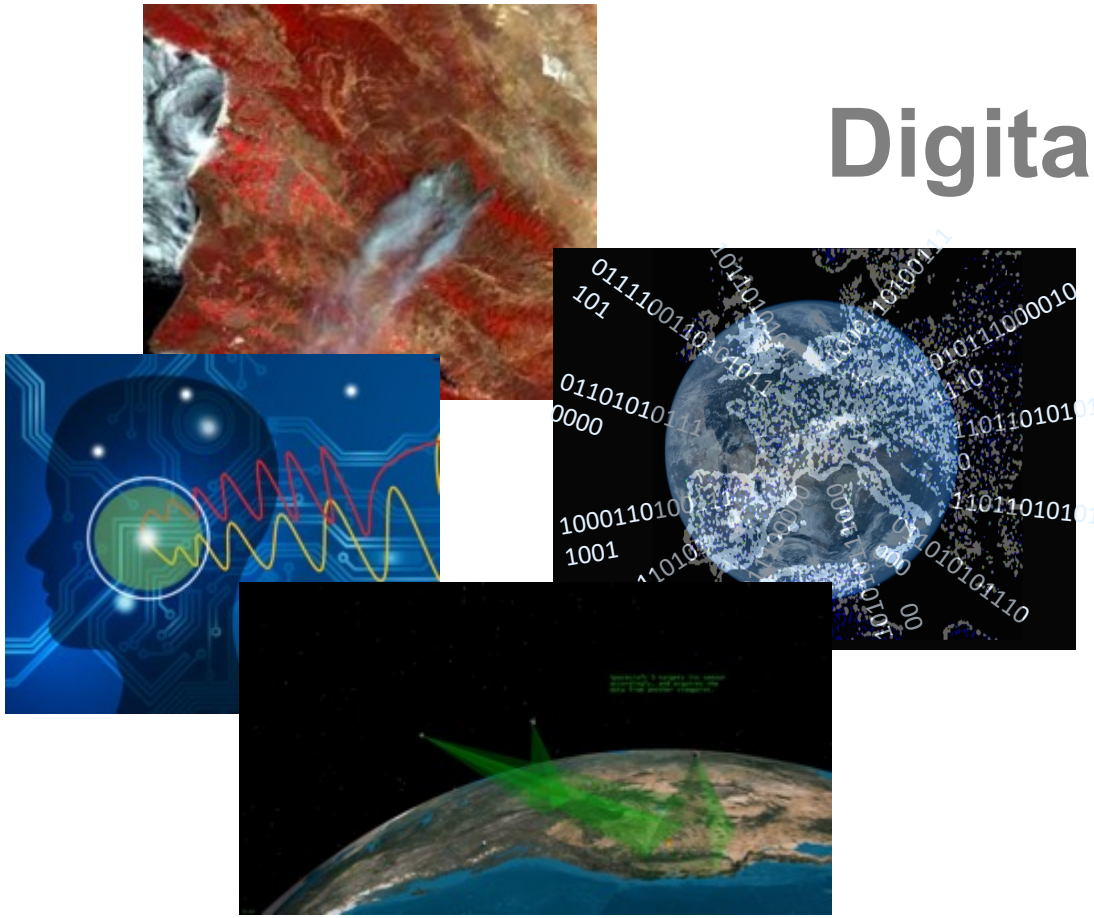


Earth System Digital Twins (ESDT) Technology for NASA Earth Science

Jacqueline Le Moigne
LPS'2022

May 26, 2022

What is a Digital Twin?
Engineering & Healthcare

- **2010/NASA–Michael Grieves and John Vickers:**
  A Digital Twin is a set of virtual information constructs that fully describes a potential or actual physical manufactured product from the micro atomic level to the macro geometrical level. At its optimum, any information that could be obtained from inspecting a physical manufactured product can be obtained from its Digital Twin.

- **2020/IBM:**
  A digital twin is a virtual representation of an object or system that spans its lifecycle, is updated from real-time data, and uses simulation, machine learning and reasoning to help decision-making.

- **(2019/NIH-Interagency Modeling Analysis WG):**
  A digital twin is a digital replica of a living or non-living physical entity, such as a manufacturing process, medical device, piece of medical equipment, and even a person. Utilizing sensor data, digital twins combine simulation and analytics to gain insight into present and future operational states of each physical twin. The resulting digital twin can be wrapped with artificial intelligence, combined with ensembles of similar twins, or used in tandem with other predictive tools to analyze and diagnose operational states and to optimize performance under real-world operating conditions.

  Digital Twin = Physical Science + Digital Thread + Data Science. Digital twins exist at the nexus of physical engineering, data science, and machine learning, and their value translates directly to measurable business outcomes and actionable predictions.

- **2020/Purdy-MIT Sloan:**
  A digital twin is a virtual replica of an object, being, or system that can be continuously updated with data from its physical counterpart.

Good Reference Website: https://www.digitaltwinconsortium.org/glossary/index.htm

Digital Twin(s) of the Earth
"Proxies of the Physical World"

- ***2020/Loekken, Le Saux & Aparicio-ESA:***
  ESA DTE = A dynamic interactive replica of the past, present and future of our planet in the digital domain based on an effective integration of observations (satellite, in-situ, IoT and socioeconomic data), Earth-system science and simulations, the bridge to impact sectors science and simulations and artificial intelligence methodologies

- ***2020/Bauer-Destination Earth, ESA:***
  A digital twin of Earth is an information system that exposes users to a digital replication of the state and temporal evolution of the Earth system constrained by available observations and the laws of physics.

