Accelerating Scientific Knowledge Discovery (ASKD) Program

Deb Agarwal, Amber Boehnlein, Rich Carlson, Michael Ernst, Ian Foster, Barbara Jennings, Scott Klasky, Kerstin KleeseVanDam, Ruth Pordes, David Skinner, May 21st 2013

Vision:

The ASKD program will significantly shorten the time needed to transform scientific data into actionable knowledge by enabling the dynamic creation of advanced discovery ecosystems.

The ASCR Accelerating Scientific Knowledge Discovery (ASKD) program will conduct the research needed to create the dynamic real-time and near-real time knowledge discovery ecosystem elements needed to drive future scientific discoveries. These elements will enhance the value of existing DOE science facilities, contribute to the creation of a new generation of user facilities, and accelerate the use of extreme-scale computational facilities. Representative research topics include: advanced science processes (mechanisms that accelerate the extraction of knowledge from raw information), knowledge management in the discovery process (methods for the distributed creation, synthesis and management of knowledge), execution environments (accelerated knowledge creation, new data storage and access mechanisms, new programming paradigms for knowledge manipulation), hybrid theoretical and experimental models for designing and optimizing discovery environments (multi-dimensional combinations of predicted and measured behaviors of the dynamic, real-time systems that can be described, compared, validated and optimized), and large-scale real-time distributed systems (exposed by increasing the dynamic, responsive, and timeliness capabilities of the exascale space). These actions will allow DOE scientists to harness the collective intellectual prowess of their peers and technological capabilities of the DOE laboratory complex to energize the discovery process for the 21st century and maintain US leadership for the future.

Objectives:

We will achieve this vision and accelerate research, development and innovation within an identified scientific community by:

- Creating new methods for rapid and scalable collaborative analysis, interpretation, modeling, simulation, and prediction of complex phenomena relevant to DOE, to create new sharable insights;
- Constructing a cyber framework that supports increasingly complex real-time analysis and knowledge navigation, integration, and creation processes that ensures timely sharing of knowledge;
- Developing a high-speed, secure, and easily traversed web of scientific user facilities, computational facilities, and data, information, and knowledge resources;
- Demonstrating broad applicability within diverse scientific communities that span DOE and academic R&D institutions, educational partners, and industry;
- Developing measures of productivity and effectiveness within scientific communities to assess the impact of increased collaboration;
- Developing the programmatic structure needed to support an appropriate mix of short (3 year), medium (6 year), and long (12 year) term research activities; and
- Establishing mechanisms for funding the development of responsive sustainable knowledge discovery efforts in an expanding number of scientific communities over time.

Background:

The winners of the increasingly global and aggressive 21st Century race for scientific discovery will be those who are more agile in formulating and testing new hypothesis through physical experiments, computational simulation, and data analysis to gain new insights. DOE manages the world's largest collection of unique scientific user facilities[1] for open science research. These facilities continue to

[1] http://science.energy.gov/~/media/_/pdf/user-facilities/Office_of_Science_User_Facilities_FY_2012_rev1.pdf

expand in capability to meet the needs of researchers tackling the world's most pressing scientific challenges. Scientists using these facilities need new adaptive mechanisms that appropriately balance automation with expert intervention. Only then will they be able to extract trustable knowledge and discovery in a timely manner against a background of growing data diversity, rates, and volumes. Full understanding of objects or processes can often only be gained through a combination of different scientific techniques. Scientists therefore require new methods to allow the formulation and execution of discovery processes across facilities. Equally important are methods for establishing a persistent synthesis of knowledge that spans facilities and disciplines—and that permit large, multidisciplinary teams located in disparate facilities to work together effectively and efficiently. ASCR's natural role in this domain is to research and develop new theories, mechanisms, tools, and services to accelerate the scientific discovery process.

Questions Attacked:

Scientific discovery requires environments that foster the integration, navigation, and creation of knowledge across geographically distributed collections of researchers, scientific facilities, and data and knowledge repositories. The creation of such environments is a core challenge facing all DOE science communities, that requires fundamental computer science research to answers to questions like:

- How can we co-design tools and technologies in order to accelerate evidence-based knowledge navigation, production, and dissemination for exascale science?
- How do we enable scientists to work efficiently with large-volume, potentially streaming data sources?
- How can we deliver immediate and reactive knowledge discovery and creation processes for research teams that span multiple locations?
- How can we develop diagnostic and predictive models to understand the time spent in the science process, and how can we reduce the bottlenecks observed in our predictions?
- How can we understand potential trade-offs when configuring and operating complex reactive real-time solution environments for scientific knowledge discovery?
- How can we rebalance the human, machine and instrument interactions that enable scientists to make the leap from data to insight to knowledge faster?
- How can we accelerate the assembly of science discovery infrastructures that effectively and efficiently combine scientific instruments, data resources, and computational resources to answer new questions at the forefront of science?

Keys to Success:

To achieve this vision the computer science research community is articulating the major challenges that inhibit or otherwise impose barriers to scientific knowledge discovery. These challenges will drive a program to bring to fruition new methods, services, and scientific processes that enhance and accelerate the scientist's ability to rapidly create and extract knowledge in a vibrant collaborative environment An important element of this research program will be the definition of metrics to measure success and the continuous evaluation of progress against these metrics. The cycle of research into new techniques followed by their adoption into active scientific programs will ensure that new challenges are continuously identified as more knowledge is gained.

Successful realization of the ASKD vision will provide the foundation for the more effective use of facilities and resources across the DOE science complex, from leadership-class exascale systems, through multi-user facilities and local community facilities, to individual investigator laptops and smart devices. Long-term sustainability will come from integrating hardened research results into the experimental facility infrastructure and scientific operating procedures, which will in turn stimulate new challenges and research for future continuing innovation.