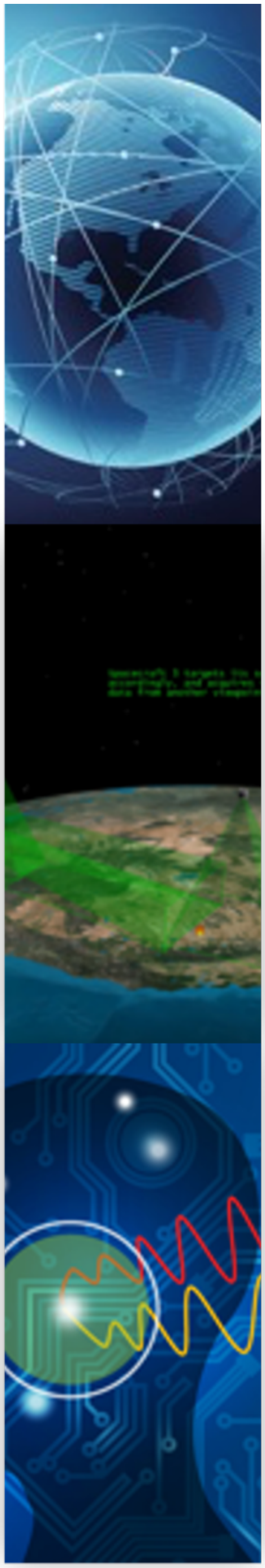


Advanced Information Systems Technology (AIST)

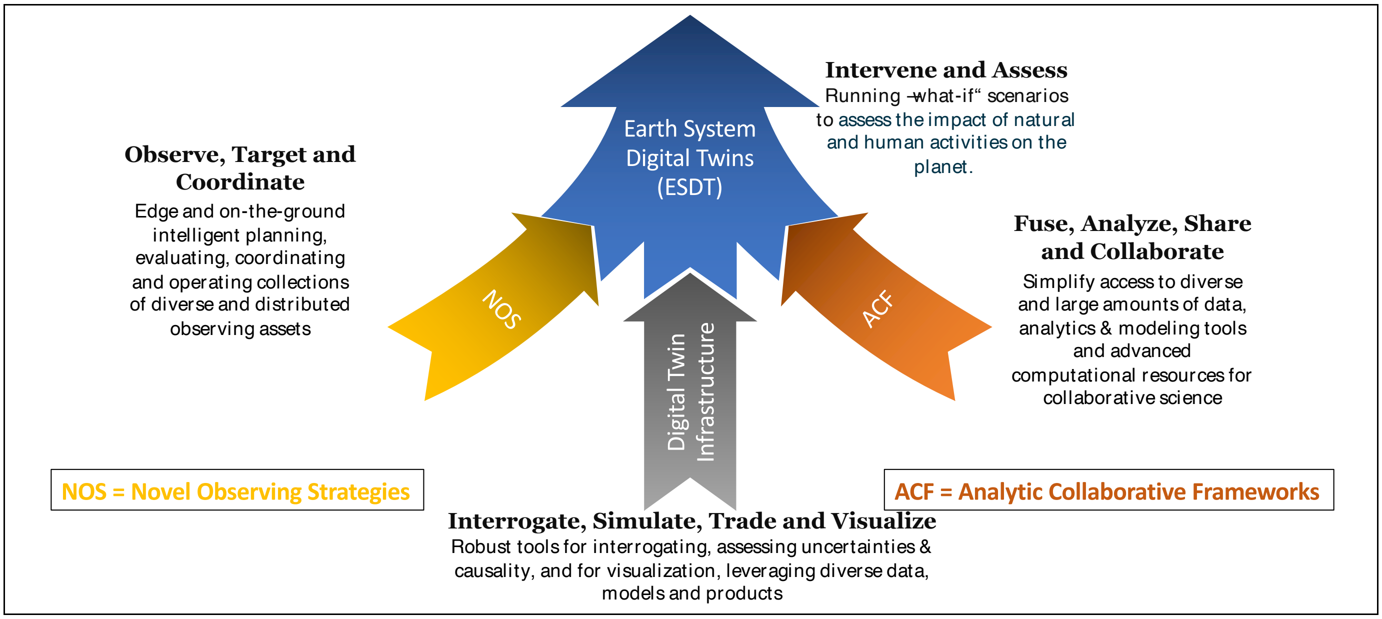
Earth System Digital Twin (ESDT) Architecture Framework

