U.S. DEPARTMENT OF ENERGY | Office of Science

Scientific Collaborations for Extreme-Scale Science Workshop

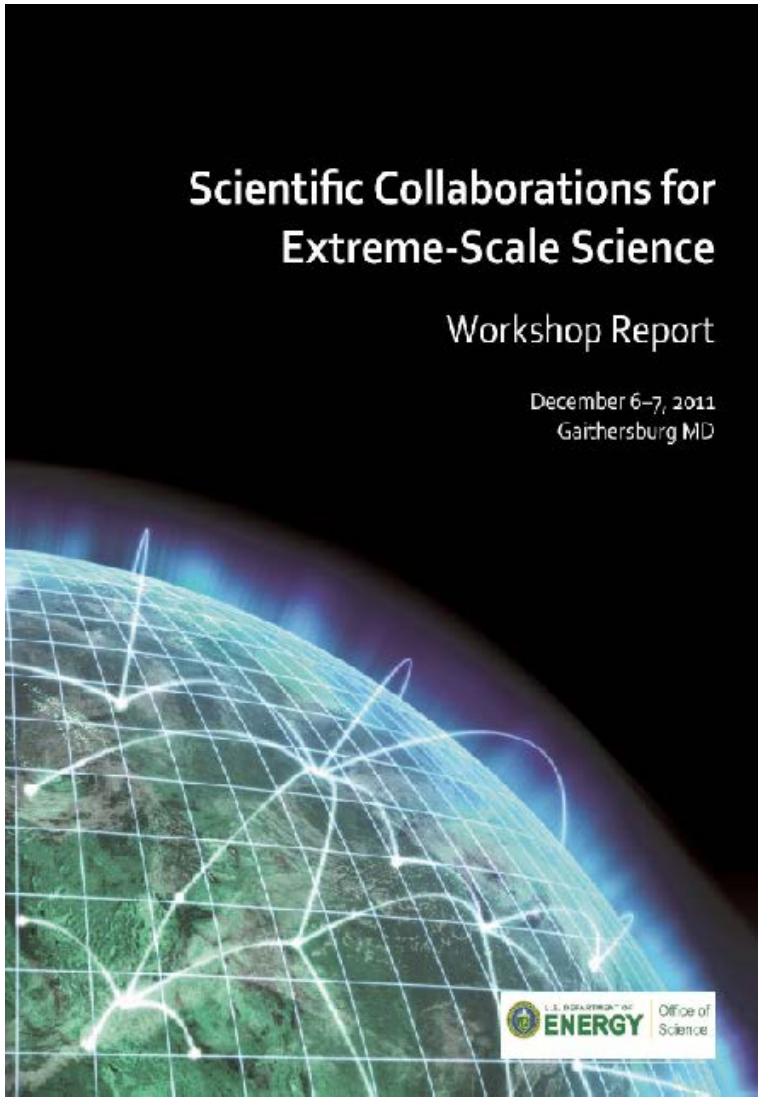
ASCAC

March 27, 2012

Richard Carlson

Richard.Carlson@science.doe.gov

Workshop Charge



The Next-Generation Networks for Science (NGNS) program in the Office of Advanced Scientific Computing Research (ASCR) in the Office of Science at DOE requests that the network and the distributed systems communities hold a workshop with the main objective of identifying the grand challenges issues inhibiting the development, deployment, and operation of scalable and secure extreme-scale scientific collaborative environments. Extreme-scale scientific collaboration systems would enable new modalities of scientific discovery involving multi-disciplinary, multi-facility, and/or distributed research teams in the next decade in the Office of Science. A committee should be appointed and tasked with preparing a first draft outline of future scientific collaboration systems as basis for discussion during the workshop.

