Federal Register Notice: 89 FR 51554, Federal Register :: Networking and Information Technology Research and Development Request for Information on Digital Twins Research and Development, June 18, 2024.

# Request for Information on the National Digital Twins R&D Strategic Plan

Brookhaven National Laboratory

# Brookhaven National Laboratory Response to: Networking and Information Technology Research and Development Request for Information on Digital Twins Research and Development

*This document is approved for public dissemination. The document contains no business-proprietary or confidential information. Document contents may be reused by the government in the National Digital Twins R&D Strategic Plan and associated documents without attribution.*

**Contributors:** Vanessa López-Marrero, Meifeng Lin, Mohammad Atif, Wei Xu, Yuewei Lin
**Date:** July 28, 2024
**Keywords:** Artificial Intelligence, Ecosystem

## 1 Introduction: Ecosystem – Mathematical, Statistical, and Computational Foundations

The success of modern digital twins is highly dependent on innovations in computational and applied mathematics, jointly with advancements in computer science, high-performance computing (HPC), machine learning (ML), and artificial intelligence (AI), that can enable simulation, prediction, inference, and decision-making for digital twins. While predictive or decision-making capabilities in digital twins may exist, this is currently an active area of research [1, 2, 3]. Hence, continuous data assimilation and model calibration, as well as predictive or decision-making capabilities in digital twins, are not necessarily mature or fully integrated with traditional simulation methods. Thus, at present, digital twins are not utilized at their full potential. Fundamental research needed to achieve the goal of practical development and adoption of such predictive digital twins includes [2, 4, 5]:

- Mathematical methods that allow for unified frameworks for probabilistic modeling, processing, and analysis of data from a wide range of data distributions,
- Scalable and portable high-fidelity simulation algorithms and software libraries,
- Efficient surrogate models to enable fast real-time simulations to support decision making,
- Methods that enable uncertainty quantification (UQ)
- Methods for validation & verification (VV) of software, algorithms, models, and predictions,
- Methods to support processing/fusion of data from diverse sources, e.g., sensor measurements, images, video, and outputs from experiments/simulations,
- Integrated workflows and frameworks to support seamless data ingestion, model calibration, control, and decision making,
- Visualization techniques and other tools to facilitate integration of and support for the human-in-the-loop component of digital twins.

Furthermore, mechanisms that inform a digital twin user (human expert or AI agent) of the expected level of accuracy of the digital twin's simulations or predictive answers, and analogous feedback from the user to the digital twin, are essential to realizing the goal of continuous coupling/integration between the digital and physical twins. Such an ecosystem is depicted in Figure 1. In the next sections we expand upon some of the aforementioned aspects and research needs.