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

1.0 Background

The Advanced Information Systems Technology (AIST) Program identifies, develops, and supports adoption of software and information systems, as well as novel computer science technologies expected to be needed by the Earth Science Division in the 5-10-year timeframe. It matures these technologies through competed research projects. One of the major AIST thrusts focuses on digital twin technologies. Over the past four competitive cycles, maturation of technologies foundational to digital twins have proceeded. Earlier AIST solicitations focused on Analytic Collaborative Frameworks (ACF) and Novel Observing Strategies (NOS) themes. As the understanding of these capabilities evolved, it became clear that they feed into the broader concept of an Earth System Digital Twin (ESDT), as depicted in Figure 1.



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

Figure 1

ESDT stems from the unification of 3 AIST thrusts

AIST strategic goals for ESDT are to:

- Develop information system frameworks to provide continuous and accurate representations of Earth systems as they change over time.
- Mirror various Earth Science and anthropomorphic 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 as well as the impact to humanity.
- Provide the tools to conduct "what if" investigations that can result in actionable predictions.

Digital Twin technologies are one of the main subjects of technology development for AIST. This guide describes a framework for alternative architectures for ESDT. A framework provides a structure for developing and evaluating alternative architectures and their contributing technologies without prematurely committing to a particular solution. Some principles in the design of an architecture include modularity of the various building blocks, components, or subsystems. In addition, the architecture should consider what open-standard interfaces should be defined to enable unplanned use. There are a wide variety of digital twin architectures being developed both in the United States and abroad and for AIST technologies to have the broadest possible applicability they must be architecturally agnostic. However, a framework can provide a basis for describing where a technology fits into the overall scheme without narrowing its utility to one architecture or another. For example, a tool for visualizing data could be useful whether the architecture specified a database or filesystem as the Digital Replica. It also enhances the ability to coordinate with other offerors, who may want to construct tools which interoperate with other proposed capabilities. Before specific components can be developed, an architecture framework is needed which focuses the design decisions and defines the specific components and their inter-relationships.

2.0 Digital Twin Concept

An ESDT is a dynamic and interactive information system that:

1. provides a *digital replica* of the past and current states of the Earth or Earth system, as accurately and timely as possible.
2. allows for computing *forecasts* of future states under nominal assumptions and based on the current replica.
3. offers the capability to *investigate many hypothetical scenarios* under varying impact assumptions.

