ACCI Recommendations on Long Term Cyberinfrastructure Issues: Building Future Development

Jeremy Fischer

Indiana University

9 September 2014

Citation:

Fischer, J.L. 2014. "ACCI Recommendations on Long Term Cyberinfrastructure Issues: Building Future Development." Indiana University, Bloomington, IN. September, 2014. Available at: http://hdl.handle.net/2022/18713

PERVASIVE TECHNOLOGY INSTITUTE
INDIANA UNIVERSITY

RESEARCH TECHNOLOGIES
INDIANA UNIVERSITY
University Information Technology Services
Pervasive Technology Institute

In 2009, the National Science Foundation's (NSF) Advisory Committee for Cyberinfrastructure (ACCI) created six task forces to make recommendations to the NSF in the strategic areas of cyberinfrastructure: Campus Bridging, Data, Grand Challenges and Virtual Organizations, High Performance Computing, Software and Tools, and Work Force Development. These task forces were led by ACCI members and consisted of members of the industrial and academic communities. Over a two-year period, through meetings, workshops, position and white papers, and other outreach, these task forces gathered ideas and information in their respective areas and composed a final report outlining recommendations for furthering cyberinfrastructure to support NSF research.

In looking at the development of the next-generation cyberinfrastructure for NSF research, it is helpful to have the recommendations of the previous ACCI task forces. These recommendations may give a strong foundation for future designs. In this paper, the relevant findings of the various ACCI task forces will be covered and included. With this guidance as the bedrock, the next generation cyberinfrastructure will benefit from the experience of a wide swath of NSF researchers and contributors.

Recommendations from the NSF Advisory Committee for Cyberinfrastructure Task Force on Grand Challenges [1]

- NSF should work with the Department of Energy and other agencies in creation of an Interagency Working Group con CS&E or generally on Computational Science and Engineering, including Data-Intensive Computing, in the spirit of other NSF-wide working groups. This broad-based Working Group could provide input leading to important interagency collaborations on new programs, particularly in HPC, and could lead to more focused and efficient use of resources to address the Grand Challenges facing our nation.
- It is recommended that NSF, through OCI, continue to give priority to funding a sustained and diverse set of HPC and innovative equipment resources to support the wide range of needs within the research community. These needs include support for the development of technologies to meet the foremost challenges in HPC, such as the power-aware and application-sensitive architectures, new numerical algorithms to efficiently use petascale and exascale architectures and data flow and data analysis at the extreme scale.

Recommendation from the NSF Advisory Committee for Cyberinfrastructure Task Force on Data and Visualization [2]

- Recognize data infrastructure and services as essential research assets fundamental to today's science and as long-term investments in national prosperity. Make specific budget provisions for the establishment and maintenance of data sets and services and the associated software and visualization tools infrastructure.

Recommendations from the NSF Advisory Committee for Cyberinfrastructure Task Force on Software for Science and Engineering [3]

- NSF should take leadership in promoting verification, validation, sustainability, and reproducibility through software developed with federal support.
- NSF support for software should entail collaborations among all of its divisions, related federal agencies, and private industry
- NSF should develop a consistent policy on open source software that promotes scientific discovery and encourages innovation.

Recommendations from the NSF Advisory Committee for Cyberinfrastructure Task Force on High Performance Computing [4]

- Develop a sustainable model to provide the academic research community with access, by 2015–2016, to a rich mix of HPC systems that:

    - Deliver sustained performance of 20–100 petaflops on a broad range of science and engineering applications;
    - Are integrated into a comprehensive national CI environment; and
    - Are supported at national, regional, and/or campus levels.
- Invest now to prepare for exascale systems that will be available by 2018–2020. NSF should consider the use of co-design partnerships to provide the HPC systems and data CI needed to enable data-driven science.
- Establish a continuing process for soliciting community input on plans for HPC investments

Recommendations from the NSF Advisory Committee for Cyberinfrastructure Task Force on Campus Bridging [5]

- As part of a strategy of coherence between the NSF and campus cyberinfrastructure and reducing reimplementation of multiple authentication systems, the NSF should encourage the use of the InCommon Federation global federated system by using it in the services it deploys and supports, unless there are specific technical or risk management barriers.
- The NSF should create a new program funding high-speed (currently 10 Gbps) connections from campuses to the nearest landing point for a national network backbone. The design of these connections must include support for dynamic network provisioning services and must be engineered to support rapid movement of large scientific data sets.
- The NSF should establish a national cyberinfrastructure software roadmap. Through the Software Infrastructure for Sustained Innovation (SI^2) or other programs, the NSF should seek to systematically fund the creation and ongoing development and support of a suite of critical cyberinfrastructure software that identifies and establishes this roadmap, including cyberinfrastructure software for authentication and access control; computing cluster management; data movement; data sharing; data, metadata, and provenance management; distributed computation / cycle scavenging; parallel computing libraries; network performance analysis / debugging; VO collaboration; and scientific visualization. Funding for personnel should be a strong portion of such a strategy.
- The NSF should fund activities that support the evolution and maturation of cyberinfrastructure through careful analyses of needs (in advance of creating new cyberinfrastructure facilities) and outcomes (during and after the use of cyberinfrastructure facilities). The NSF should establish and fund processes for collecting disciplinary community requirements and planning long-term cyberinfrastructure software roadmaps to support disciplinary community research objectives. The NSF should likewise fund studies of cyberinfrastructure experiences to identify attributes leading to impact, and recommend a set of metrics for the development, deployment, and operation of cyberinfrastructure, including a set of guidelines for how the community should judge cyberinfrastructure technologies in terms of their technology readiness. All NSF-funded cyberinfrastructure implementations should include analysis of effectiveness including formal user surveys. All studies of cyberinfrastructure needs and outcomes, including ongoing studies of existing cyberinfrastructure

