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

Anonymous NAS DT

RFI Response: Digital Twins R&D Plan

We regard Digital Twins to correspond to a virtual environment that encodes an end-to-end physical experiment, incorporating theory/modeling of materials, processes and controls involved in the experiment. Digital twins can constitute a transformative concept in scientific research, offering a powerful approach to understand, model, and optimize materials and processes. They also provide a framework to enhance collaboration and knowledge sharing in the scientific community among different sub-disciplines. Digital twins enable researchers to bridge the gap between the physical and digital worlds, offering numerous benefits such as: (1) providing an in-depth understanding of materials at various scales. When coupled with the growing availably of high-fidelity experimental data, they can be used to create a dynamic, digital representation of materials, allowing for an in-depth analysis of their characteristics; (2) facilitating predictive modeling as they allow researchers to simulate how materials behave under different conditions, such as varying chemical composition, structure, as well as environmental effects like temperature, pressure, or stress; (3) enabling the optimization of materials and processes, by iteratively refining the virtual model based on real-world data and simulations.

Taking large scientific user facilities (SUFs) as an example, experiments can be time-consuming and require carefully planned sequences of steps, including formulating the scientific hypothesis, performing the experiment(s), modeling the results, and theory-experiment matching with data analytics to draw conclusions. Thus, a key challenge is to optimize experimental planning at both the individual and facility levels to reduce time to discovery of scientific knowledge, minimize redundancy, and maximize physics knowledge from each experiment. Examples of planning and optimization challenges include determining optimal synthesis conditions for a new material; selecting the correct set of multimodal experiments to solve a structural inverse problem; optimizing the parameters of specific tools to achieve computational and experimental end goals and generating precise, continuously calibrated models of accelerators for data analysis and interpretation. Key to achieving this objective is the need for a digital twin of each SUF that enables users to design, operate, and optimize experiments in a safe, virtual environment so they can seamlessly transition to the real facility, reducing the time to scientific discovery.

R&D topic areas in which the strategic plan for Digital Twins should focus with respect to SUFs are: *Long Term* investments, *Artificial Intelligence (AI)*, *Sustainability*, *VVUQ*, *Trustworthy*, *Ecosystem*.

Artificial Intelligence (AI): Modern foundational AI architectures are capable for very complex tasks, including the ability to generate new data, but require tuning billions of learnable parameters. Digital Twins provide an efficient platform to train such large foundational AI models that can be deployed in the physical world. End-to-end solutions can be enabled when Digital Twins of physical experiments are coupled with different theoretical and simulation approaches. As an example, in the nanosciences, it is possible to automate nanoscale experiments, such as a scanning tunneling microscope (STM), to perform routine measurements and manipulation. While the human tasks are automated, the speed with which a physical STM can create new complex structures hosting a completely quantum states is still going to be too slow in the real-world. But training such AI models on digital twin data of an STM, comprising an interconnected set of coupled multi-fidelity and multi-scale materials and process modeling, with controls that mimic the input/output to a physical STM chamber, can be extremely efficient. Generative AI models trained on big data coming from such digital twins of the STM can then be deployed on the edge in automated physical STMs to accelerate discovery of new quantum structures and perform measurements confirming the existence of exotic quantum states. In this scenario, few physical measurements come into play mainly to validate and improve the components of the digital twin in an iterative model, and finally to confirm the predictions made by trustworthy generative AI models trained on the digital twin datasets.

VVUQ (relevant for trustworthiness): Digital twins of complex experiments, especially incorporating new materials and new processes is a challenging task. Since discovery always occurs at the fringes of human understanding of matter and energy, theoretical models in digital twins need to be validated and corrected, and uncertainty in predictive models need to be quantified using physical measurements so to increase our trustworthiness of the digital twins. An iterative active learning loop is required that is general purpose to refine multi-fidelity and multi-scale theoretical models using the least possible number of multi-modal physical measurements, to make digital twins trustworthy. In the concrete example of developing digital twins for a scanning tunneling microscope (STM), an active iterative learning loop that validates, verifies the models composing the digital twins, estimates robust uncertainties on outcomes in real physical experiments for new input parameters and incorporate these uncertainties to further refine the digital twins and AI models is a necessity to efficiently utilize the least number of physical experiments to perform new scientific discoveries and breakthroughs.

Long Term (relevant for Ecosystem): Creation and refinement of digital twins is a continuous process, that involves connecting different types of digital twins across different scientific facilities and over different generations of experiments. For nanoscience experiments, digital twins will be deployed on the edge so that it can be actively used by experimentalists to test and validate their ideas on-the-fly, while the digital twin itself learns and improves from the physical measurements. Digital twins will have several layers or components that can be independently validated and verified by experiments across different DOE user-facilities, so that over time the composite digital twin model for specific planned experiments can be actively configured from these improved model components and tested in real-time by an experimenter. The 4 V's of data-science equally hold for digital twins, due to the need for large *V*olumes of experimental datasets, different *V*arieties of data to train the digital twins and subsequent AI models, the *V*eracity of the components composing the digital wins as well as the physical validation and verification data coming from different locations and experimentalists, and *V*elocity of data of data to actively fine tune digital twin models. This requires sustained long-term planning of resources, storage and high bandwidth national network communication, with scientific digital twin end-stations in each facility that can crosstalk continuously based on the latest information. These end-stations can be part of a larger digital twin ecosystem connecting national facilities across different scientific domains and principles.