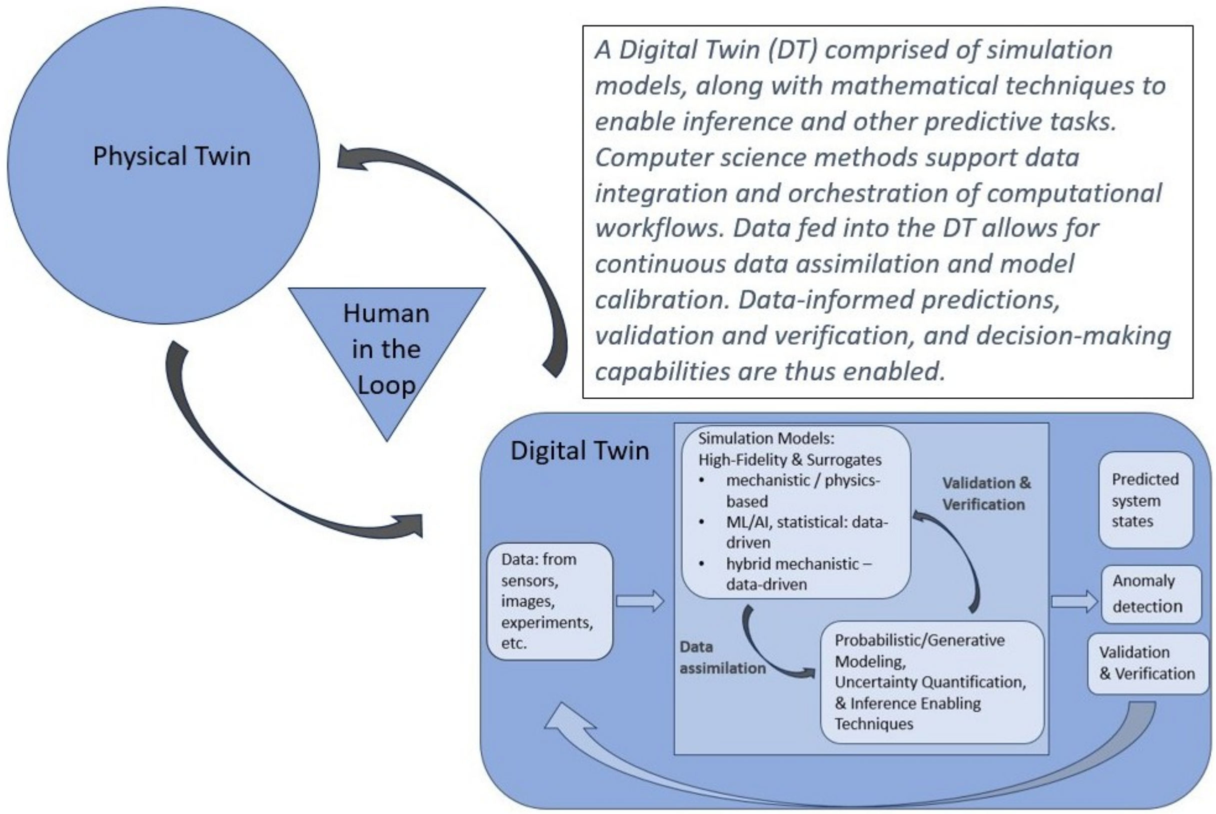


Figure 1: Exemplar digital twin ecosystem

## 2 Predictive Modeling and Simulation

As illustrated in Figure 1, methods from computational and applied mathematics are central to digital twins (DTs). The success of modern, predictive DTs is thus highly dependent on advancements in computational and applied mathematics that can enable simulation, data assimilation, inverse problem solution, prediction, inference, and decision-making for DTs [2, 4, 5]. Also essential is integration of such methods with computer science techniques for workflow orchestration and frameworks that support seamless processing of data from diverse sources.

A hierarchy of models for simulating phenomena associated with the physical twin is one core component of a DT's mathematical ecosystem. The collection of models may include both deterministic and stochastic models and should offer various degrees of fidelity and levels of required computing resources so as to allow for a range of needs, from real-time or interactive processing to execution of non-real-time simulation workflows. The models can be (a) mechanistic, such as those governed by (systems of) ordinary/differential equations or differential algebraic equations, and should include high-fidelity models and their surrogates of lower levels of fidelity, but which are also less computationally demanding, (b) purely data-driven, such as machine learning (ML), artificial intelligence (AI), or other statistical models, and (c) hybrid mechanistic – ML/AI/statistical models. Via judicious algorithmic choices, surrogate (mechanistic/ML/AI) models can be used together with (or in lieu of) high-fidelity models in order to reduce computing costs, while retaining acceptable levels of accuracy in the results [6, 7].

Probabilistic and statistical methods, modeling frameworks accounting for both additive (i.e., probabilistic) measures of uncertainty as well as for non-additive ones (i.e., imprecise probabilities), and other inference enabling techniques [8, 9, 10, 6, 11, 12, 1, 13, 14, 3, 15, 16] are another class of methods essential for predictive DTs. These include methods that allow for unified frameworks for generative/probabilistic modeling and analysis/assimilation of data from a wide range of distributions, Gaussian and non-Gaussian alike, such as those grounded on the theory and application of optimal measure transport. Equally important are data assimilation techniques as well as methods that enable uncertainty quantification (UQ) and validation & verification (VV) for the assessment of (a) models, (b) simulation of quantities of interest, and (c) predictions resulting from DTs. In conjunction with the aforementioned simulation capabilities, as well as with computer science methods for workflow coordination and data management, they provide core capabilities to enable predictive modeling, simulation, and decision making for DTs.

## 3 Long-Term Research Needs for Computing

To support digital twins' diverse use scenarios - from real-time monitoring and control, to offline design optimizations of the physical systems - there is a need to provide the computational readiness through the integration and application of modern advanced computing methodologies. The essential components of a responsive digital twin would require the scale-appropriate modeling and simulation approaches and fast, real-time, inference and prediction based on the data streams captured by the digital twin, all of which require considerable computational resources.