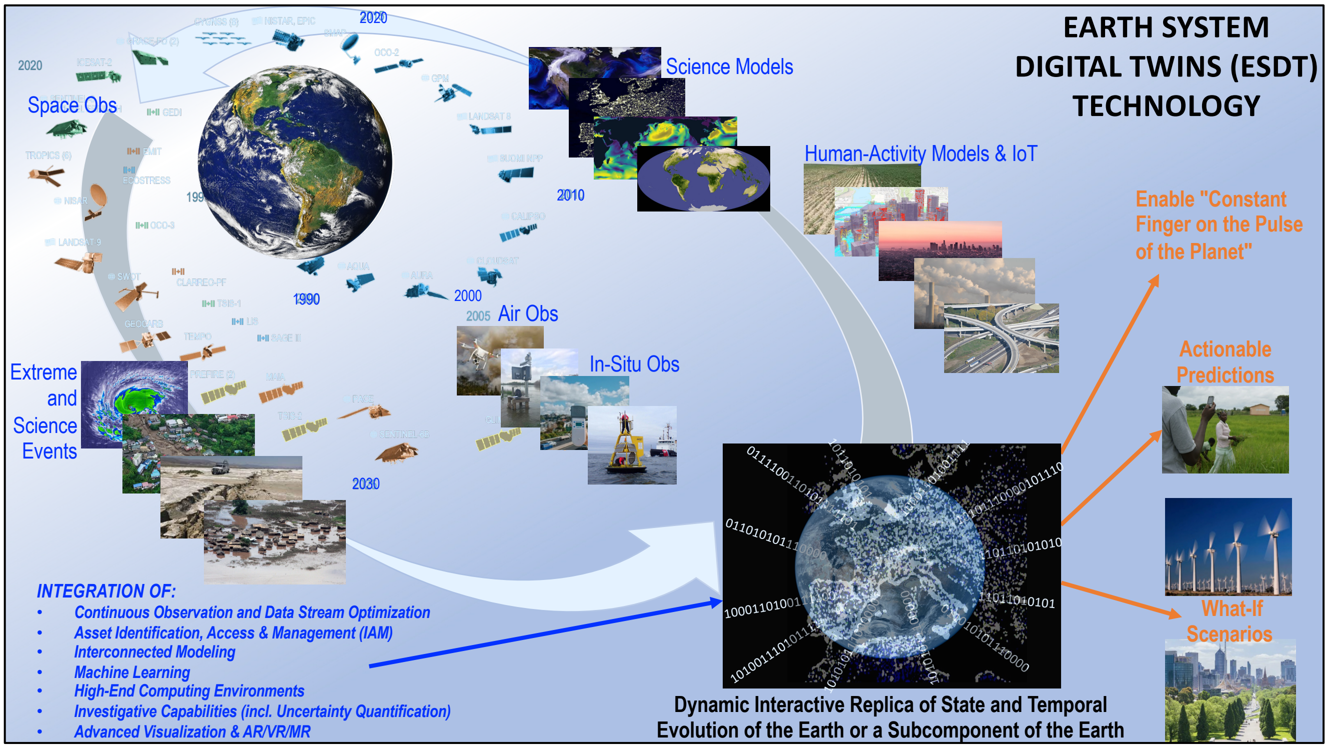


Figure 2

Earth System Digital Twin (ESDT) Technology Concept

Figure 2 provides an operational view of the 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, emulators, 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 series 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 from naturally occurring and/or human activities on physical and natural environments.

A given ESDT should support different types of user communities and their needs. Different functional users include:

- Policy makers
- Decision makers
- Earth scientists
- Public customers

An ESDT contains ACF-type functionality and also provides modeling capabilities but is far more than either. Figure 1 describes how the ESDT is a beneficiary of the previous work in each of the NOS and ACF threads. For example, some of AIST-21 projects (e.g., AIST-21-0003, AIST-21-0018, AIST-21-0031 and AIST-21-0091, just to name a few, and which are more fully described in the ESTO Technology Portfolio) couple state-of-the-art ML with NASA (and other) Earth observing data, and provide various forecasting, sensitivity analyses, and counterfactual "what if'' experiments.