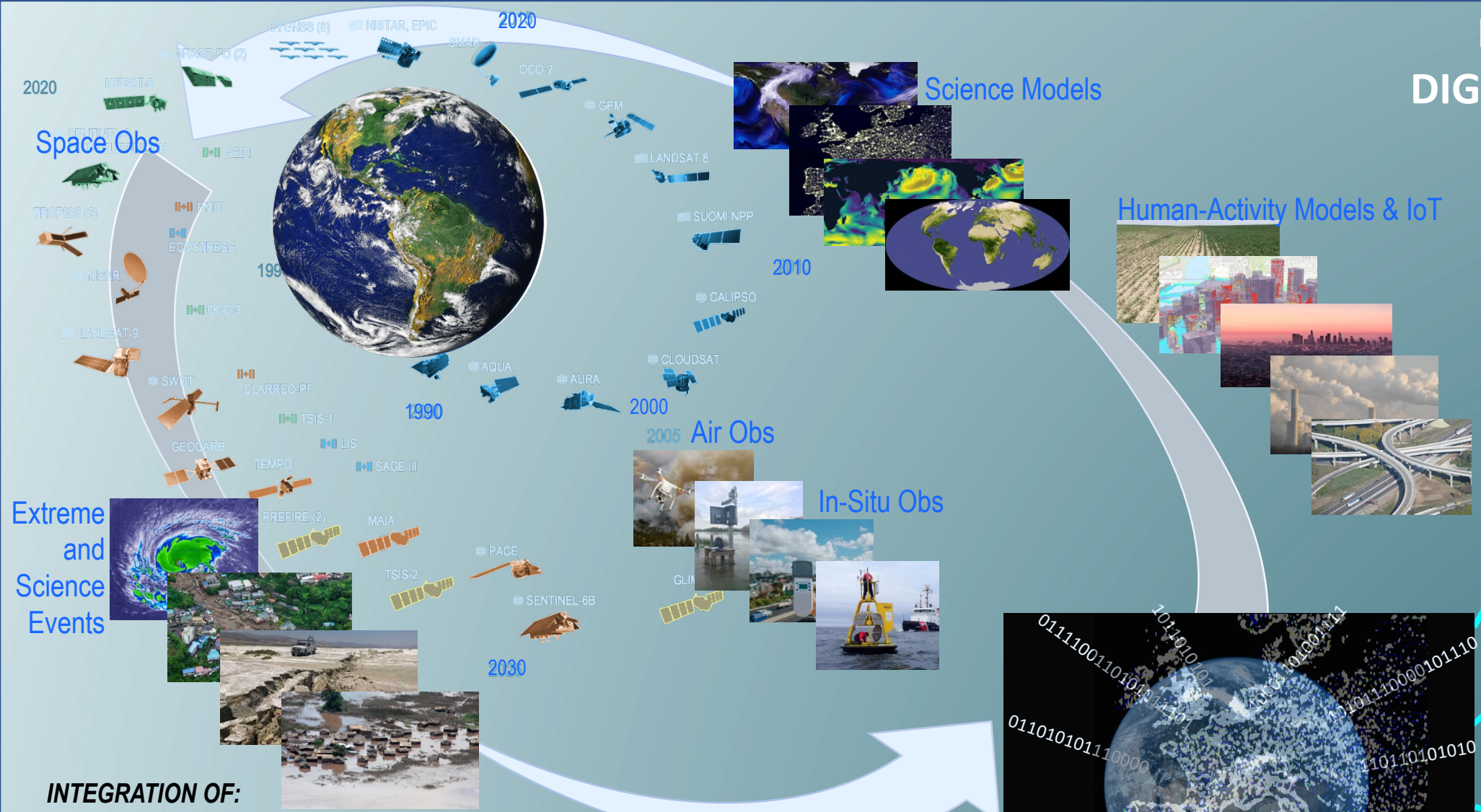
  "Destination Earth" – ESA + ECMWF/EUMETSAT/ETH Zurich/CSCS, 10-year Initiative (Bauer et al., March 2021):
  1. Create a highly accurate digital model of the Earth to map climate development and extreme events as accurately as possible in space and time.
  2. Continuously incorporate observational data and relevant human activities data
  3. This digital twin of the Earth is intended to be an information system that develops and tests scenarios to show more sustainable development and better inform policies.

AIST Definition: An **Earth System Digital Twin (ESDT)** is an interactive and integrated multidomain, multiscale, digital replica of the state and temporal evolution of Earth systems. It 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; as well as
- Analytics and artificial intelligence tools.

An ESDT is a type of integrated information system that ***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. For example, it enables continuous assessment of impact from naturally occurring and/or human activities on physical and natural environments.*

EARTH SYSTEM DIGITAL TWINS (ESDT) TECHNOLOGY



Enable "Constant Finger on the Pulse of the Planet"

