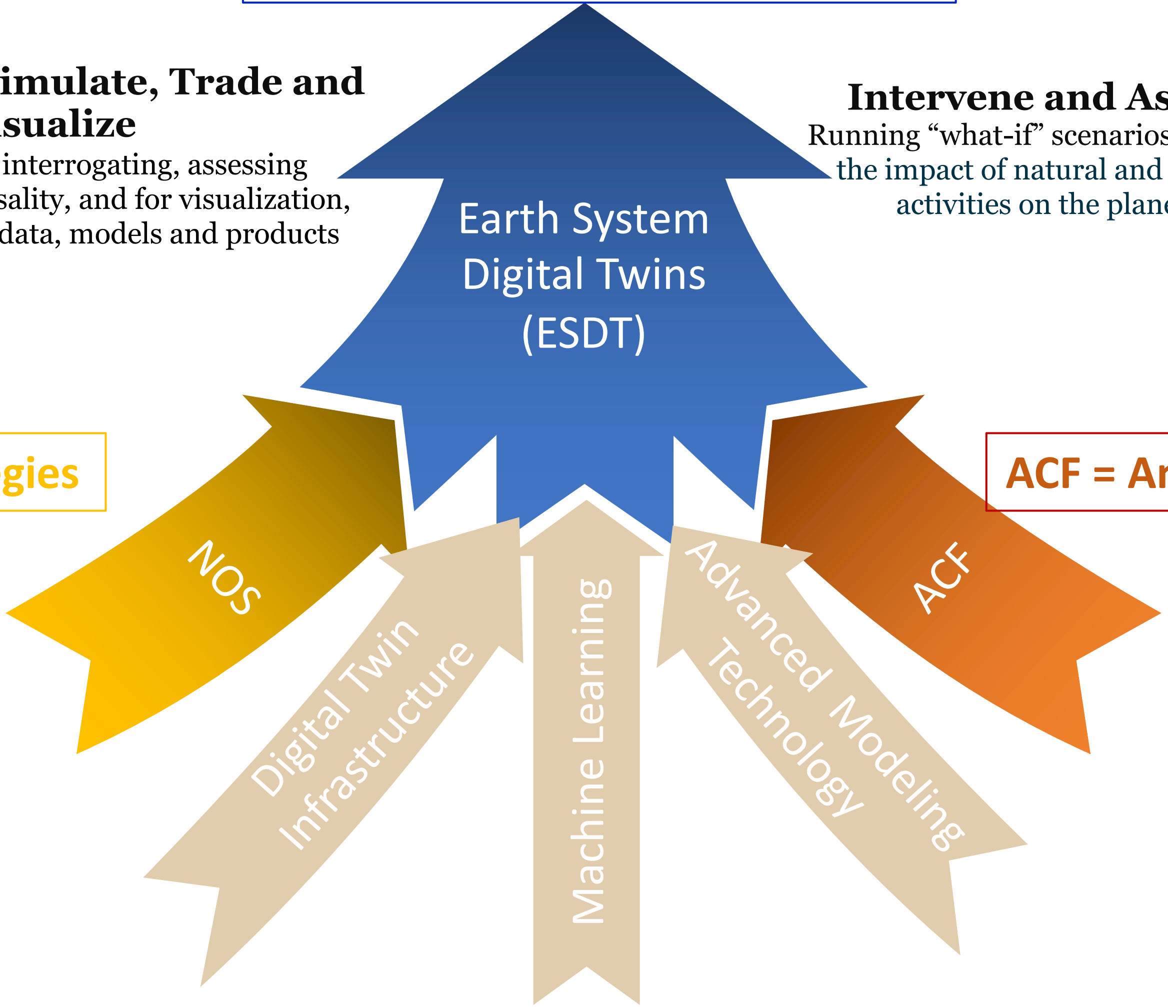
Software and Information Systems Technology for ESDT

ESTO
Earth Science Technology Office
IST

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.