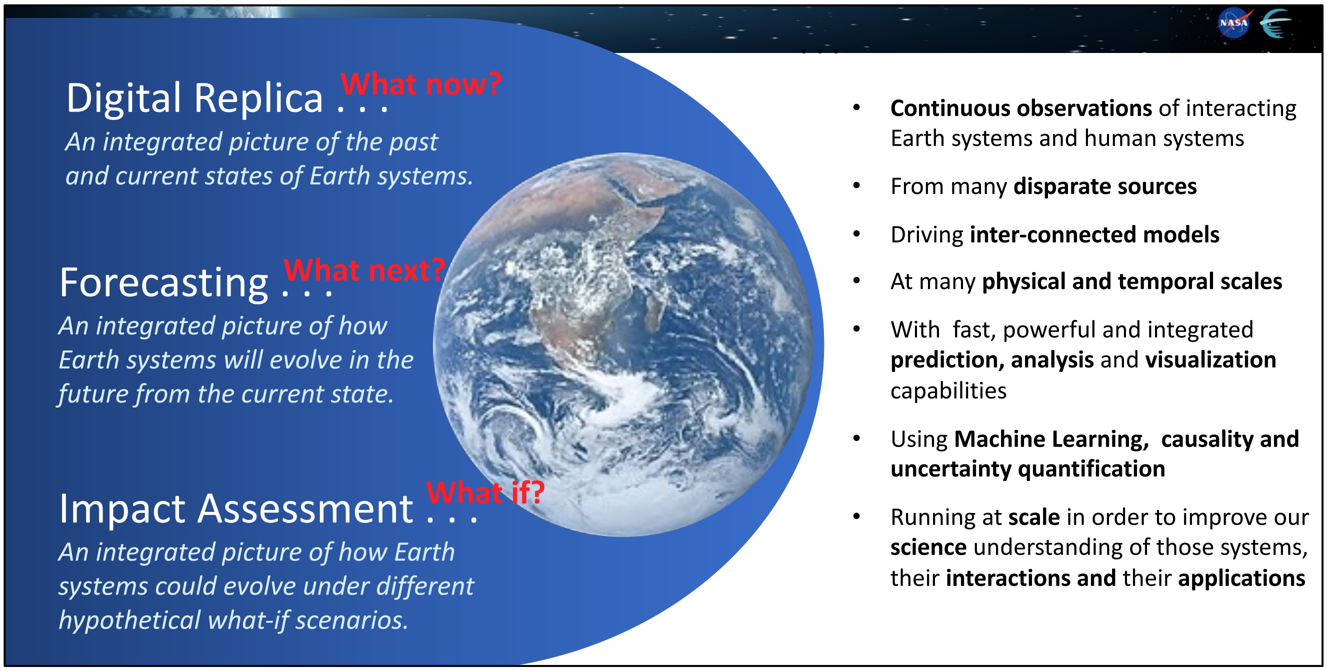
3.0 Functional Concept

Figure 3 defines the ESDT concept from a functional point of view.

Some **key features and capabilities** that are needed include:

- *Flexibility* to promote re-use
- *Modularity*, including what interfaces should be built in
- *"Provide the glue around" capabilities* provided by the other components of NASA's Earth Science Division (ESD), Research & Analysis (R&A), Applications and Earth Science Data Systems (ESDS)
- *User interfaces* that can be used with a range of entry skill levels and interests (i.e., usable "from Farmer to Scientist")
- *Ability to evolve* over time

- *Ability to ingest data and information from multiple sources*, not only NASA, but Other Government Agencies (OGAs), industry, academia, human and infrastructure orgs, various formats, structured and unstructured
- *Ability to incorporate new components* "without breaking the system"
- *Compliant with Open-source Science*, per NASA SPD-41A
- *Interoperability* with other Digital Twins of related phenomena
- *Modeling of coupled Earth systems across temporal scales* (hindcast, nowcast, near-term forecasts, long-range projections)
- Enable *analysis of Earth science processes across coupled systems, spatial scales, and temporal scales*
- *Intuitive* interfaces and analysis.



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**

Figure 3

Functional Definition of an ESDT

4.0 Digital Twin Components

Figure 4 describes the major notional components of the system and some key relationships, while providing room for a variety of architectures to respond to them. It provides a generic system diagram of an ESDT, shown in blue, including the interfaces to the external observing systems, shown in green.

4.1 Observational Data Repository (ODR)

These external systems, depicted in green in Figure 4, provide the satellite observations, model output and in situ measurements that characterize the set of physical phenomena and natural processes being studied by the digital twin. The DT needs all observations and measurements to be reliably available and constructed by science data processing (SDP) systems using algorithms and codes validated and approved by the instrument science teams. The DT needs these external sources to share uncertainty quantification (UQ) characteristics and appropriate metadata. Design decisions for the DT Architecture include not only how to load historic data into the Digital Replica and possibly into the Digital Record, but also how to collect updates to the data. Because the DT is possibly a near real time (NRT) system, the observational data is retrieved from the output of the SDP directly or as quickly as it is available in the archive. Consideration, when making design decisions for the DT Architecture, should be given to tradeoffs in parameter sources and how sensitive the needed DT capability is to latency, accuracy, validation, etc. Each instrument has different latency characteristics, which must be considered along with the UQ. For an archive to be usable, the data must be retrievable for at least a year to permit both time-series analysis and replication of results.