However, not all the computational requirements can be met through the large-scale processing at traditional data centers and supercomputing facilities, as the overhead of transferring the data into or out of these centralized facilities may be too large for certain digital twin applications. Edge computing devices and other novel computing architectures may be needed to support digital twins with low-latency requirements. Research and development is needed on the performance, programmability, and energy efficiency of new hardware architectures for compute, storage and data transferring fabrics to support digital twin deployments. In particular, user-friendly and portable programming models need to be developed to ensure the efficient utilization of the new architectures. Software tools, including performance analysis, monitoring and benchmarking, debugging frameworks and workflow middleware will need to be developed or adapted from existing ones to support the unique requirements of digital twin applications.

## 4 Sustainable and Extensible Software Ecosystem

The advancement and adoption of digital twins in sciences, engineering, and healthcare necessitates the development of a sustainable and extensible software framework. The framework should identify crosscutting creational design patterns and provide base classes for data storage, data refinement, system components, computational models, machine learning methods, and visualization. It should seamlessly combine the aforementioned components with workflow orchestration technologies for efficient execution on HPC machines as well as allow for the bidirectional interaction with a human-in-the-loop.

Thus, a unified software framework encompassing all software needs for a digital twin is an outstanding challenge. This framework must be modular, extensible, portable, and user-friendly to enable widespread adoption by the domain specialists in various applications. Currently, several such frameworks from the industry as well as research labs are undergoing development [17, 18]. The existing tools provide mechanism for asynchronous multi-host execution; however, they lack the interface to workflow scheduling, and execution tools on HPC machines [19, 20, 21]. Moreover, a successful digital twin must scale on the world-class extreme computing platforms which emphasizes performance and energy consideration of each individual component. Such a framework should be built around performance portable programming models [22, 23, 24, 25] to enable utilization of the HPC machines with minimal human effort for software redevelopment.

The digital twins rely on integrating heterogeneous computing components such as AI/ML models with HPC simulations. Currently, most scalable physics-based models are primarily developed in high-performance computing languages like C++. In contrast, the data-driven methods largely rely on Python frameworks such as PyTorch/TensorFlow/JAX. There is also an imperative need of integrating libraries that interface statistical models for uncertainty quantification [26, 27]. Digital twins require load-balancing algorithms to interact with high-performance computing resources, monitoring, logging, and recovery mechanism, and a lightweight 3D rendering for a real-time interactive visualization. The software landscape is complex which makes the integration quite challenging even if several excellent components exist. The adaptation of digital twins into different applications relies on a low entry barrier software framework. This framework should act as a portable and easy-to-use abstraction layer that unifies the various components of a digital twin.

## 5 Visualization

The development of digital twins (DTs) offers a viable opportunity to leverage human expertise and intuition to guide experimentation design and processes. From the perspective of visualization, such innovative DTs must render realistic physical systems and sensor data, support human interactions with the systems and capture that in real-time, so that generating intervention strategies can be autonomous or via human-in-the-loop. This revolutionary transformation requires orchestrating state-of-the-art visualization, AI, and HPC techniques.

There are three major visualization tasks for DTs [28]: 1) monitor and analyze the data from the physical assets to identify important events and trends in various visualization forms; 2) display simulations that are crucial to identify improvement opportunities and optimize the performance of physical assets and systems, allowing users to explore scenarios in real time, test alternatives, and design more efficient solutions; and 3) generate predictive maintenance alerts, highlighting potential issues before they become critical so that proactive actions can be taken to improve the reliability and efficiency of their equipment.

To realize the digital twin through visualization, there are three different techniques: 1) rendering a virtual 3-D representation (such as using NVIDIA's Omniverse [29]) and allowing users to interactively explore the virtual scenarios that might not be possible with physical systems; 2) situated visualization [30] using augmented reality to overlay 3-D renderings with computer-generated imagery directly in their physical locations; and 3) virtual reality supporting human-physical system interaction with gaze, head, and hand movements with haptic feedback.