In revisiting these recommendations from 2009, examining the progress in the years since, and finding the gaps between 2009's recommendations and 2014's practices, we can establish an updated framework for building the next generation cyberinfrastructure. Part of Indiana University's ongoing commitment to the XSEDE project and notably the Campus Bridging program is to help extend the scientific cyberinfrastructure seamlessly into more organizations as well as into organizations that are currently not well served or represented. Designing the next generation of cyberinfrastructure with that in mind might allow for both broader and deeper research that goes beyond the traditional organizations represented.

Indiana University has also significantly contributed in other ways to the recommendations the ACCI task forces provided. IU has partnered with the Texas Advanced Computing Center (TACC) to further the goals of creating collaborative space for data analysis and management. [6] Wrangler will provide fast, reliable, and replicated storage for researchers, a necessary part of cyberinfrastructure for managing and sharing data.

IU has also been a major contributor to the Apache Airavata project. In contributing to Airavata, IU is helping to create sustainable, easily implemented, free software for "executing and managing computational jobs and workflows on distributed computing resources including local clusters, supercomputers, national grids, academic and commercial clouds." [7] In implementing Airavata, IU and its collaborators are helping to lower the barriers to research computing and to simplify specific workflows.

As part of the goals of supporting grand challenges, IU houses and supports one of two computation and data storage clusters in support of the ATLAS Midwest Tier 2 Center (MWT2). These systems provide "a high-throughput, data-intensive analysis and simulation facility for proton collisions recorded by the ATLAS detector at the Large Hadron Collider (LHC), based at the CERN Laboratory in Geneva, Switzerland." [8] In combination with the Open Science Grid, whose Grid Operations Center is based at IU, these resources are interconnected with other Tier 1 and Tier 2 centers, creating a single service for researchers.

Creating an infrastructure that is more easily accessible to a broader range of researchers at all types of universities should be a primary goal of any design ideas. In that process, designers must take into account the jumps in data size and the need to transport it quickly and easily for collaboration. Current state of the art focuses on reliability and robustness, and has made significant gains in providing solid services in a fashion familiar to a very large number of users. In order for the national cyberinfrastructure ecology to evolve, it must reach new research communities, including the social sciences and the humanities and support novel types of analyses. The design goals for this process must keep an emphases on reducing barriers to collaboration, dealing with the increasing scale of data sets and analyses, and supporting new and infrastructures for analyses in a broad range of scientific domains. The end result should be a design for the next generation that concentrates on being inclusive, reliable, fast, and seamless.

References

[1] NSF Advisory Committee for Cyberinfrastructure Task Force on Grand Challenges. Final Report. March 2011. http://www.nsf.gov/cise/aci/taskforces/TaskForceReport_GrandChallenges.pdf

[2] NSF Advisory Committee for Cyberinfrastructure Task Force on Data and Visualization. Final Report. March 2011. http://www.nsf.gov/cise/aci/taskforces/TaskForceReport_Data.pdf

[3] NSF Advisory Committee for Cyberinfrastructure Task Force on Software for Science and Engineering. Final Report. March 2011. http://www.nsf.gov/cise/aci/taskforces/TaskForceReport_Software.pdf

[4] NSF Advisory Committee for Cyberinfrastructure Task Force on High Performance Computing. Final Report. March 2011. http://www.nsf.gov/cise/aci/taskforces/TaskForceReport_HPC.pdf

[5] NSF Advisory Committee for Cyberinfrastructure Task Force on Campus Bridging. Final Report. March 2011. http://www.nsf.gov/cise/aci/taskforces/TaskForceReport_CampusBridging.pdf

[6] IU plays key role in two new supercomputing systems. IT News. January 14, 2014. http://itnews.iu.edu/articles/2014/iu-plays-key-role-in-two-new-national-supercomputing-systems.php

[7] Apache Airavata. Retrieved September 9, 2014. http://airavata.apache.org/

[8] MWT2 : ATLAS Midwest Tier 2 Center. Laboratory for Advanced Numerical Simulations. Retrieved September 9, 2014. http://www.mcs.anl.gov/research/LANS/research/index.php?p=proj_detail&id=86

Acknowledgements

This document was developed with support from National Science Foundation (NSF) grant OCI-1053575 for "XSEDE: eXtreme Science and Engineering Discovery Environment." Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the NSF.

More information is available at http://xsede.org

This document was also made possible with support by the Indiana University Pervasive Technology Institute.

More information is available at http://pti.iu.edu