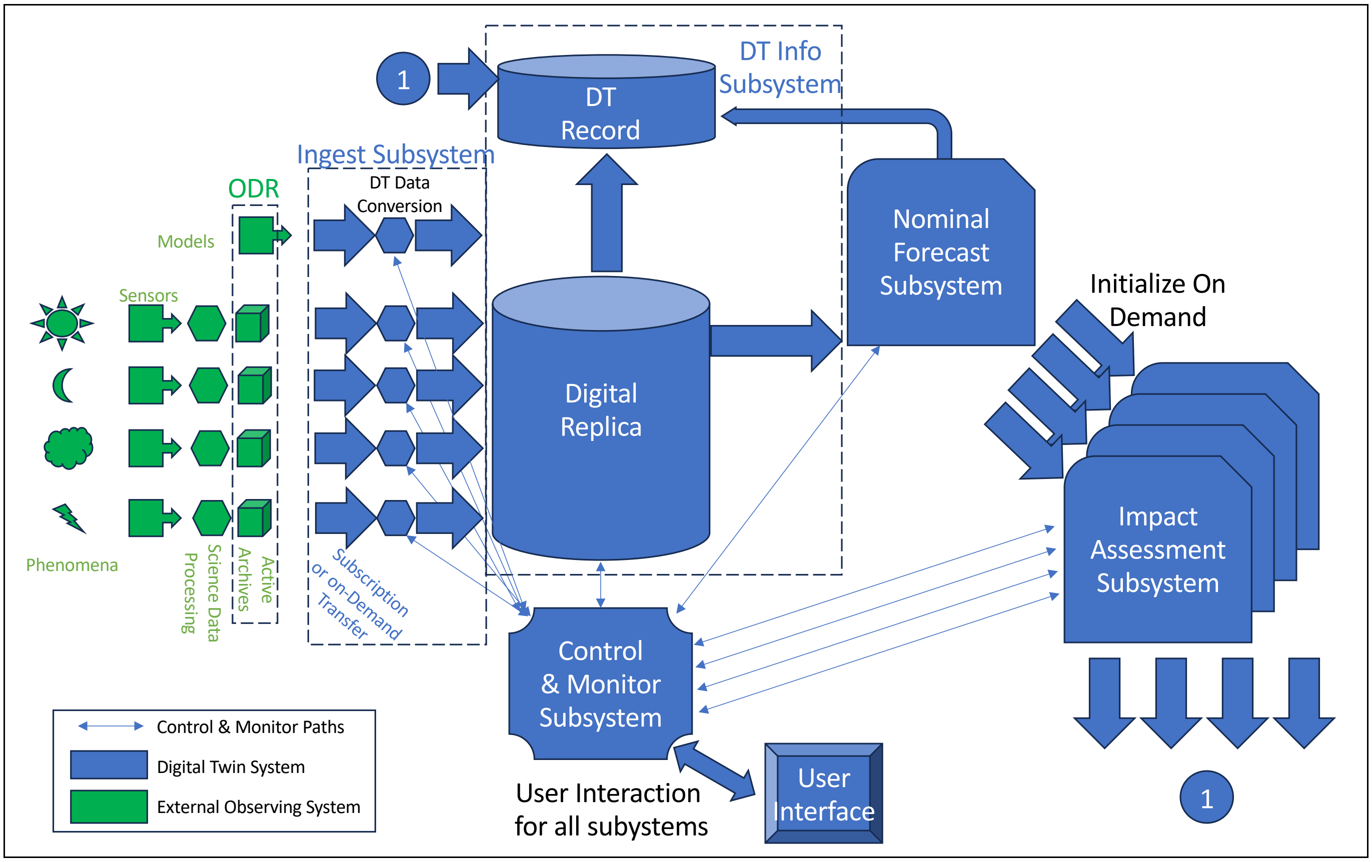


Figure 4
ESDT System Diagram

4.2 Ingest Subsystem (IS)

The Ingest Subsystem obtains data and model output from the external sources either through subscriptions or on demand, leaving the authoritative source as the original repository, such as the DAAC. The architecture design may collect both primary and alternate sources to provide redundancy or improved error correction on sources. The IS performs whatever restructuring of the data is needed to fit it onto the selected Digital Replica frame of reference. It performs quality assurance before ingesting and reports status to the Control and Monitor Subsystem (CMSS). Extract, transform, and load operations are some of the functions performed by the IS.