The opportunities for enhancing current visualization techniques for DTs include several directions. First, efficient digitization techniques are required to provide high visual fidelity in appearance of the virtual environments with realistic human-system interactions. Even from limited images, it requires to photorealistically digitize complex physical environments. The challenge is to reserve the underlying physical principles or feedback realism of human-environment interaction, which can limit the applications for capturing natural and accurate human activities in scientific experiments. Second, it is essential to develop methods for effectively presenting the data to adapt seamlessly to various experimental conditions. A typical 3D scene visualization system is limited to pre-captured images. Once the experimental conditions vary, the regeneration triggers the need of full-blown experiments, image capture, 3D reconstruction, and rendering. It remains a challenge to bypass the process and result in a significantly faster and more interactive experience. Third, a new approach is needed to understand the rationale behind the decisions made by DT and increase assurance for humans to trust DT's findings [31]. This will facilitate humans-in-the-loop fashion enabled by adding a human cognitive dimension to DT's dynamic data-driven modelling and simulation, by involving human cognition at different simulation stages or in various formats. Finally, capturing human interaction with DTs will be vital for understanding and streamlining the experiment process, allowing us to identify key aspects to observe; improve information synthesis; and, ultimately, optimize decision-making. Currently, the information typically flows from the physical system to the users [32], but the feedback mechanism going the other way remains manual and tedious.

## 6 Artificial Intelligence (AI) Foundation Models

Foundational models (FMs) have revolutionized AI by providing a unified approach to handling different data modalities and performing a variety of tasks. In the field of digital twin research, FMs have great potential to enhance the development, accuracy, and usefulness of digital twins across multiple scientific fields. For example, in materials science, combining FMs with digital twins enables researchers to simulate and predict material behavior under different conditions, facilitating the design and testing of new materials with desirable properties. This approach can be equally applied to biomedical research, climate science, and nuclear physics, where accurate simulation and predictive capabilities are equally important.

The benefits of combining foundational models with digital twins include:

- **Strong predictive capabilities:** Foundational models trained on broad and diverse datasets demonstrate strong predictive capabilities. In the scientific field, this means more accurate simulations of complex systems under a variety of conditions, facilitating the creation of highly predictive digital twins. These enhanced predictions can inform design, testing, and optimization processes, reducing the need for expensive and time-consuming physical experiments.
- **Multimodal data integration:** Digital twins require the integration of data from a variety of sources, such as sensor data, experimental results, and theoretical models. Foundational models excel at handling multimodal data, allowing for seamless integration and interpretation of diverse data sets. This capability ensures that the digital twin is always updated with the latest data, maintaining its relevance and accuracy.
- **Accelerate scientific discovery:** Foundational models can greatly accelerate the discovery process by providing insights into the complex behavior and properties of the system being studied. By incorporating these insights into the digital twin, researchers can simulate and

predict how new designs or materials will perform under different conditions. This acceleration of discovery is critical to solving pressing challenges in areas such as energy storage, environmental sustainability, and advanced manufacturing.

## References


[1] R. Ward *et al.* Continuous calibration of a digital twin: Comparison of particle filter and bayesian calibration approaches. *Data-Centric Engineering*, 2021.

[2] National Academy of Engineering and National Academies of Sciences, Engineering, and Medicine. *Foundational Research Gaps and Future Directions for Digital Twins*. 2024.

[3] C. Agrell *et al.* Optimal sequential decision making with probabilistic digital twins. *SN Appl. Sci.*, 2023.

[4] K. Willcox and B. Segundo. The role of computational science in digital twins. *Nat Comput Sci*, 2024.

[5] H. Antil. Mathematical opportunities in digital twins (MATH-DT), 2024. arXiv.2402.10326.

[6] B. Peherstorfer *et al.* Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Rev*, 2018.

[7] D. Kochkov *et al.* Machine learning-accelerated computational fluid dynamics. *Proc. Natl. Acad. Sci. U.S.A.*, 2021.

[8] T. Hastie *et al.* *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Series in Statistics. Springer, 2009.

[9] F. Santambrogio. *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*. Birkhäuser, Cham, 2015.

[10] Y. Marzouk *et al.* Sampling via measure transport: An introduction. In *Handbook of Uncertainty Quantification*. Springer International Publishing, Cham, 2016.

[11] F. Cuzzolin. *The Geometry of Uncertainty*. Springer, 2021.

[12] G. Papamakarios *et al.* Normalizing flows for probabilistic modeling and inference. *J. Mach. Learn. Res.*, 2021.

[13] R. Baptista *et al.* On the representation and learning of monotone triangular transport maps. *Foundations of Computational Mathematics*, 2023.