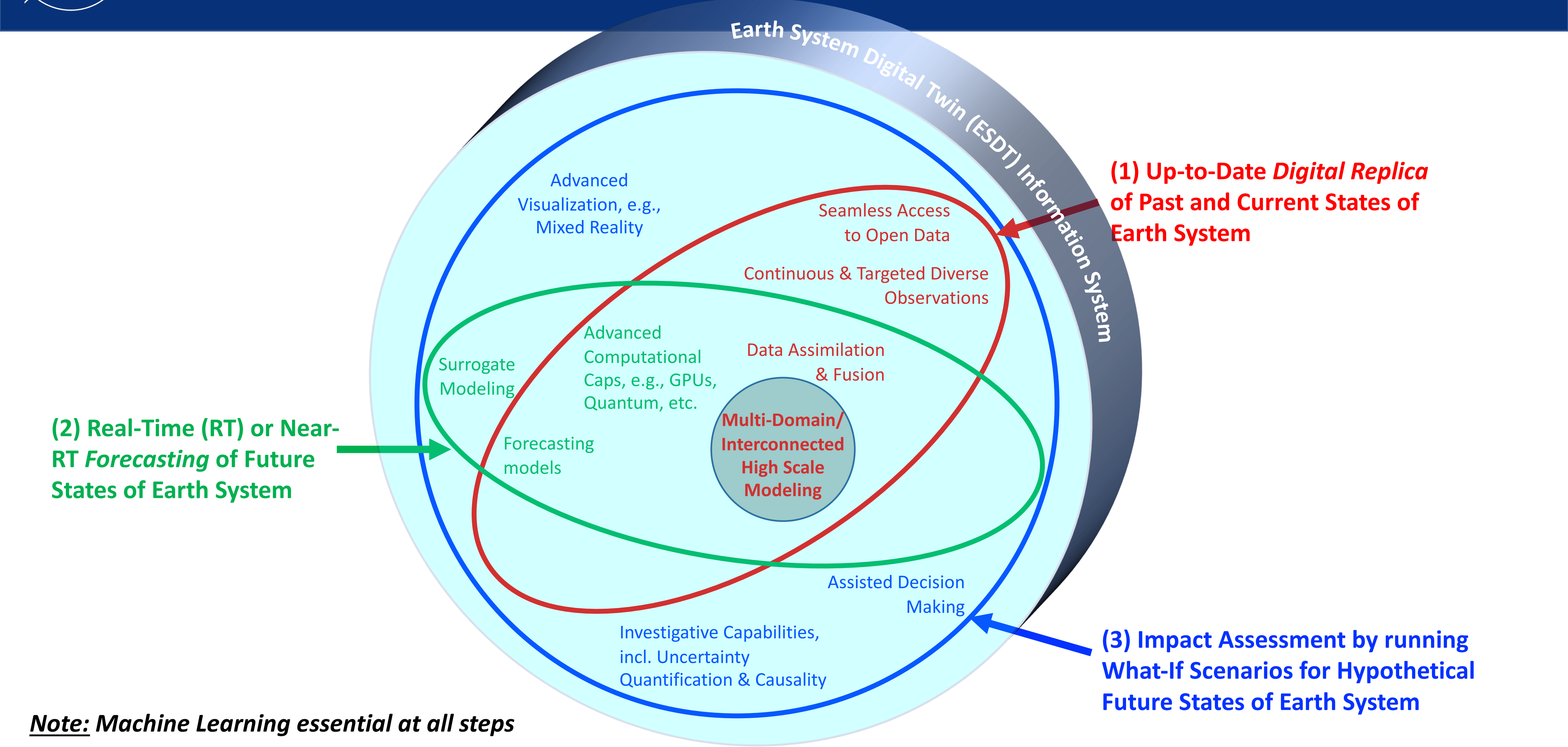
NOS = Novel Observing Strategies

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

ESDT Capabilities and Technologies

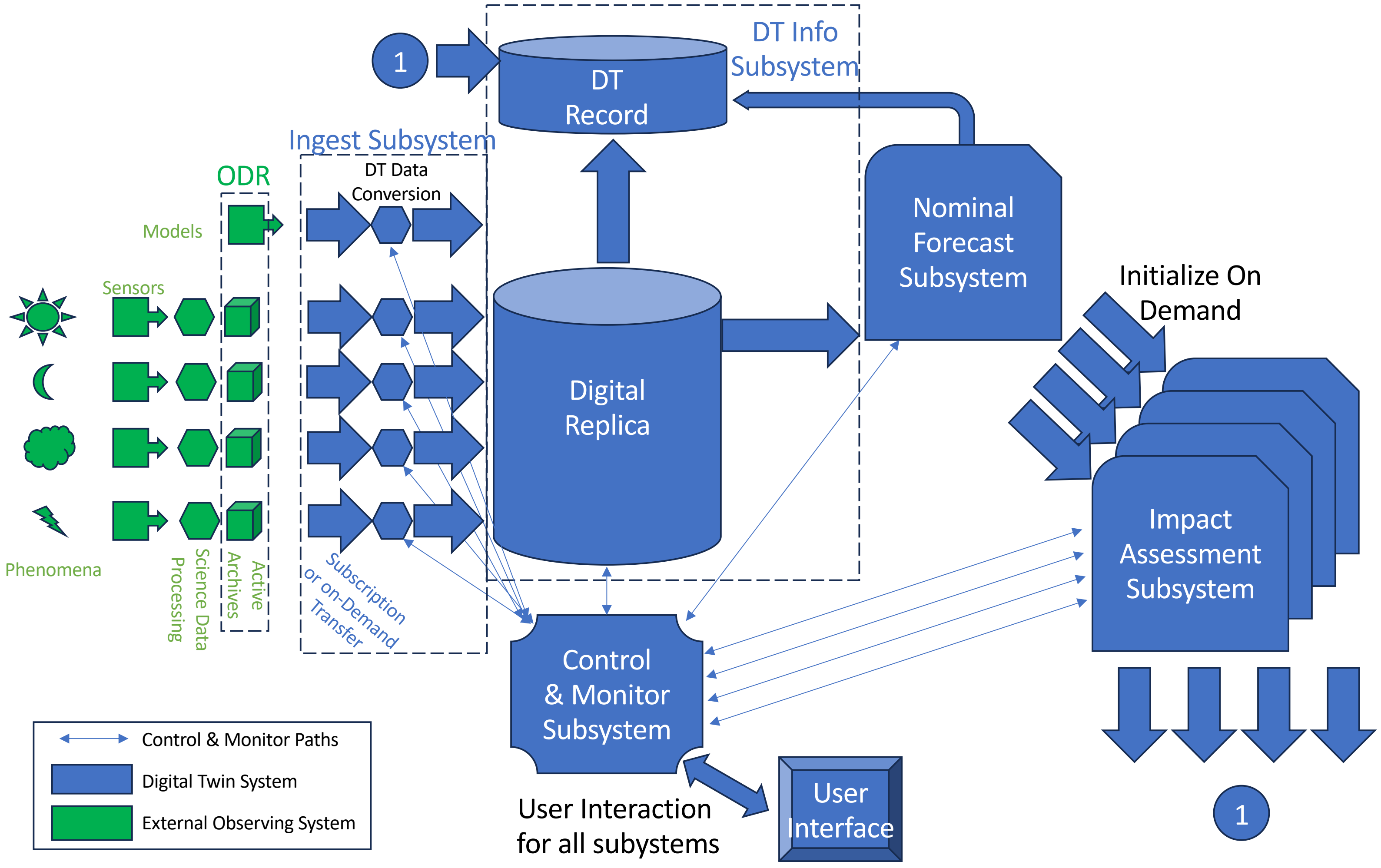


Note: *Machine Learning essential at all steps*

ESDT Conceptual System Diagram

ESTO Earth Science Technology Office IST

An ESDT architecture must consider the full range of components and their relationships



