FYI's from NSF Large Facilities Meeting

Frank Würthwein
OSG Executive Director
Professor of Physics
UCSD/SDSC

Open Science Grid

UCSD

The purpose of this presentation is to make the Council aware of the 2nd "NSF Large Facilities Cyberinfrastructure Workshop"

http://facilitiesci.org

Some Slides from Irene Qualter's Presentation

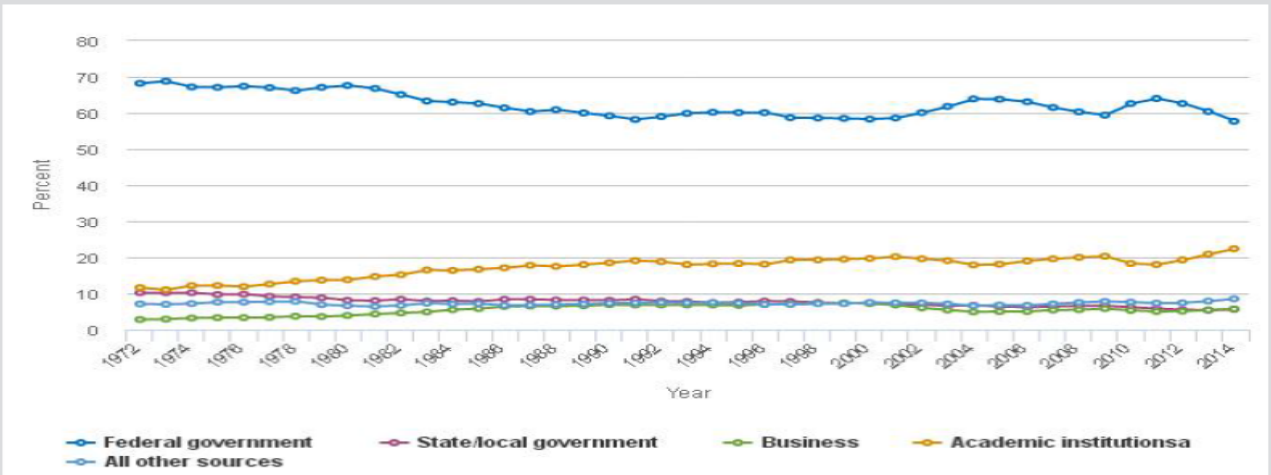
University Investment is the growing slice of the pie.

Trends in Academic R&D Funding of Science and Engineering

National Science Board | **Science & Engineering Indicators 2016**

Figure 5-1

Academic S&E R&D expenditures, by source of funding: FYs 1972–2014



[a] Academic institutions' funds exclude research funds spent from multipurpose accounts.

SOURCE: National Science Foundation, National Center for Science and Engineering Statistics, Higher Education Research and Development Survey.

Science and Engineering Indicators 2016

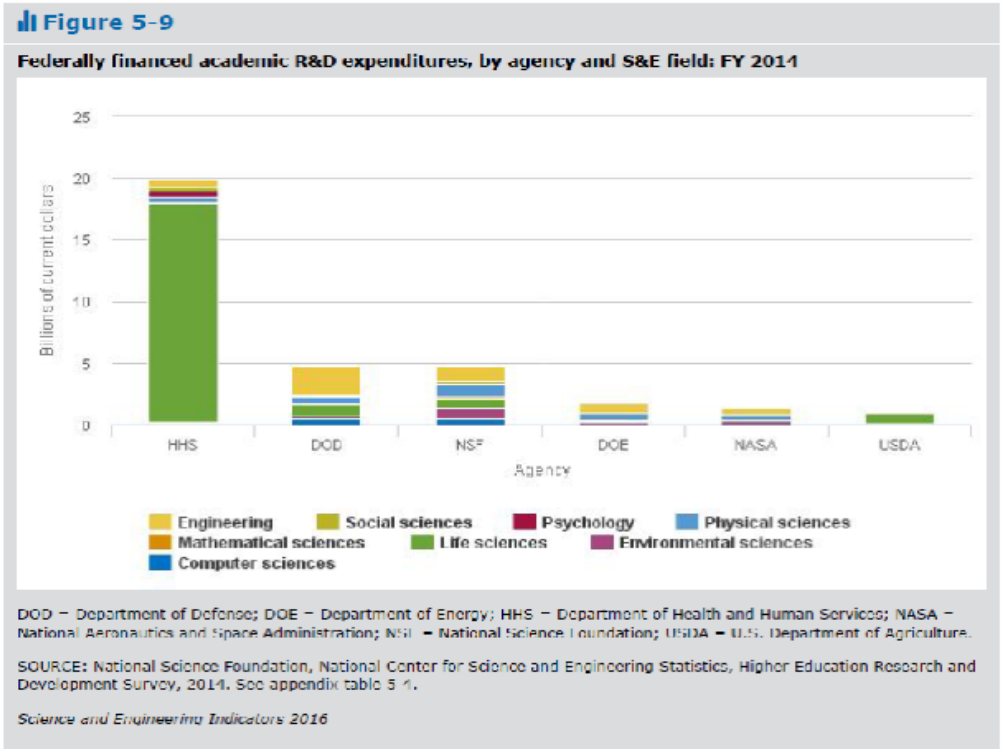
Clear long term and short term trend

I.Q. slide 3

3 Tiers: NIH, DOD&NSF, Others

I.G. slide 5: Points out that there are 3 tier's in scale, and that NSF is the most broad among the agencies.