Actionable Predictions

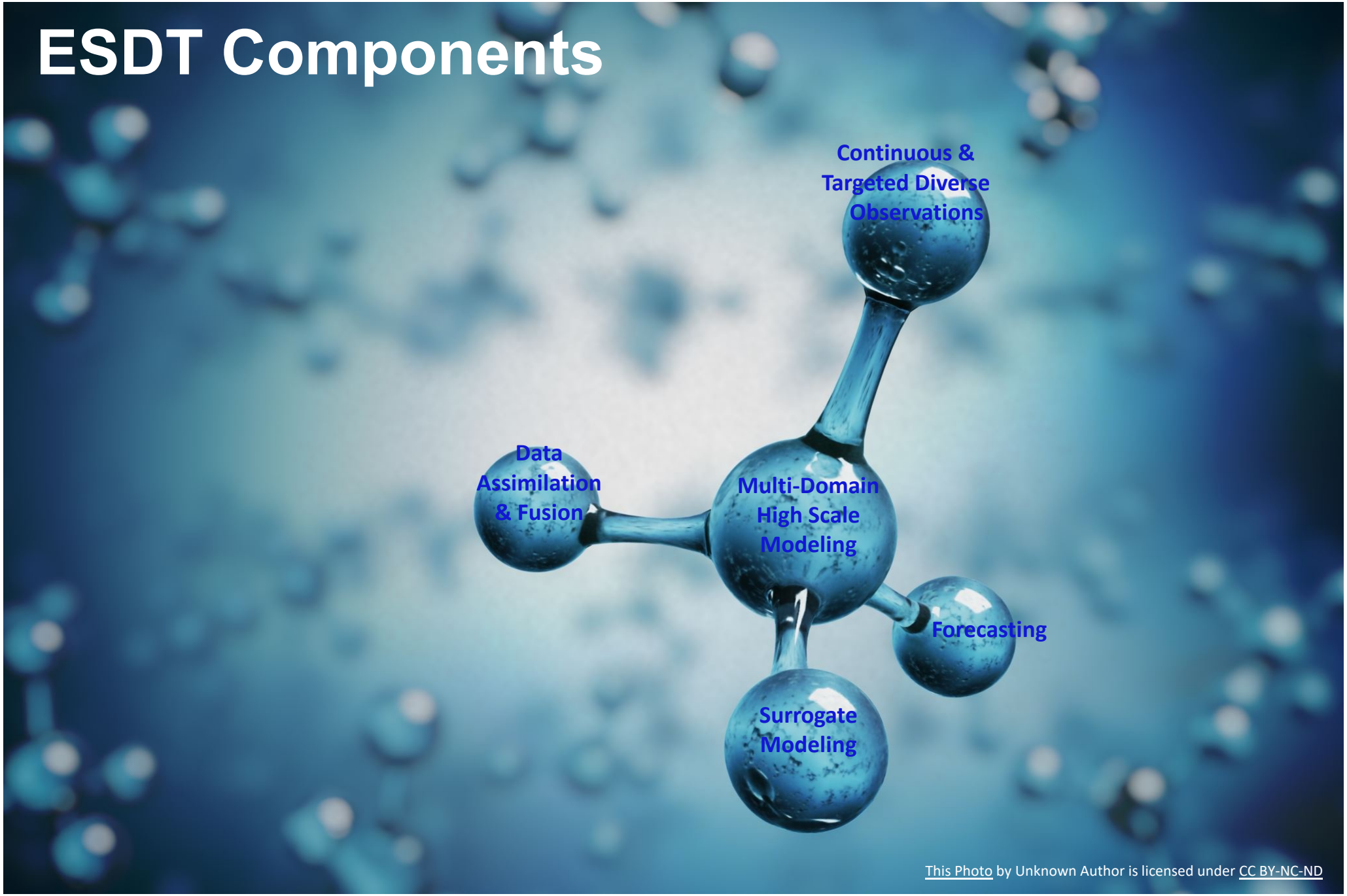
What-If Scenarios

Dynamic Interactive Replica of State and Temporal Evolution of the Earth or a Subcomponent of the Earth

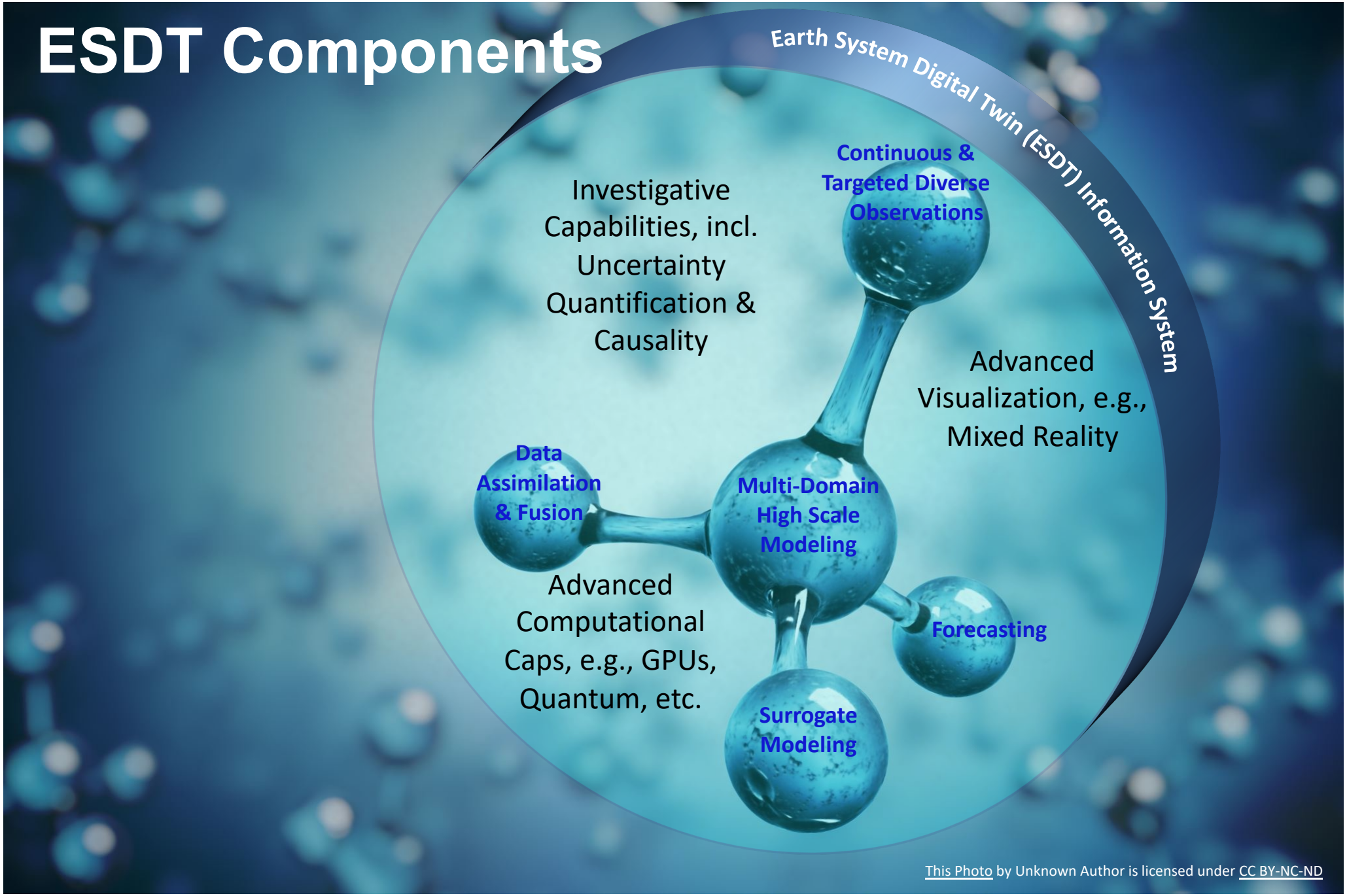
INTEGRATION OF:

+ ***Continuous Observation and Data Stream Optimization***
+ ***Asset Identification, Access & Management (IAM)***
+ ***Interconnected Modeling***
+ ***Machine Learning***
+ ***High-End Computing Environments***
+ ***Investigative Capabilities (incl. Uncertainty Quantification)***
+ ***Advanced Visualization & AR/VR/MR***