Functional components:

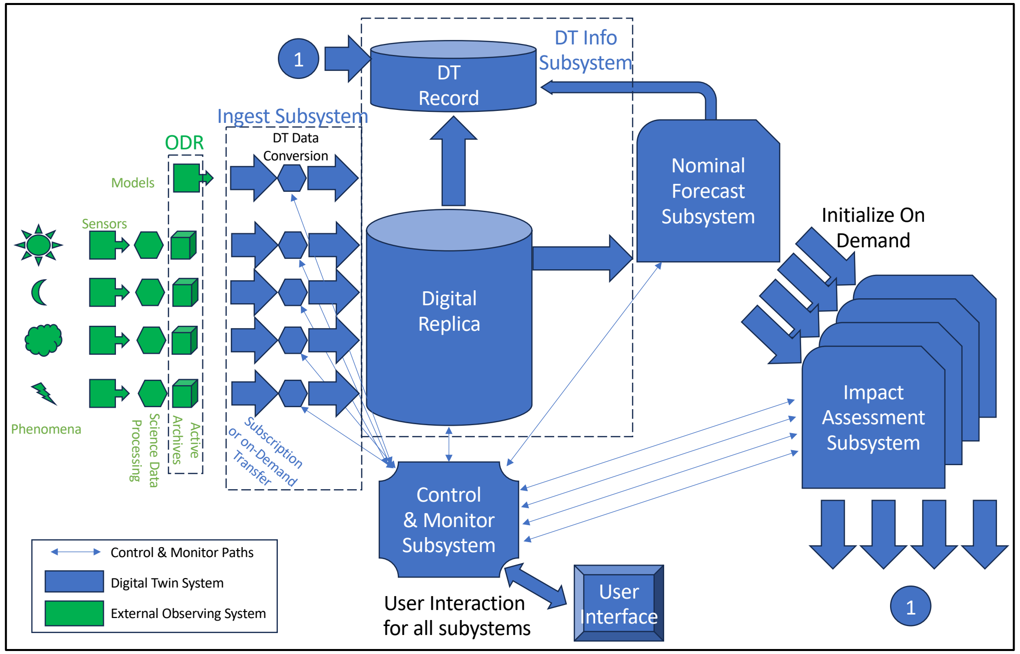
- Observational Data Repository (ODR)
- Ingest Subsystem (ISS)
- DT Information Subsystem (DISS)
- Nominal Forecast Subsystem (NFSS)
- Impact Assessment Subsystem (IASS)
- Control and Monitor Subsystem (CMSS)
- User Interface Subsystem (UISS)

Architecture design may combine components or group them differently

ESDT Architecture Framework
Considerations

IST

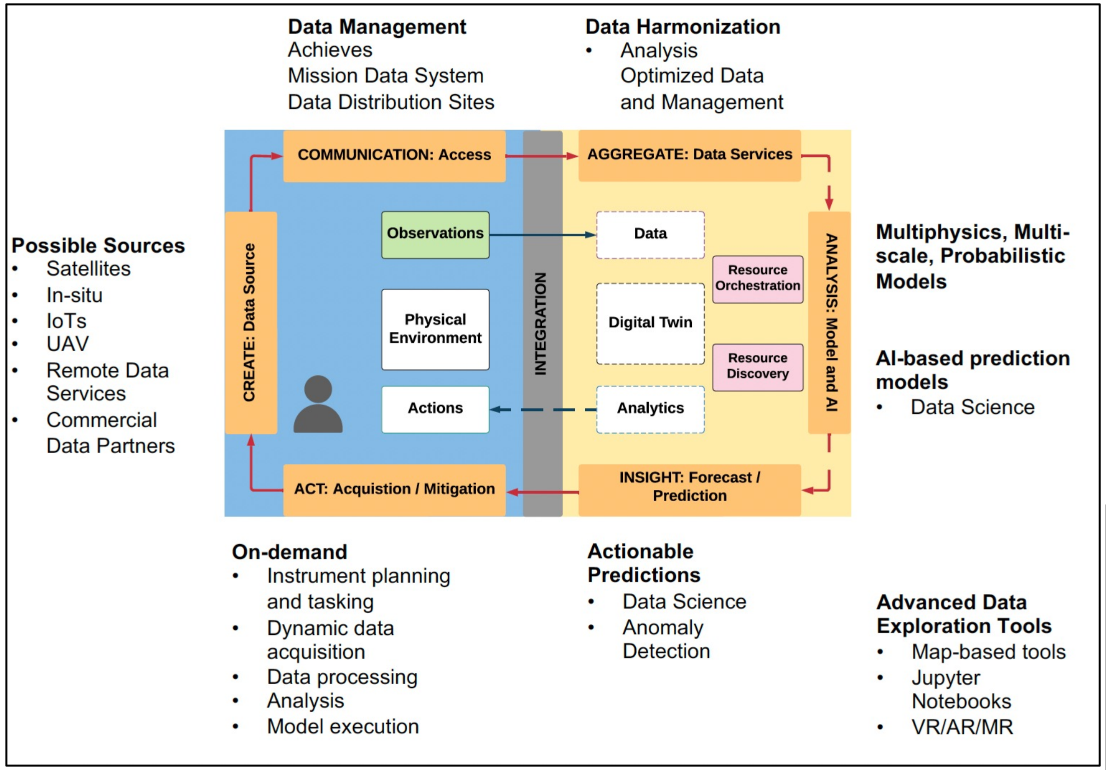
- **Consider architecture principles**
  - Modularity
  - Process automation and error checking
  - Comply with Open-Source Science principles from SPD-41A
  - Permit evolution of concepts and uses and reasonable addition of new components
  - Provide the Glue to stitch together all ESD capabilities
  - 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