IS also includes diagnostic tools, accessed through the CMSS to troubleshoot any anomalies which invariably arise in observational systems. The IS should provide the means to attach subsequent analysis, links to published papers and other downstream results to the DT.

4.3 DT Information Subsystem (DISS)

The information storage system consists of three components, the Digital Replica (DR), the DT Record and separately, the external repositories for the source data, when they are provided and in an appropriate format.

The core computational component of a ESDT is the Digital Replica (DR). It is the unified representation of the current state of the subject Earth system. That representation reflects the observational data and is aligned to a common spatial and time reference system (e.g., common grid). Where warranted, the DR may interpolate observational data to provide a seamless view over time and space. For example, nowcasting, reconstruction, and/or interpolation to fill in observational gaps in space or time.

The DR includes not only the data and model output, but also the interpretation that transforms data into scientific description of the phenomena. It provides UQ and is traceable back to the original sources. The data and output contained therein has been pre-processed to be analysis ready for the visualization, analysis, and modeling tools in the DT and to supply data upon request with the lowest possible latency.

Architectural design decisions lie in data storage and localization to minimize latency; fusion of diverse data sources and model outputs; and managing computational complexity, including high-dimensionality and non-linearity. AI technologies may play a role in data fusion, interpolation, surrogate modeling, causal impacts, and interpretation of the data, among others.

The DT Record is the repository for the source materials, along with appropriate metadata, which are not otherwise reliably archived. The external ODRs continue to be the authoritative source of the data in original form. The DT Record also preserves the provenance of the steps in the DTs evolution. A design decision is needed to assess sources and their repositories to rely on them or preserve them in the DT Record. Further design decisions may require the DT Record to preserve intermediate results temporarily to support checkpointing or diagnostics. It may also hold data for which time-series are more cost-effectively stored in a format other than that supplied by the external archives. The DT Record retains any forecast or impact assessment data determined to be necessary by the DT owner. Log data is also be stored in the DT Record, as appropriate to reconstruct the state of the Digital Replica at any time in the past, characterize the data and code used for forecasts and for Impact Assessments. Careful consideration of the traceability and downstream reproducibility of experiments will influence the design of the DT Record.

The Digital Replica and the DT Record both have interfaces to the Control and Monitor Subsystem (CMSS) described in section 4.6. Careful consideration should be given to logging of key events to be fed into the DT Record for reproducibility, troubleshooting and audit purposes. User interaction occurs through the CMSS.

4.4 Nominal Forecast Subsystem (NFSS)

A digital twin produces nominal forecasts of the future state(s) of the phenomena and processes. These forecasts may take a number of forms, including near- and long-term forecasts, hindcasts and nowcasts, as appropriate for the science in the domain of interest. Design decisions must consider the computational load, the frequency at which meaningful updates to the Digital Record are supplied and how often the phenomena change requiring a new forecast. An architectural decision may rely on traditional physics-based models or use AI-based surrogate models, depending on available computational resources, limits on latency and accuracy, uncertainty quantification and confidence in the alternatives.

The Nominal Forecast Subsystem permits user interaction, including console, display, and APIs, for diagnostic, analysis, and visualization purposes through the CMSS. It also connects to the Control and Monitor Subsystem to provide status and to respond to prioritization in resource allocations.

4.5 Impact Assessment Subsystem (IASS)

When users need to investigate the impact of conditions of the phenomena and processes on other phenomena or humans, they can open a session on the Impact Assessment Subsystem through the CMSS. The development of a DT architecture is afforded considerable latitude, based on the user needs for this subsystem, including how much the DT provides and how much the user must supply. They would initialize their analysis with the current state in the Digital Replica, or some future state as predicted by the Nominal Forecast System and provide some provenance data to explain to others the key issues of the impact study. Surrogate AI modes should be considered to accelerate timing and also reduce the cost of computation. The IASS should support both individual studies of a single session or a campaign of multiple sessions with different settings. A design decision of the DT Architecture is what to store from the impact analysis; it could be everything including intermediate results to reduce the restart time for follow-on and to enable stepwise debugging/analysis. Or it could be to save only the final result. IASS should also permit multiple users to conduct analysis independently with resources assigned through the DT or by exporting the required initialization information for studies run on external platforms.