ESDT Components



ESDT Components



AIST Objectives

Innovate in technologies that enable:

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

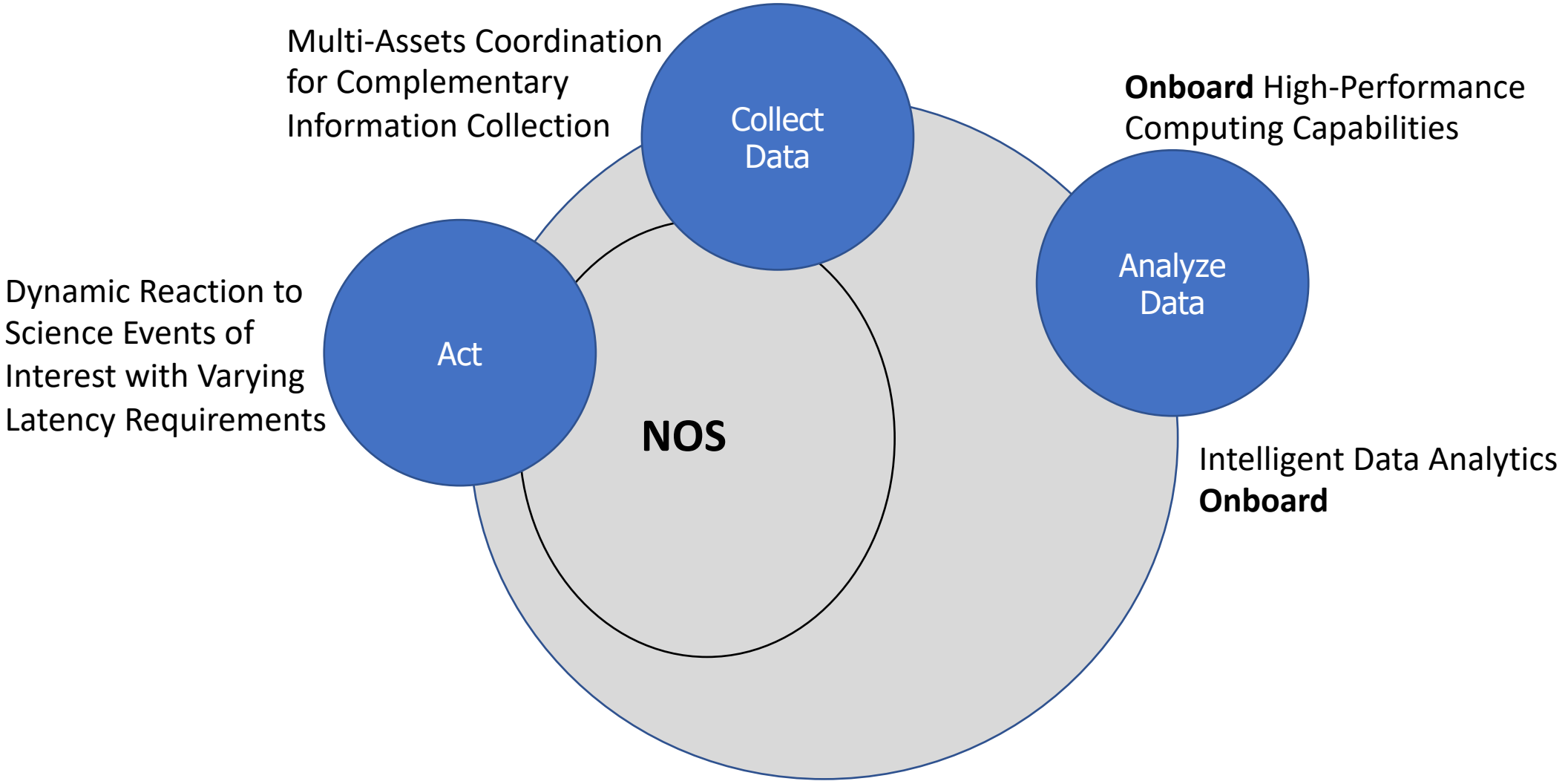
O2. Agile science investigations that fully utilize the large amount of diverse observations using advanced analytic tools, visualizations, and computing environments, and that interact seamlessly with relevant observing systems
=> Analytic Collaborative Frameworks (ACF)

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)