NSF view of Science and Engineering Frontiers is Multidisciplinary

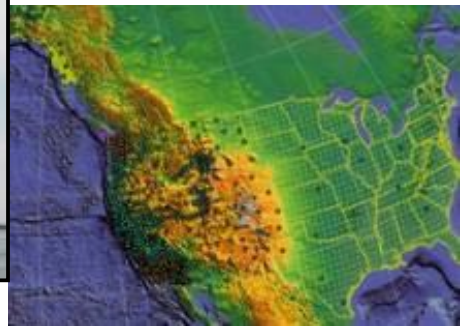


I.Q. slide 12

Facilities are Increasingly CI Intensive … and dependent on robust, reliable, and highly connective CI





I.Q. slide 17

NSF "Big Ideas" – each have demanding CI implications

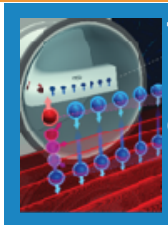
RESEARCH IDEAS

Harnessing Data for 21st Century Science and Engineering

Work at the Human-Technology Frontier: Shaping the Future

Navigating the New Arctic

Windows on the Universe: The Era of Multi-messenger Astrophysics



The Quantum Leap: Leading the Next Quantum Revolution

Understanding the Rules of Life: Predicting Phenotype

PROCESS IDEAS

Mid-scale Research Infrastructure

NSF 2050: Seeding Innovation

Growing Convergent Research at NSF

NSF-INCLUDES: Enhancing Science and Engineering through Diversity

OSG has an obvious role to play here as we support CI for 2/3 of the pictured instruments.