[14] A. Spantini *et al.* Coupling techniques for nonlinear ensemble filtering. *SIAM Rev.*, 2022.

[15] R. Legin *et al.* Posterior sampling of the initial conditions of the universe from non-linear large scale structures using score-based generative models, 2023. arXiv.2304.03788.

[16] V. López-Marrero *et al.* Density estimation via measure transport: Outlook for applications in the biological sciences. *Stat. Anal. Data Min.*, 2024.

[17] twinflow. `https://github.com/aws-samples/twinflow`. Accessed: 2024-07-15.

[18] interTwin. `https://www.intertwin.eu/`. Accessed: 2024-07-15.

[19] Andre Merzky, Matteo Turilli, Mikhail Titov, Aymen Al-Saadi, and Shantenu Jha. Design and performance characterization of radical-pilot on leadership-class platforms. *IEEE Transactions on Parallel and Distributed Systems*, 33(4):818–829, 2022.

[20] QCG-PilotJob. `https://github.com/psnc-qcg/QCG-PilotJob`. Accessed: 2024-07-15.

[21] eFlows4HPC. https://github.com/eflows4hpc/hpcwaas-api. Accessed: 2024-07-15.

[22] Meifeng Lin, Zhihua Dong, Tianle Wang, Mohammad Atif, Meghna Battacharya, Kyle Knoepfel, Charles Leggett, Brett Viren, and Haiwang Yu. Portable programming model exploration for LArTPC simulation in a heterogeneous computing environment: OpenMP vs. SYCL. *arXiv preprint arXiv:2304.01841*, 2023.

[23] Mohammad Atif, Meghna Battacharya, Paolo Calafiura, Taylor Childers, Mark Dewing, Zhihua Dong, Oliver Gutsche, Salman Habib, Kyle Knoepfel, Matti Kortelainen, et al. Evaluating portable parallelization strategies for heterogeneous architectures in high energy physics. *arXiv preprint arXiv:2306.15869*, 2023.

[24] Tianle Wang, Mohammad Atif, Zhihua Dong, Charles Leggett, and Meifeng Lin. A new portable random number generator wrapper library. In *EPJ Web of Conferences*, volume 295, page 11001. EDP Sciences, 2024.

[25] Mohammad Atif, Zhihua Dong, Charles Leggett, Meifeng Lin, and Vakhtang Tsulaia. Porting ATLAS fast calorimeter simulation to GPUs with performance portable programming models. In *EPJ Web of Conferences*, volume 295, page 11018. EDP Sciences, 2024.

[26] Robin A Richardson, David W Wright, Wouter Edeling, Vytautas Jancauskas, Jalal Lakhlili, and Peter V Coveney. EasyVVUQ: a library for verification, validation and uncertainty quantification in high performance computing. *Journal of open research software*, 8(1), 2020.

[27] Brian M Adams, William J Bohnhoff, Keith R Dalbey, Mohamed S Ebeida, John P Eddy, Michael S Eldred, Russell W Hooper, Patricia D Hough, Kenneth T Hu, John D Jakeman, et al. Dakota, a multilevel parallel object-oriented framework for design optimization, parameter estimation, uncertainty quantification, and sensitivity analysis: version 6.13 user's manual. Technical report, Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2020.

[28] João L. D. Comba, Nicolau O. Santos, Jonathan C. Rivera, Regis K. Romeu, and Mara Abel. Data visualization for digital twins. *Computing in Science & Engineering*, 25(2):58–63, 2023.

[29] NVIDIA. Nvidia omniverse. https://www.nvidia.com/en-us/omniverse/.

[30] Sungbok Shin, Andrea Batch, Peter William Scott Butcher, Panagiotis D. Ritsos, and Niklas Elmqvist. The reality of the situation: A survey of situated analytics. *IEEE Transactions on Visualization and Computer Graphics*, 30(8):5147–5164, 2024.

[31] Parisa Niloofar, Sanja Lazarova-Molnar, Femi Omitaomu, Haowen Xu, and Xueping Li. A general framework for human-in-the-loop cognitive digital twins. In *Proceedings of the Winter Simulation Conference*, WSC '23, page 3202–3213. IEEE Press, 2024.

[32] Hangbin Zheng, Tianyuan Liu, Jiayu Liu, and Jinsong Bao. Visual analytics for digital twins: a conceptual framework and case study. *Journal of Intelligent Manufacturing*, pages 1–16, 2023.