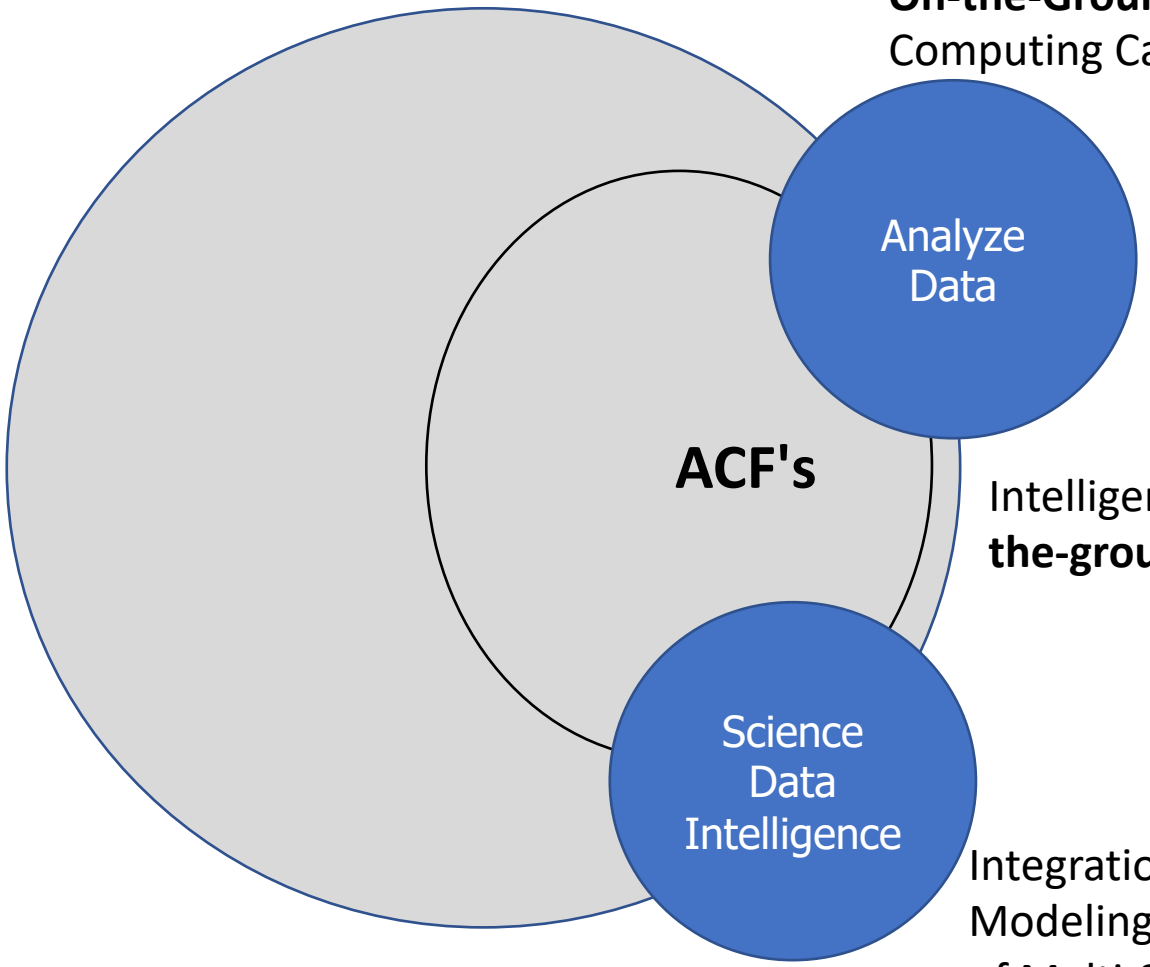
More generally, build the tools that provide "Science Data Intelligence"

* "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 New Observing Strategies (NOS)



AIST Analytic Collaborative Frameworks (ACF)

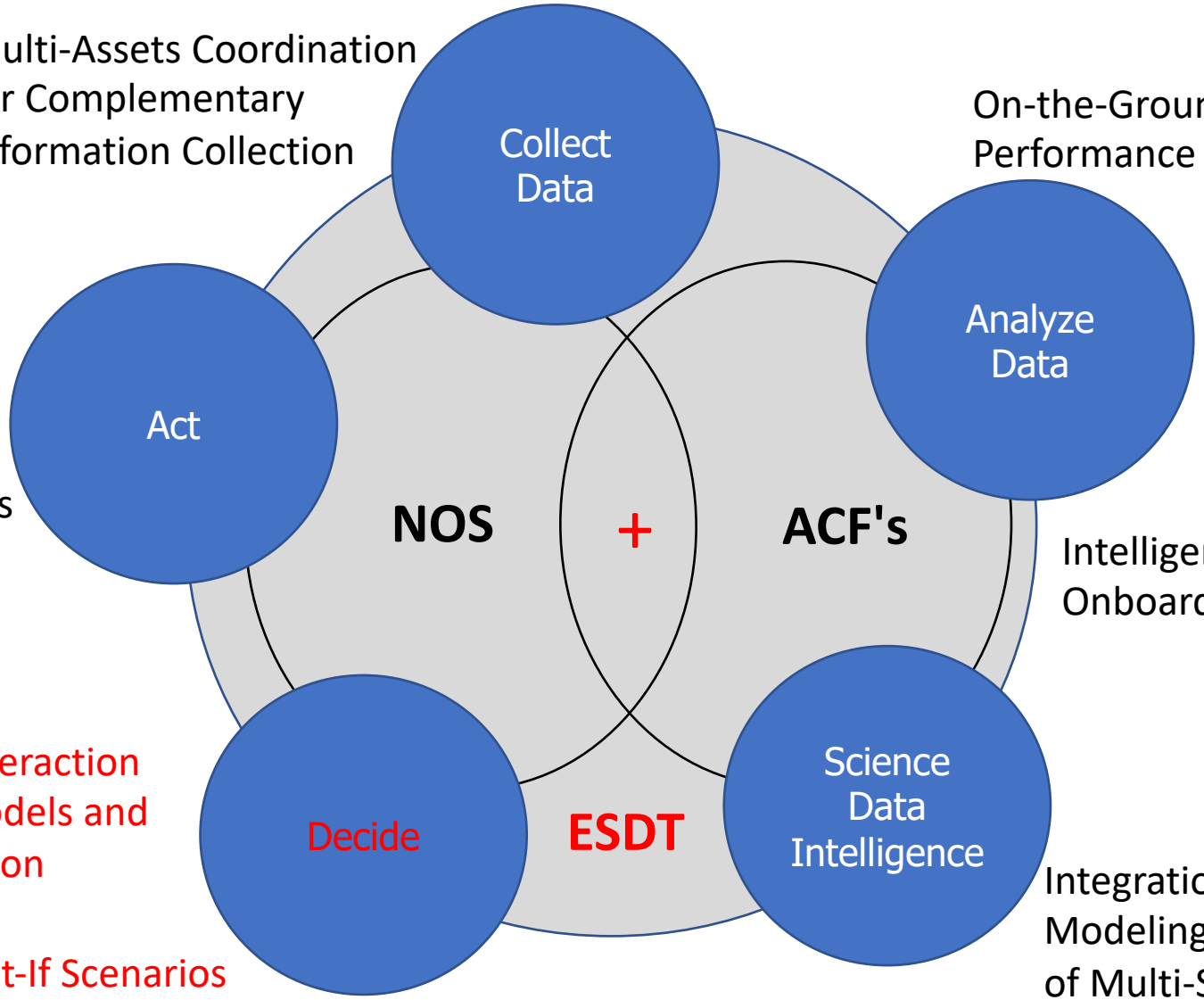


On-the-Ground High-Performance Computing Capabilities

Intelligent Data Analytics **On-the-ground**

Integration, Fusion, Visualization, Modeling and Analysis of Large Amounts of Multi-Source Data

Earth System Digital Twins (ESDT): New AIST-21 Thrust – *Continuous Integration of NOS and ACF Techs*