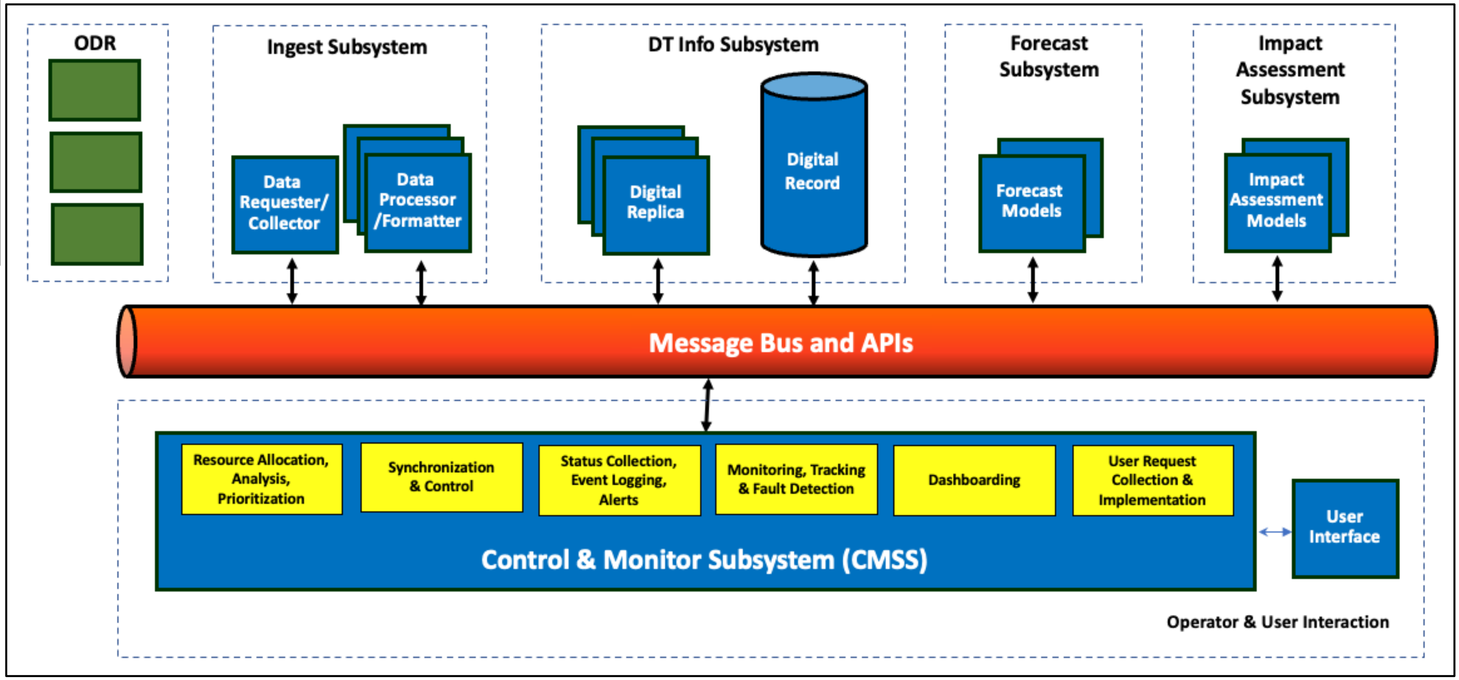
ESDT Architecture Examples

ESTO Earth Science Technology Office
IST



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

ESDT Message Bus – Based Framework
Continuous Digital Replica



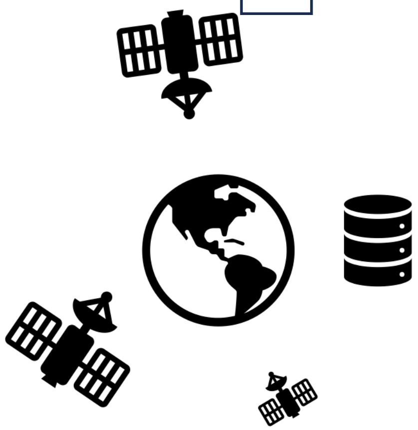
Earth Science Modeling (ESM)

Simplified Pipeline

ESTO
Earth Science Technology Office
IST

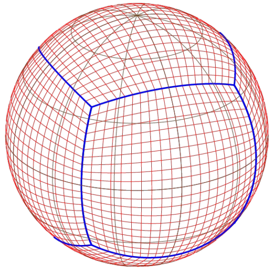
Earth system models (ESMs) combine process-based models of different sub-systems of the Earth system into an integrated numerical model that yields for a given state of a system at time t a prediction of the system state for time t + 1.

Use observation data to evaluate ESM performance



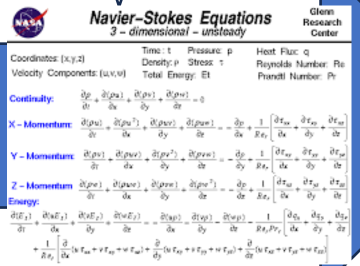
Use data from observations to generate initial state

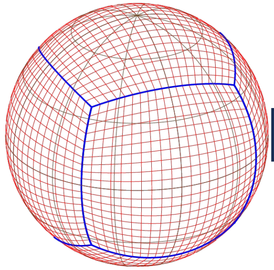
Grid model: Discretization of the continuous equations of motion



