A Vision for 2050: A Seamless Integrated Science Complex

Lavanya Ramakrishnan, Daniel Gunter, Shreyas Cholia, Deborah Agarwal
Data Science and Technology Department, Lawrence Berkeley National Laboratory
Contact Email: LRamakrishnan@lbl.gov

As Yogi Berra said, "It's tough to make predictions, especially about the future". However, the way science is done will move beyond the social and technology barriers that exist today. Science will fundamentally be multi-disciplinary. A user will be able to access science user facilities, data facilities, supercomputers, networks, data and software in a seamlessly integrated environment from their desks. A fundamental paradigm shift from today will be the existence of data facilities that provide long-term, well-curated data repository services and inter-disciplinary data analytics services. User facilities will be more automated than we can imagine such that the line between real-world synthesis and experiment through automation and simulated/modeled experiment will be blurred. Similarly, the distinction between data and model/simulation will disappear and there will be a continuum where models are used to inform data and data is used to inform/create models.

A trained scientist will be able to find data and workflows akin to how we find information on the internet today. A scientist will be to find data, publications and the software all linked together. A scientist will be able to read and interpret the data analytics process and reproduce the analytics used by another researcher as easily as they can learn a new skill on the internet today. A user will be able to contribute data to data facilities using existing standards and tools that will check for data compliance and quality and automatically tag the data with appropriate metadata. Artificial intelligence will be mainstream aiding users write their workflows and find data, enable data fusion across datasets from different domains and guide users to find correlations, helping staff at facility managing instruments and machines.

Many earlier ASCR reports capture the R&D innovations that will be possible in the next decade that will help us realize the vision. Rapid technical advances in hardware, software will be accompanied by innovations in workflow, data, and resource management. While automation will be an important theme, the role of humans in the systems (when, where, how, what) will be an important research area. Science and ASCR facilities will provide more building blocks, many of the tasks that are manual today will be automated e.g., data curation, QA/QC. Every file will be a knowledge base that will include metadata, links to papers and workflow scripts.

An often-neglected discussion when talking about future advances is the role of humans and their interaction with the technologies and artifacts. The nature of scientific collaborations and relationships will change significantly. For example, ASCR HPC facilities will be serving a large number of users who will not have direct access to the machines (akin to how we use the network today) and thus user support will be across facilities and projects. Education and career paths will look different; teams will evolve to be even more multi-disciplinary; data citations will be valued in promotions. It will not be sufficient to just support technological advances to realize the vision for the future. Adoption will require shifts in the culture. Thus, it will be necessary to develop a deep understanding of the social aspects of science collaborations using ethnography and other user research methods that are fairly commonly used in industry in the design and launch of innovative products.