Multi-Assets Coordination for Complementary Information Collection

On-the-Ground and Onboard High-Performance Computing Capabilities

Dynamic Reaction to Science Events of Interest with Varying Latency Requirements

Intelligent Data Analytics Onboard and On-the-ground

Seamless Interaction between Models and Data Collection

What-If Scenarios

Integration, Fusion, Visualization, Modeling and Analysis of Large Amounts of Multi-Source Data

ESDT Technologies Requested in AIST-21

- Technologies for agile interaction and interoperability between measurement acquisition (NOS or NOS-like) and science investigations (ACF or ACF-like)
- Frameworks that enable data ingest from multiple, integrated models, and/or moving from mono-discipline to multi-discipline inter-related systems
- Leveraging of Model-Based System Engineering (MBSE) frameworks for the development and sustainment of Earth Systems Digital Twins, especially based on the integration and coordination of NOS and ACF systems
- Digital Thread developments to link all digital twin capabilities (design, performance data, product data, operational status data, event status data), to enable design requirements, records, provenance, and system reorientations to be easily reviewed and address issues within the digital twin system
- Concepts and technologies for developing "federated ESDTs" in which multiple individual ESDTs interact and can be integrated as the layers of broader ESDTs
- Novel AI (not limited to ML) techniques enabling systems to quickly request, integrate, and fuse diverse and timely Earth observations into ESDTs
- Investigative technologies to facilitate "what-if" investigations inherent to ESDT systems, including but not limited to:
  - Multi-scale simulations, statistics, uncertainty quantification, and causality methodologies
  - Computational algorithms and methodologies involving high-end computing, such as GPUs or other hardware systems that will enable running large permutations of what-if scenarios using large amounts of data and high-resolution and high-fidelity models
  - Statistical methodologies that optimize the computational efficiency of such "what-if" investigations
  - Innovative and simple user interfaces and visualization methods based on Augmented Reality (AR) and Mixed Reality (MR) techniques and capable of visualizing complex systems of systems

AIST-21 ESTD Awards

https://esto.nasa.gov/project-selections-for-aist-21

- **ESDT Infrastructure**

PI's Name	Organization	Title	Synopsis
Tanu Malik	De Paul University	Reproducible Containers for Advancing Process-oriented Collaborative Analytics	Aims to establish reproducible scientific containers that are easy-to-use and are lightweight. Reproducible containers will transparently encapsulate complex, data-intensive, process-oriented model analytics, will be easy and efficient to share between collaborators, and will enable reproducibility in heterogeneous environments.
Thomas Grubb	NASA Goddard Space Flight Center (GSFC)	Goddard Earth Observing System (GEOS) Visualization And Lagrangian dynamics Immersive eXtended Reality Tool (VALIXR) for Scientific Discovery	Proposes to develop a scientific exploration and analysis mixed augmented and virtual reality tool with integrated Lagrangian Dynamics (LD) to help scientists identify, track, and understand the evolution of Earth Science phenomena in the NASA GEOS model. It will provide both a scientific discovery tool and a model analysis and improvement tool.
Matthias Katzfuss	Texas A&M University (TAMU)	A scalable probabilistic emulation and uncertainty quantification tool for Earth-system models	Proposes to develop a fully automated toolbox for uncertainty quantification in Earth-system models, to provide insight into the largest and most critical information gaps and identify where potential future observations would be most valuable. It would allow interpolation between observed covariate values and running extensive what-if scenarios.
Thomas Clune	NASA Goddard Space Flight Center (GSFC)	A Framework for Global Cloud Resolving OSSEs	Will enable global, cloud-resolving Observing System Simulation Experiments (OSSEs) by addressing key computational challenges to enable existing technologies to scale to the spatial resolutions needed by the end of decade, e.g., extending parallel I/O capabilities, adopting a 2-phase Nature Run approach and a flexible API for customization.

- **AI-Surrogate Modeling for ESDT**

PI's Name	Organization	Title	Synopsis
Christopher Keller	Morgan State University (MSU)	Development of a next-generation ensemble prediction system for atmospheric composition	Proposes to develop a next-generation modeling framework for the real-time simulation of reactive gases and aerosols in the atmosphere. Will deploy computationally efficient parameterizations of atmospheric chemistry and transport and will develop generative models based on machine learning (ML) to predict model uncertainties.
Jouni Susiluoto	NASA Jet Propulsion Laboratory (JPL)	Kernel Flows: emulating complex models for massive data sets	Proposes a general-purpose, versatile emulation tool to provide fast, accurate emulation with little tuning, to scale up to very large training sets, and to provide uncertainties associated with outputs. This tool set will facilitate large-scale implementation of forward modeling and retrievals, and of UQ at production scales.

AIST-21 ESTD Awards (cont.)