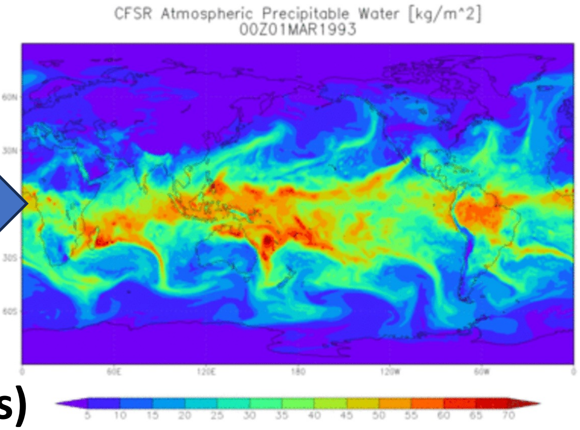
- **Dynamical core + parameterizations***

Run model forward in time





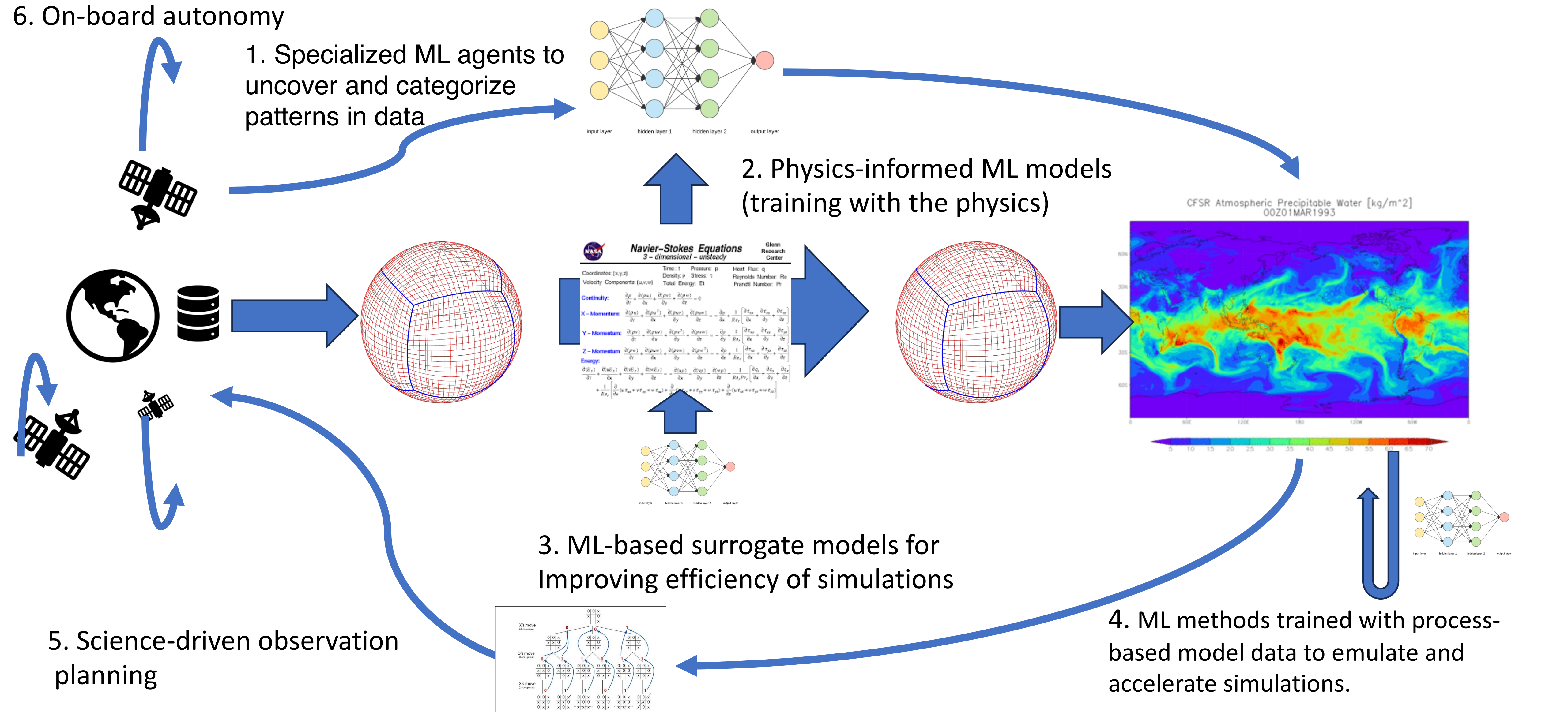
- **Use physics model (e.g. Navier-Stokes)**
- **Equations of motion do not exist for some processes (e.g. physiological processes governing vegetation)**



Data Products

*Simplified representations of natural phenomena used in ESMs. For example, to estimate sub-grid scale processes.

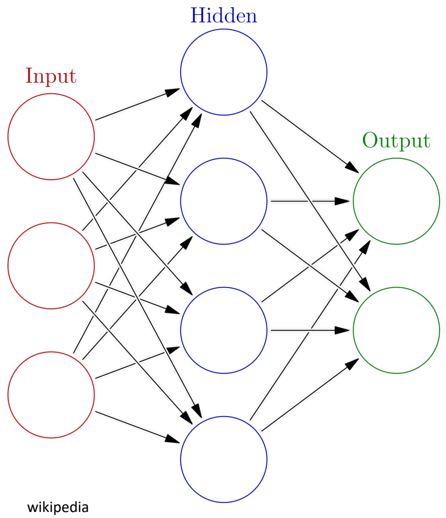
Machine Learning for Earth Science Modeling