At the start of each session or campaign, the materials to be archived should be defined so that they can be stored in the DT Record, as appropriate. A design decision for a DT architecture would be to determine the flexibility for each internal user session in terms of data, tools and computational resources needed for analysis. This includes what AI surrogate models may be available for their use. The user may be permitted to supply their own models and tools and visualization assets to perform the analysis internally, each of which should be logged into the code repository. Interfaces to external github-type code bases should be included to avoid unnecessary duplication and configuration management problems. Users may choose to perform their analysis on separate systems and only export the initialization data.

Consideration of the business aspects of the IASS are important as costs could be high to run models and analysis and should be attributed to the end user's funding. There should be a clear delineation as to which costs are funded by the DT organization and which are funded by the end user. Security and privacy of the user session should be balanced with Open Science policies.

4.6 Control and Monitor Subsystem (CMSS)

The DT is an extremely complex system requiring a high level of internal coordination. While human owners could manually operate a simple version of the system, CMSS is a key element of automating the coordination so as to be efficient and to provide human owners, operators, and users with the information they need to make decisions about effective use. The CMSS acts to glue all the components together and to synchronize their operations. It is the master planning system and reports when observational data is not received on time that might affect processing, forecasting and impact assessments. It monitors system load, capacity and for system malfunctions and reschedules processing to optimize around a reduced capacity. It also provides all alerts and user interface notifications of the state of the digital twin's operations. The CMSS provides an automated prioritization of resource tasking with a user interface permitting override by the system owner. A key element of the design of the CMSS is the human interaction; design decisions must be made to the CMSS to be a semi-autonomous system with human supervision or an operator aid.

The CMSS also logs appropriate events and provides analysis and visualization tools for the system operator and owner to evaluate system performance and faults. The CMSS User Interface is the Code, tools, APIs, and auxiliary interfaces needed to assess the state of the Replica and the Record. As a minimum, it must support troubleshooting DT behavior, but it may also be used to analyze the data being used. Visualization of the contents of the Digital Replica, of the results of the NFSS and IASS as well as analysis should be supported.

When a user logs into the DT to perform impact analysis, CMSS User Interfaces offer the individual resources available at the time and to plan the interaction to assess their impact on system resources and provides notices of any outages that might impact their assessment. It then schedules the resources for their use and sets priorities.

4.7 User Interface

In AIST's concept, all user interaction occurs through a single user interface in order to provide unified user authentication and authorization. The user role and authority determine what controls and functions are available to the user, whether checking for data in the DR, identifying status of jobs, extracting a subset of the nominal forecast, or creating an impact assessment. Each subsystem design must consider user interaction with it, what functions it provides and through what API and how they are managed by the CMSS so as to avoid disruption of nominal processing.

5.0 Conclusion

Earth System Digital Twins have the potential to revolutionize the way Earth Science research will be conducted in the future, but also how results and knowledge from this research will provide information to support decision making and yield impactful societal benefits. While there will be a handful of global Digital Twins of the Earth providing essential global information, they will co-exist with a hierarchy or an ecosystem of many various smaller and/or thematic digital twins which will require cross-cutting technologies and will need to interoperate and to be federated. This framework is the first step towards supporting this ecosystem and enabling these various digital twins to be coordinated and collaborative.

References

- AIAA Digital Engineering Integration Committee, “Digital Twin: Reference Model, Realizations & Recommendations,” January 2023, https://www.aia-aerospace.org//wp-content/uploads/Digital-Twin-Implementation-Paper_Dec_2022.pdf.

- K. M. Alam and A. El Saddik, “C2PS: A Digital Twin Architecture Reference Model for the Cloud-based Cyber-Physical Systems,” *IEEE Access*, Vol. 5, pp. 2050–2062, DOI: 10.1109/ACCESS.2017.2657006, January 2017.