https://esto.nasa.gov/project-selections-for-aist-21

- **ESDT Prototypes**

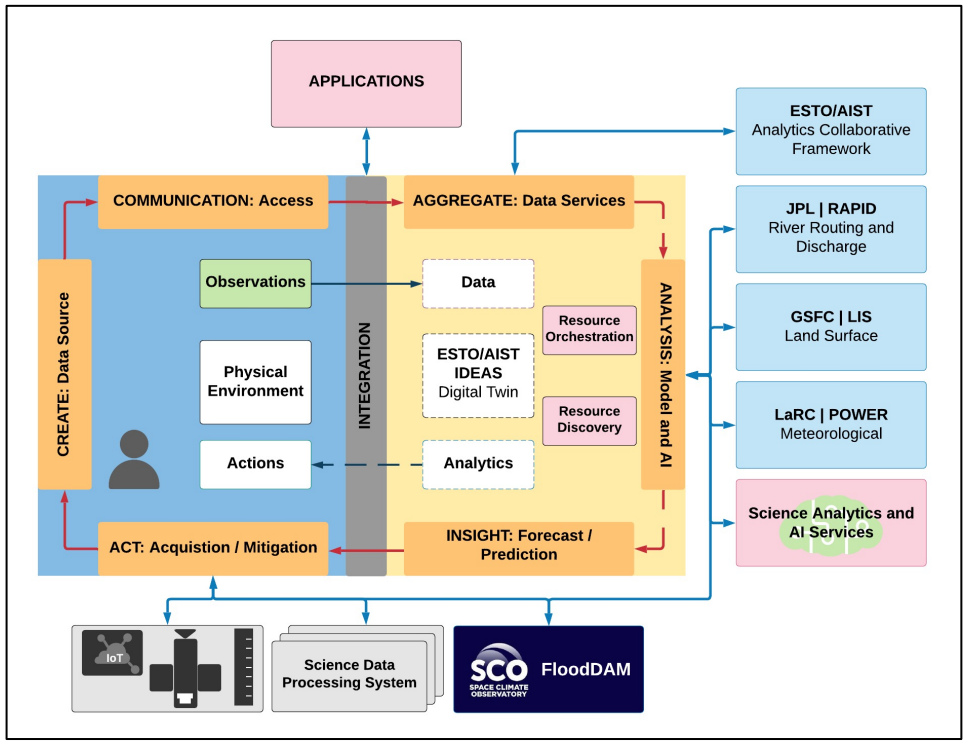
PI's Name	Organization	Title	Synopsis
Milton Halem	University of Maryland, Baltimore County (UMBC)	Towards a NU-WRF based Mega Wildfire Digital Twin: Smoke Transport Impact Scenarios on Air Quality, Cardiopulmonary Disease and Regional Deforestation	Will develop and implement a Regional Wildfire Digital Twin (WDT) model with a sub-km resolution to enable the conduct of mega wildfire smoke impact scenarios at various spatial scales and arbitrary locations over N. America. WDT will provide a valuable planning tool for impact scenarios by season, location, intensity, and atmospheric state.
Rajat Bindlish	NASA Goddard Space Flight Center (GSFC)	Digital Twin Infrastructure Model for Agricultural Applications	Will develop an agriculture productivity modeling system over Continental United States as an example of incorporating representations of infrastructure-oriented process, for the understanding, prediction, and mitigation/response of Earth system process variability, with application to crop growth, yield, and agricultural production information, critical to commodity market, food security, economic stability, and government policy formulation.
Craig Pelissier	(SSAI)	Terrestrial Environmental Rapid-Replicating Assimilation Hydrometeorology (TERRAHydro) System: A machine-learning coupled water, energy, and vegetation terrestrial Earth System Digital Twin	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.

- **ACF Towards ESDT**

PI's Name	Organization	Title	Synopsis
Arlindo Da Silva	NASA Goddard Space Flight Center (GSFC)	An Analytic Collaborative Framework for the Earth System Observatory (ESO) Designated Observables	Will develop an Analytic Collaborative Framework for the Earth System Observatory (ESO) missions, based on realistic, science-based observing system simulations and the Program of Record (PoR), tied together in a cloud-based cyberinfrastructure. Create a 3D, holistic view of Earth with all ESO unique satellites.
Thomas Allen	Old Dominion University	Pixels for Public Health: Analytic Collaborative Framework to Enhance Coastal Resiliency of Vulnerable Populations in Hampton Roads, Virginia (VA)	Proposes to design and operationally demonstrate a system linking the VA Open Data Cube, a socio-spatial-health information “Digital Neighborhood” (Hampton Roads Biomedical Research Cons.), hydrodynamic models, and in-situ flood sensor network. Will connect observational and physical environmental domains with human vulnerability.
Thomas Huang	NASA Jet Propulsion Laboratory (JPL)	Fire Alarm: Science Data Platform for Wildfire and Air Quality	Proposes to advance AIST's Air Quality Analytics Collaborative Framework (AQACF) to establish a wildfire and air quality ACF, Fire Alarm, focusing on the prediction and analysis of wildfire, burned area and the air quality as an integrated platform to guide decision-makers, science researchers, and first-responders.

Integrated Digital Earth Analysis System (IDEAS) – *AIST Collaboration with CNES*

- **IDEAS is a NASA ESTO/AIST Earth System Digital Twin** project that bridges the physical environment and its virtual representation by continuously assimilating new observations to improve forecast and prediction for integrated science and decision support
- Using water cycle and flood analysis as the prototype application to integrate NASA, CNES, and Space Climate Observatory (SCO) data and science
- Multi-Agency and Multi-Center partnership
- Advanced numerical models and analysis
  - **JPL's RAPID**: Routing Application for Parallel computation of Discharge
  - **GSFC's LIS**: Land Information System
  - **LaRC's POWER**: Prediction of Worldwide Energy Resources
  - **CNES and SCO's FloodDAM:** Automated service to reliably detect, monitor and assess flood events globally
  - Integration with **NASA IPCC Sea Level Prediction** data for coastal flooding
  - Joint developed and trained flood detection and prediction machine learning algorithms
  - Promote and advance interoperable standards
- Improve the Machine Learning flood prediction model (Huang *et al.* 2020) from the JPL-CNES Joint Data Science pilot
- Scenario-based prediction for infrastructure and population impacts
- Ongoing formulation and planning with CNES and SCO's FloodDAM