Foundation Models

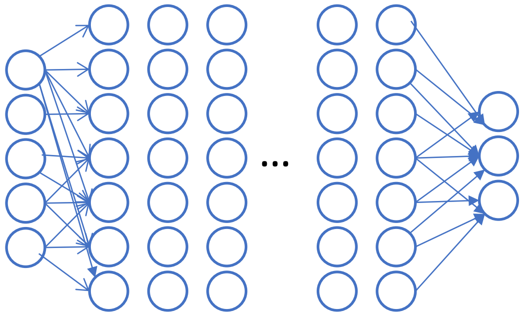
Really Big Neural Networks

IST



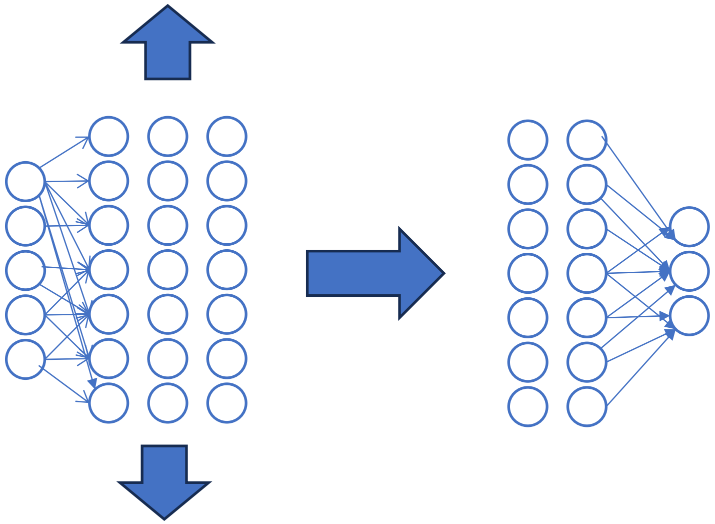
Early Neural Networks

Small networks.
Performed ok but often
eclipsed by other methods.



Deep Neural Networks

10k ~ M parameters
Lots of (labeled) training data
Great performance



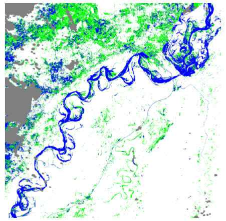
Foundation Models

Millions to Trillions of parameters
Too much training data to hand-label
Expensive to train (but only once)
Newsworthy performance

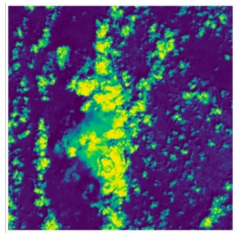
Some ES Applications of Foundation Models

ESTO
Earth Science Technology Office
IST

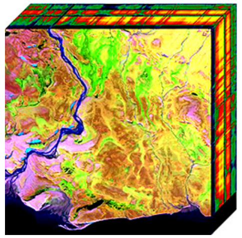
LU/LC



Surface Water Extent



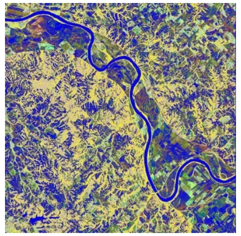
Cloud Masks



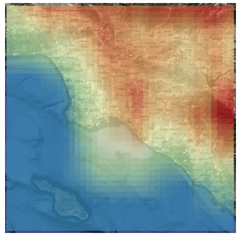
Spectral Unmixing

Recognition and Classification

On the ground and onboard

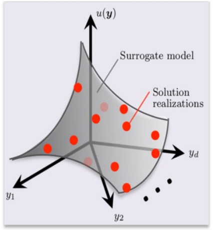


Hydrology



AQ

Forecasting and Nowcasting



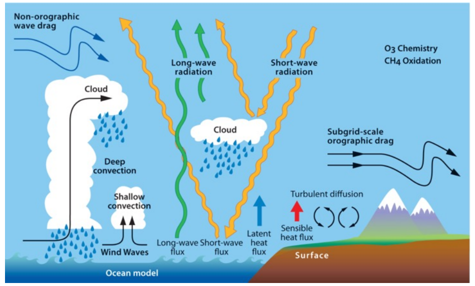
Fast Inversions

Four Cast Net

Fast Models

Surrogate Models

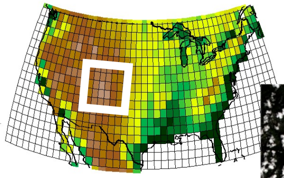
Boundary Conditions



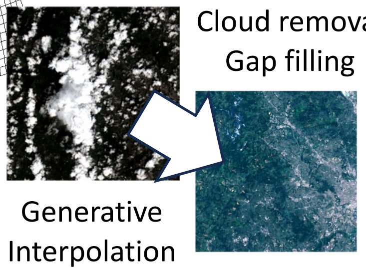
Governing Equations

Parameterizations

Model Understanding and Physics-Inspired ML



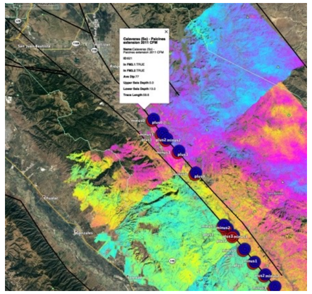
Downscaling



Cloud removal
Gap filling

Generative Interpolation

Interpolation and Reconstruction



Data Fusion

Show me the relationship between SST and algae blooms over the last decade.