Co-Chairs:

Richard Mount (SLAC)

David Skinner (LBNL)

Workshop Report

- **ASCR Web site**
  - http://science.energy.gov/ascr/research/next-generation-networking/
  - http://science.energy.gov/ascr/news-and-resources/program-documents/

- **Workshop web site**
  - https://indico.bnl.gov/conferenceDisplay.py?confId=403

Workshop Objectives

- **Identify the critical science drivers/challenges/opportunities for extreme-scale and data-intensive scientific collaboration in the DOE Office of Science**
- **Identify the technical challenges encountered in deploying and operating these collaborative technologies.**
- **Leverage the state-of-art Internet collaboration technologies and lessons learned from previous ASCR collaboratory programs (e.g., DOE National Collaboratories, SciDAC, DOE2000) to specify the needs and capabilities of future scientific collaborations**
- **Carry out a synthesis of the exascale series workshop reports to ensure that scientific collaboration requirements for exascale science, data-intensive science, SciDAC-III, co-Design, and terabit networks are integrated into future scientific collaboration capabilities**
- **Produce a well-documented workshop report that articulates the research opportunities and challenges facing scientific collaborations in the era of extreme-scale data-intensive science.**

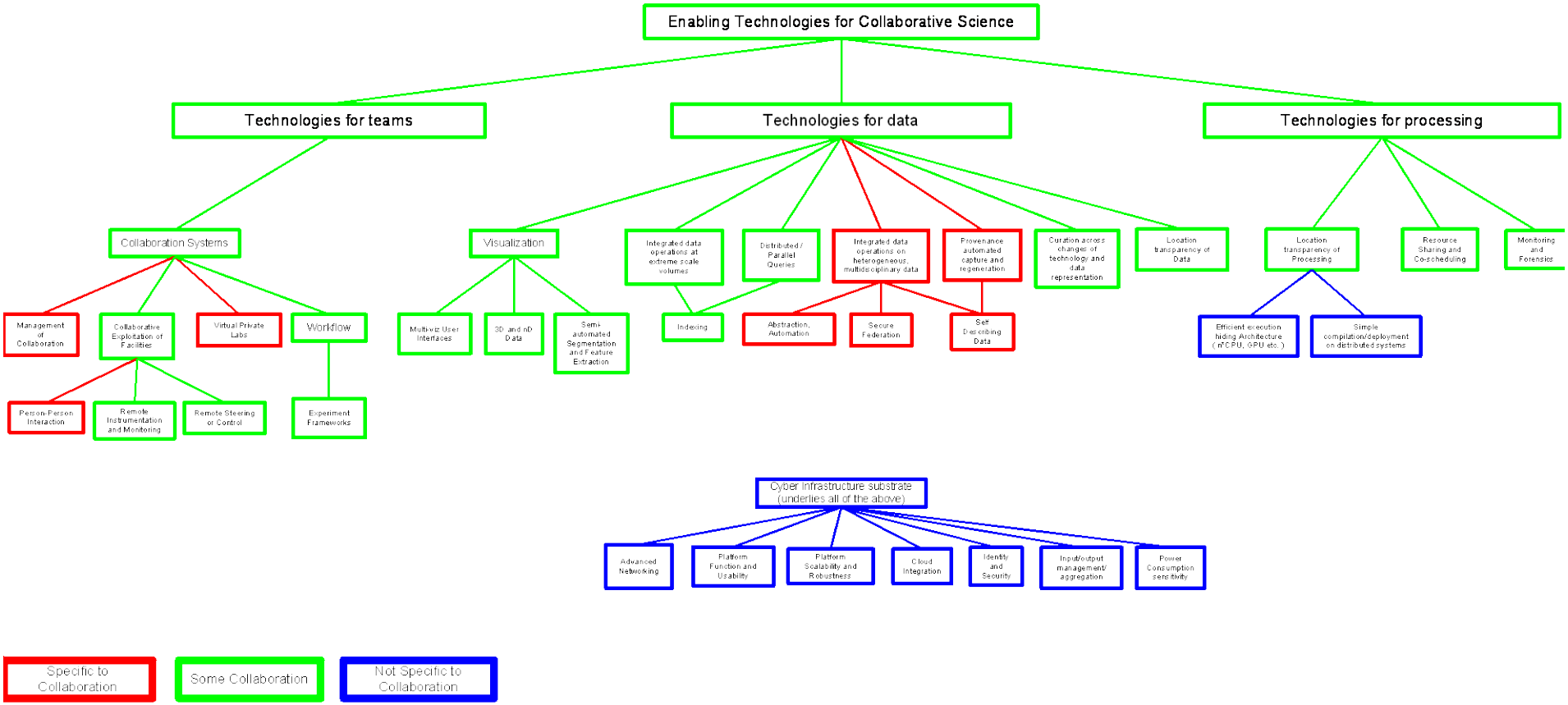
Workshop Details

- **December 6-7, 2011**
- **40 individuals on Organizing Committee**
- **88 Researchers identified as potential attendees**
  - From all 6 DOE-SC program offices
  - From other agencies
- **54 co-authors on the report**
- **3 breakout sessions**
  - Technologies for Processes
  - Technologies for Data
  - Technologies for Teams

Workshop Report

- **General Findings and Recommendations**
  - Unimpeded collaboration accelerates and empowers extreme-scale science
  - Impediments to Collaboration for extreme-scale science can be readily identified
  - Removing the impediments and empowering collaboration requires advances in several areas
- **2 Findings and Recommendations for**
  - Technologies for Team
  - Technologies for Data
  - Technologies for Processing
  - Focusing the SC Strategy for the Support of Collaboration

That beautiful figure …



Organizing Committee for Scientific Collaboration for Extreme-Scale Science Workshop -Draft/Provisional

Interested Science Communities

- **Accelerator Science**
- **Astronomy – LSST**
- **Astrophysics Simulations**
- **Biosciences**
- **Climate Prediction**
- **Combustion**
- **Environmental Molecular Science**
- **Fusion Energy Science**
- **Lattice Field Theory for High Energy and Nuclear Physics**