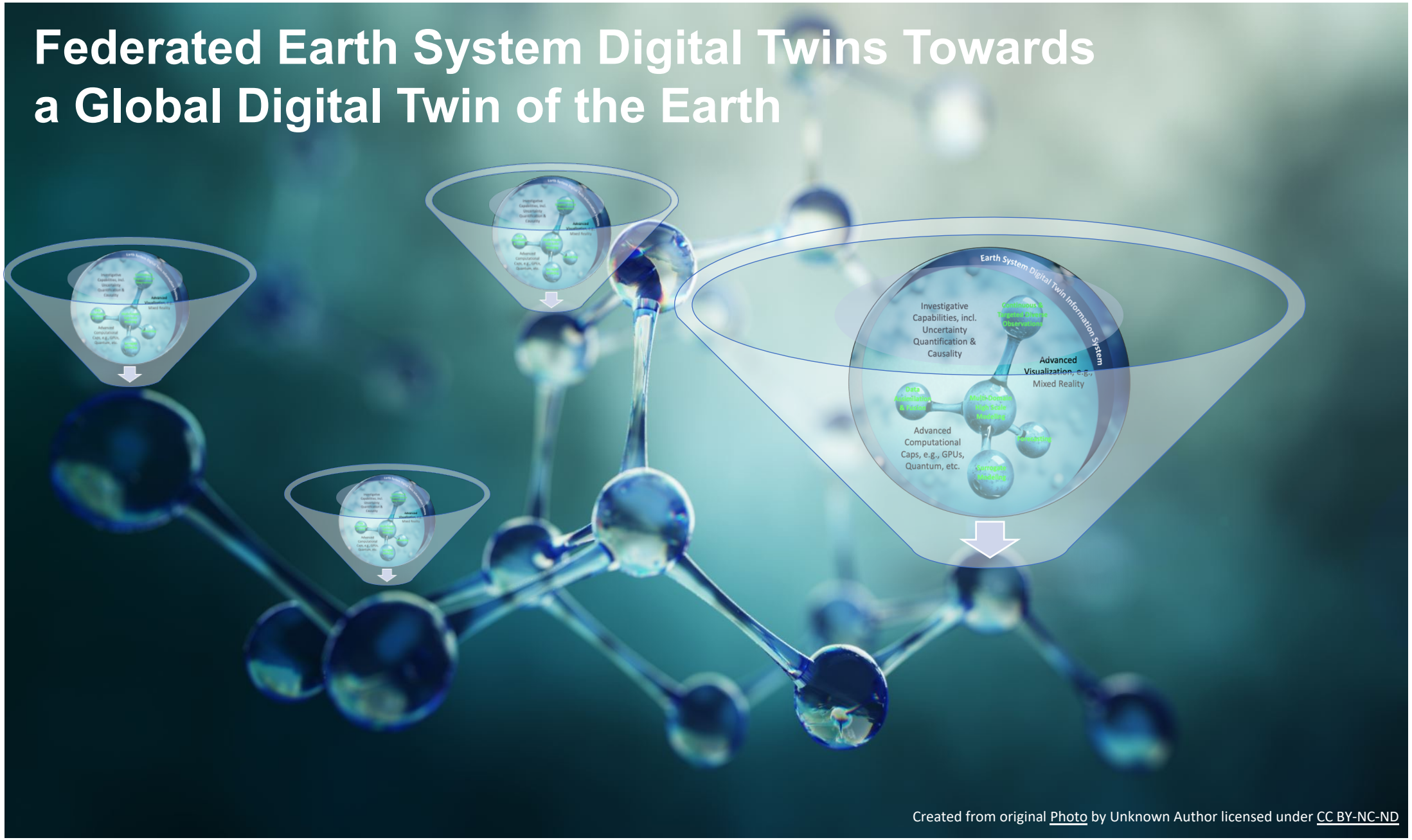
IDEAS – Digital Twin for Water Cycle and Flood Detection and Monitoring

Jet Propulsion Laboratory
California Institute of Technology

Goddard
SPACE FLIGHT CENTER

Langley
Research
Center

Federated Earth System Digital Twins Towards a Global Digital Twin of the Earth



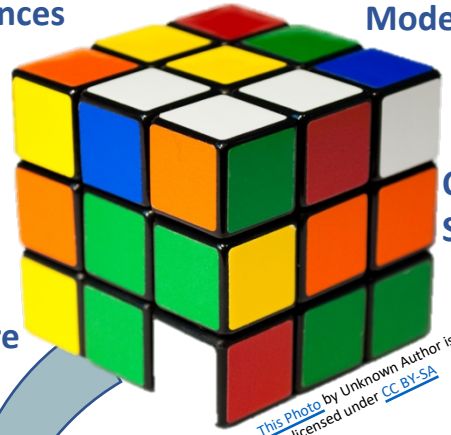
Earth Systems Digital Twins (ESDT) Technology for NASA Earth Science: Creating Proxies of the Earth – Past, Present and Future

Applied Sciences

Earth Science Models

Open Data Systems

Information Systems Infrastructure



High-End Computing

Actionable Information and Predictions (What If's)

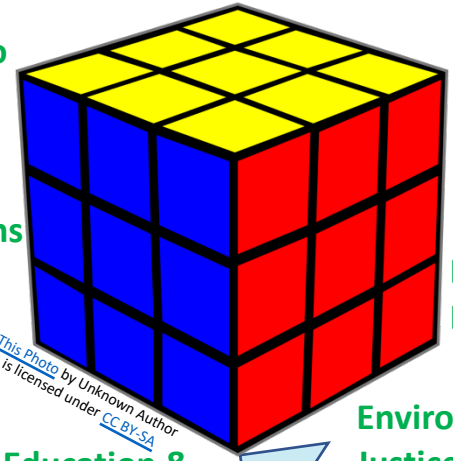
From Local to Regional to National to International Collaborations and Policy Decisions

Climate Change Mitigation

Environmental Intelligence

Environmental Justice

Education & Outreach



Connect Disparate Tools

Accessibility, Efficacy, Interoperability to Quickly and Efficiently Answer Questions

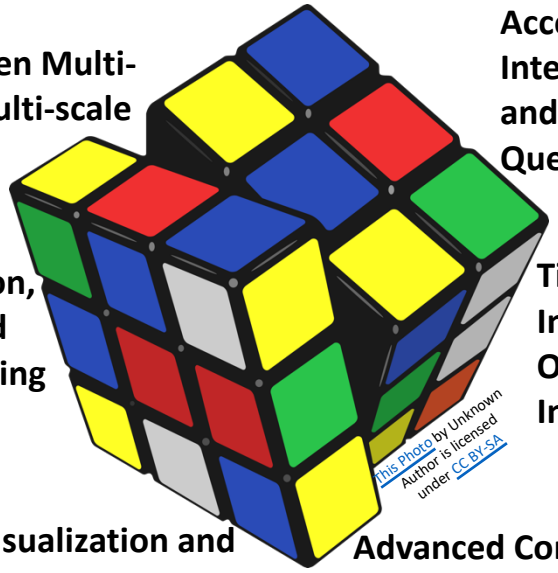
Timely, Near-Real-Time Ingestion of New Observations and Information

Advanced Computational Environments, e.g., GPUs, Quantum, DNA

Advanced Visualization and Mixed Reality

AI & ML for Fusion, Assimilation, and Statistical Modeling Surrogates

Bridge between Multi-discipline/ Multi-scale Models




Any Questions?