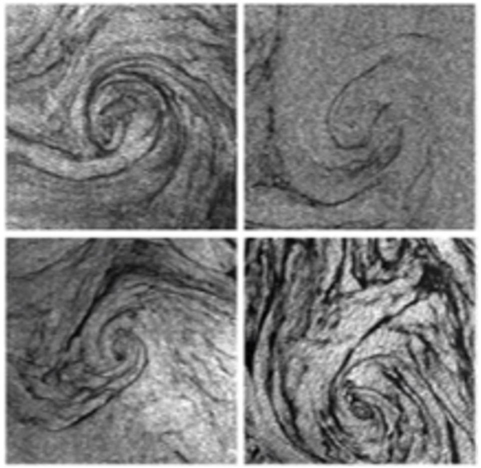
Bespoke NL interfaces

IST Foundation Models R&D

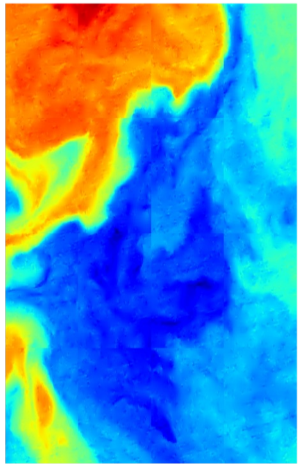
ESTO Earth Science Technology Office IST

SLICE: Semi-supervised Learning from Images of a Changing Earth

Wilson, JPL; AIST-21-0025



Eddy Detection from SAR imagery

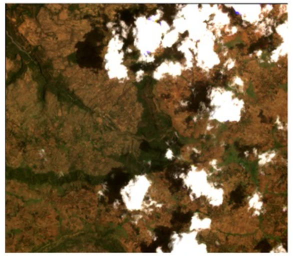


SST reconstruction under clouds

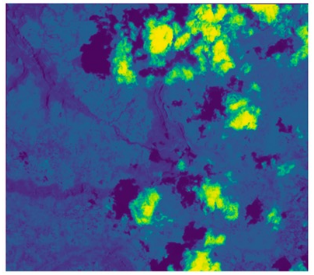
Applications of **Vision Transformers** and **semi-supervised ML** to enable hard remote sensing problems and increase performance despite scarce labeled data.

Coupled Statistics-Physics Guided Learning

Xie, UMD; AIST-21-0068



original



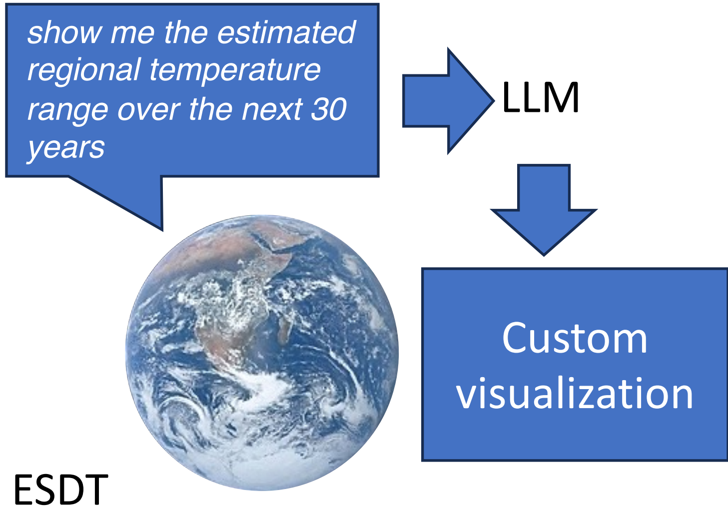
Cloud masks

reconstruction

Semi-supervised learning; physics-guided; and heterogeneity-aware learning.

Digital twin technologies for climate projections

Schmidt, GISS; AIST-QRS-23-0005



Large Language Models (LLM) to generate bespoke data visualizations for user queries

Computational Challenges for ESDT

CZ-DT Example

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

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.

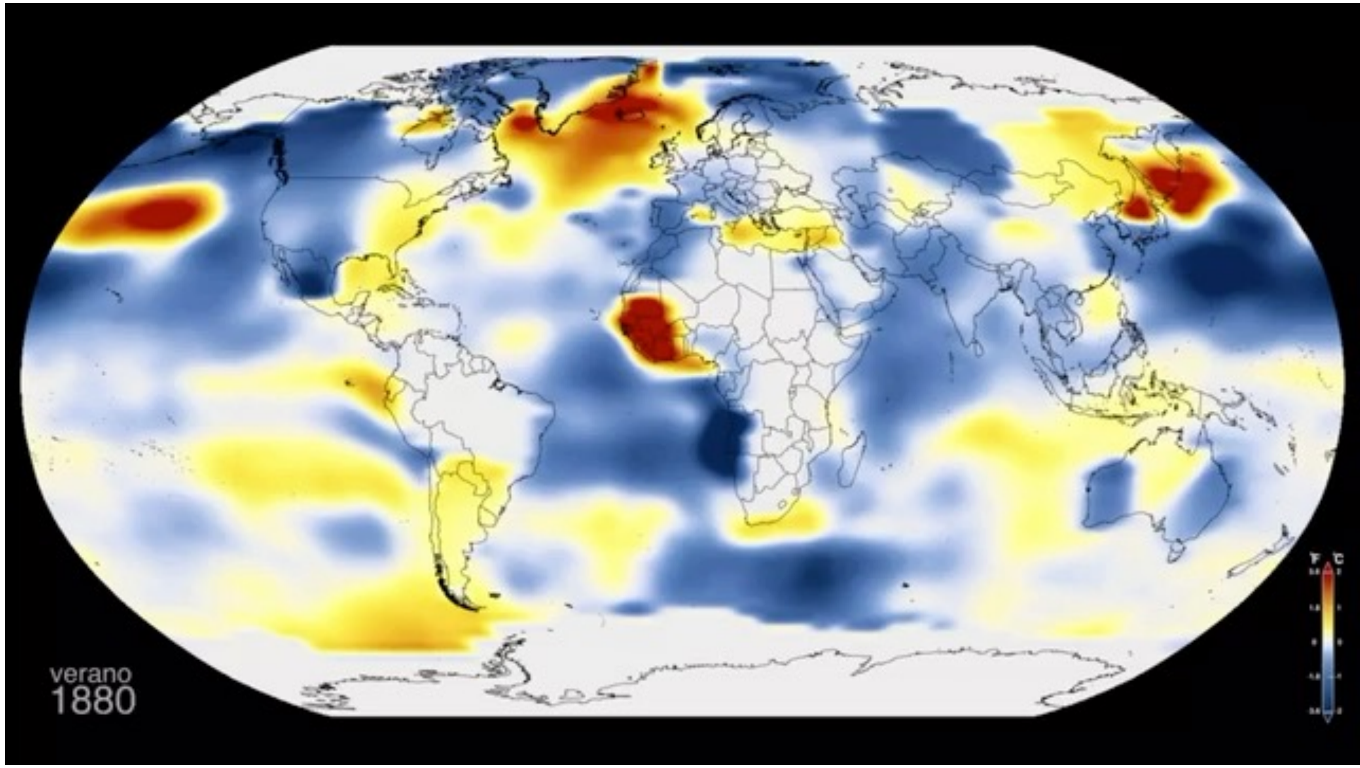
Function	Assumptions
Pre-training (Initial)	24 runs for experiments + 1 final run w/ 800TB data (INTERACTIVE) AI models are constructed using a lot of experiments to try the model settings to get the best results.
Pre-training Updates	Repeat 1 pre-training run and no experiments once per month (INTERACTIVE)
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)