- **Materials Genome**
- **Open Science Grid**
- **Earth Systems Grid**

Primitive Collaboration Systems Hinder Discovery

DOE investigators and facility users are all information workers, engaged in consuming and producing data. Yet:

- **Journal article is a primary information exchange method**
- **The paper notebook is widely used to document research**
- **The desktop drive is a primary data storage method**
- **Most computational results cannot be reproduced**
- **Email, telephone, airplane are primary collaboration tools**
- **Security concerns are a frequent obstacle to collaboration**
- **Much modeling and simulation is performed using spreadsheets and proprietary packages**

A Digital Laboratory System that Accelerates Discovery

- **All data, code, and documents system-wide are accessible, discoverable, reusable, reproducible, computable, …**
- **… and are linked by a distributed knowledge base that permits automated navigation of content and connections**
- **Advanced software and computational processes are available on-demand and used routinely by every researcher**
- **Collaboration occurs within spaces that people want to use even when they are not collaborating**
- **Intrinsic and proactive security mechanisms encourage rather than discourage collaboration, while protecting against attacks**
- **These capabilities are as intuitive, flexible, and collaborative as the best consumer software (Amazon, Apple, Google, …)**

Unifying Abstractions

- **Discovery:** All resources are easy to find and understand.
  - "I cannot use resources that I do not know exist!"
- **Centrality:** Standardized services reduce costs and encourage commonality.
  - "Don't make me install software or learn arcane details to collaborate!"
- **Portability:** Resources are widely usable in a transparent fashion.
  - "If I can't use your data or software, it isn't science!"
- **Connectivity:** Where information came from, and what other information it relates to, are easy to find.
  - "No information exists in isolation!"

Conclusions

- **DOE-SC communities need a new suite of collaboration tools and services**
- **ASCR NGNS program contains several elements**
  - Fundamental research to identify and develop new theories, algorithms, and basic components needed to support extreme-scale science
  - Base research to identify and develop new tools and services to support current and future collaborations
- **Strategies to transition NGNS research to science communities and long term sustainability**
- **Workshop report provides finding and recomendations on how to move forward**