- P. Bauer, P. D. Dueben, T. Hoefler, T. Quintino, T. C. Schulthess and N. P. Wedi, “The Digital Revolution of Earth-system Science,” *Nature Computational Science*, Vol. 1, pp. 104–113, DOI: 10.2038/s43588-021-000023-0, February 2021.

- I. Özkaya, “Architectural Concerns of Digital Twins,” *IEEE Software*, Vol. 39, Issue 2, pp. 3–6, DOI 10.1109/MS.2021.3130872, February 2022.

- R. Berkheimer, S. Formel, J. Larocque, and R. Baldwin, “DTE-FS: A Model Based System Specification for Use in Construction of an Interoperable Digital Twin Earth Framework”, *NOAA Technical Memorandum* and *AGU Fall Meeting 2022 Presentation*, Chicago, IL, Dec. 12-16, 2022, id. IN42B-0330.

- P. LaPlante, “Trusting Digital Twins,” *Computer,* Vol. 55, pp. 73–77, DOI: 10.1109/MC.2022.3149448, July 2022.

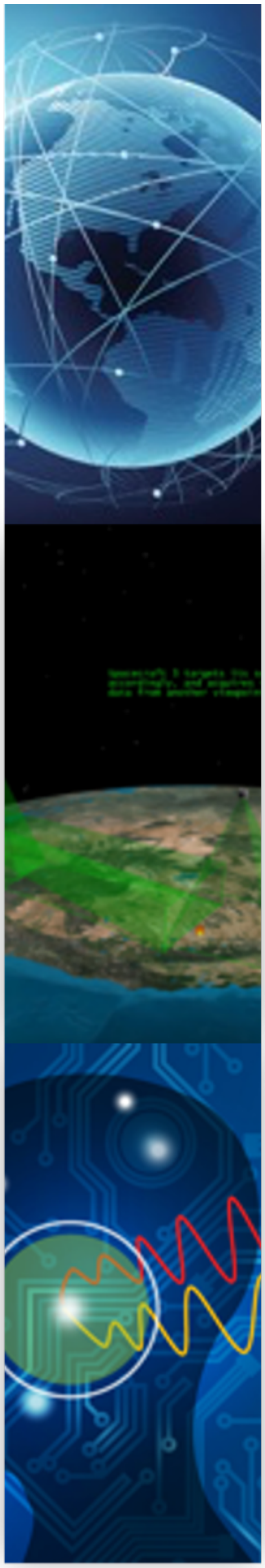
- J. Le Moigne and B. Smith, Eds, “Advanced Information Systems Technology (AIST) Earth System Digital Twins (ESDT) Workshop Report,” *ESIP-NASA AIST ESDT Workshop*, Washington DC, Oct. 26–28, 2022, https://esto.nasa.gov/files/ESDT_Workshop_Report.pdf.

- National Academy of Science, “Opportunities and Challenges for Digital Twins in Atmospheric and Climate Sciences”, *Proceedings of a Workshop in Brief*, Washington, DC, The National Academies Press. https://doi.org/10.17226/26921, February 2023.

- European Space Agency, “Working Towards a Digital Twin of Earth,” *2021 ESA EO Phi-Week*, October 2021, https://www.esa.int/Applications/Observing_the_Earth/Working_towards_a_Digital_Twin_of_Earth.

Acronyms

AIST	Advanced Information Systems Technology
CMSS	Control and Monitor Subsystem
ESDT	Earth System Digital Twin
ESTO	Earth Science Technology Office
DAAC	Digital Active Archive Center
DT Record	Digital Twin Record
DR	Digital Replica
DT	Digital Twin
DTIS	Digital Twin Information Subsystem
IASS	Impact Assessment Subsystem
IS	Ingest Subsystem
NASA	National Aeronautics and Space Administration
NFSS	Nominal Forecast Subsystem
NRT	Near Realtime
ODR	Observational Data Repository (external)
OISS	Observation Ingest Subsystem
SDP	Science Data Processing
UQ	Uncertainty Quantification



ESTO

Earth Science Technology Office