Visualization Challenges for ESDT

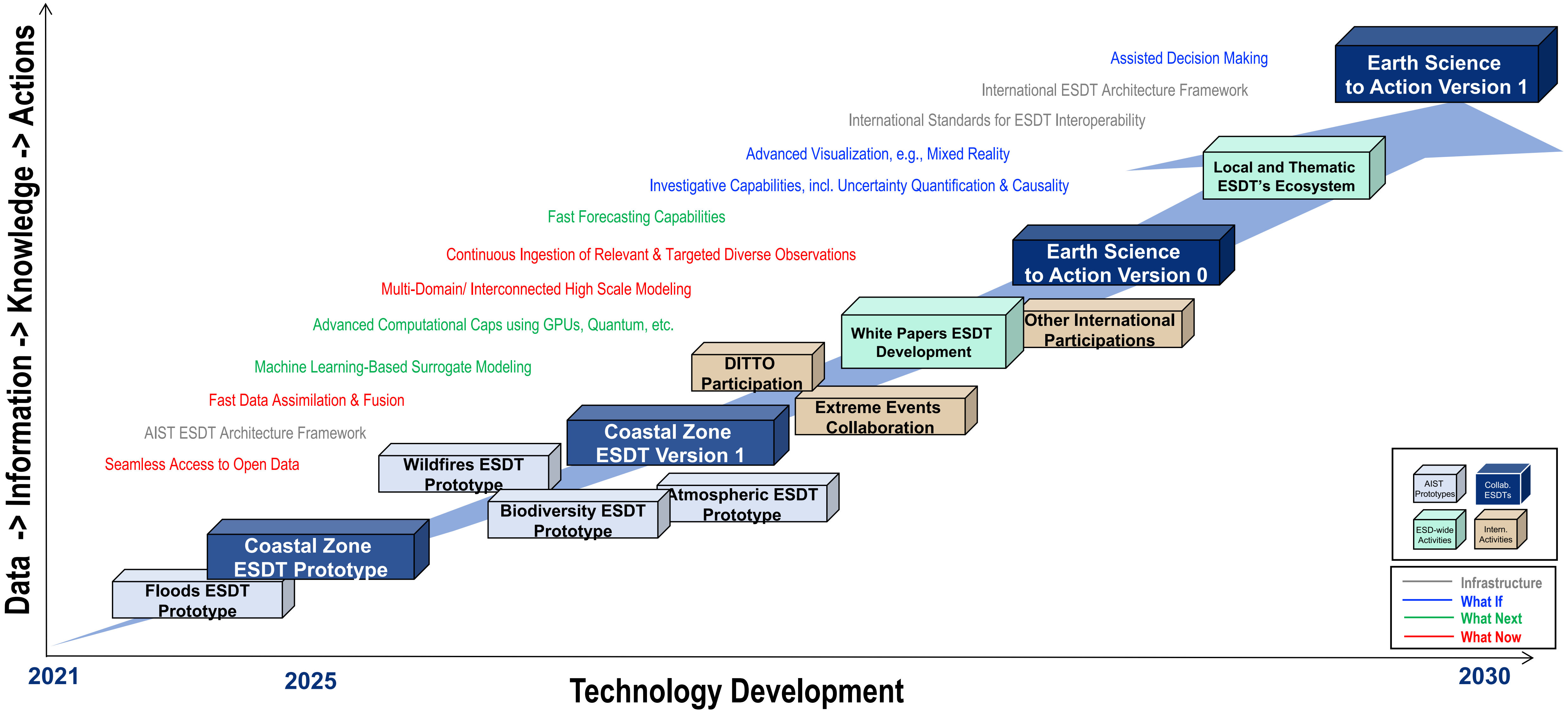
- **User interfaces and visualizations for a range of skill levels and interests (i.e.,"from farmer to scientist")**

- **Various outputs and interactions:**
  - Raw data to complex data products, forecasts and impacts (e.g., risk maps, comparisons, etc.)
  - Real-Time, Near-Real-Time or
  - Regularly produced or on-demand
  - Pre-computed videos and visualizations at selected locations, regions of interest, timeframes, etc.
  - Snapshot Digital Replica at TBD-hours refresh rate
  - Interactive maps, graphs and results with a choice of variables, rendering, etc.
  - Virtual or Extended interactive reality to explore the data and data products

Summer 2023 Record High Global Temperature
NASA's Scientific Visualization Studio
(M. Subbarao, G. Schmidt, E. Hawkins, L. Schuler, I. Jones, E. Kaplan, P. Jacobs)
https://svs.gsfc.nasa.gov/5161/



Earth System Digital Twins (ESDT)
Roadmap



IST


Experience our changing planet
NASA's Earth Information Center
Open: Mon–Fri
NASA

Earth Science Technology Office