NSF "Big Ideas" – each have demanding CI implications

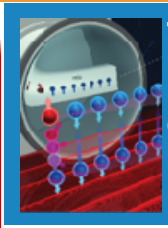
RESEARCH IDEAS

Harnessing Data for 21st Century Science and Engineering

Work at the Human-Technology Frontier: Shaping the Future

Navigating the New Arctic

Windows on the Universe: The Era of Multi-messenger Astrophysics



The Quantum Leap: Leading the Next Quantum Revolution

Understanding the Rules of Life: Predicting Phenotype

PROCESS IDEAS

Mid-scale Research Infrastructure

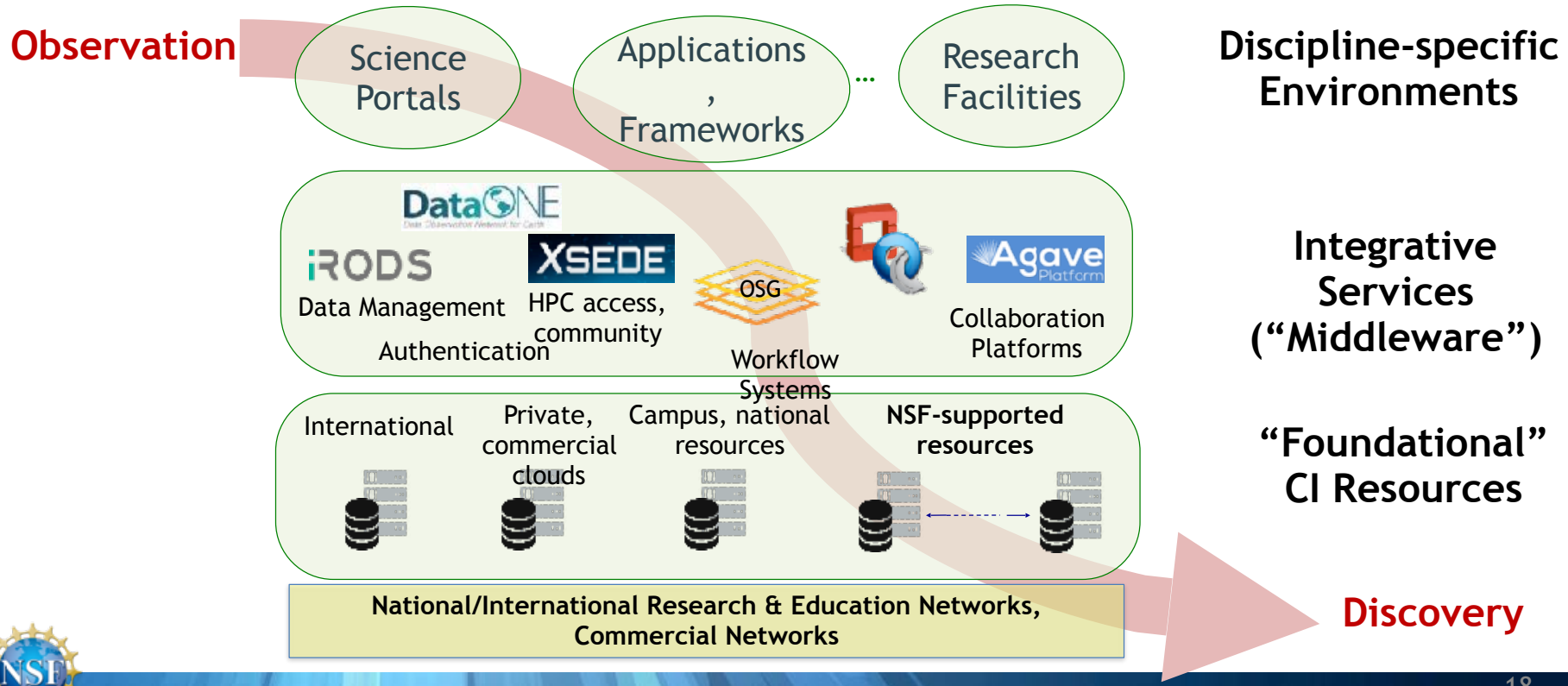
NSF 2050: Seeding Innovation

Growing Convergent Research at NSF

NSF-INCLUDES: Enhancing Science and Engineering through Diversity

I.Q. slide 18

Dynamic discovery pathways at scale: Architecture view



A Process to Redefine Goals

Community input is informing NSF's strategic planning for Advanced CI

- ✓ NSF Advisory Committee for Advanced Cyberinfrastructure (ACCI)
- ✓ National Academies report on NSF Advanced Computing (2016):
  - *Future Directions of NSF Advanced Computational Infrastructure to Support US Science in 2017 - 2022*
- ✓ NSF RFI on Future Needs for Advanced Cyberinfrastructure to Support Science and Engineering Research (NSF CI 2030), (2017)
- ✓ Joint agency assessment of the NSCI Exascale RFI (2015):
  - *NSF Assessment of Responses to the Request for Information (RFI) on Science Drivers Requiring Capable Exascale High Performance Computing*
- ✓ Workshops and Reports, e.g.:
  - Software Infrastructure 2017 PI Workshop ,
  - *Building a Materials Data Infrastructure*

Aside on NAS Report

Summary

The National Science Foundation (NSF) asked the National Academies of Sciences, Engineering, and Medicine to provide a framework for future decision-making about NSF's advanced computing strategy and programs. Advanced computing refers here to the advanced technical capabilities, including computer systems, software, and expert staff, that support a wide range of science and engineering research and that are of a large enough scale and cost that they are typically shared among multiple researchers, institutions, and applications. Advanced computing encompasses support for data-driven research as well as modeling and simulation.

The recommendations of the Committee on Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science in 2017-2020 are aimed at achieving four broad goals: (1) positioning the United States for continued leadership in science and engineering, (2) ensuring that resources meet community needs, (3) aiding the scientific community in keeping up with the revolution in computing, and (4) sustaining the infrastructure for advanced computing.

POSITION THE UNITED STATES FOR CONTINUED LEADERSHIP IN SCIENCE AND ENGINEERING

Large-scale simulation and the accumulation and analysis of massive amounts of data are revolutionizing many areas of science and engineering research. Increased advanced computing capability has historically enabled new science, and many fields today rely on high-throughput computing for discovery. Modeling and simulation, the historical focus of high-performance computing, is a well-established peer of theory and experiment. Data-driven research, a complementary "fourth paradigm" for scientific discovery, needs data-intensive computing capabilities and resources. To support this research, NSF is a major provider of the advanced computing used for U.S. basic science, for not only its own grantees but also in support of research sponsored by other agencies, such as the National Institutes of Health and the Department of Energy.

Meeting future needs will require systems that support a wide range of advanced computing capabilities, including large-scale parallel systems and data-intensive systems. Approaches that combine large-scale computing and data resources in "converged" systems can play a role; more specialized systems may also be needed to meet some requirements. Commercial cloud computing offers certain advantages and can play a role in NSF's advanced computing strategy. However, NSF computing centers already exploit economies of scale and load sharing, and commercial cloud providers do not currently support very large, tightly coupled parallel applications, especially for high-end simulation workloads. For other applications, especially data-centric workloads and communities that share data sets, cloud computing is positioned today to play a growing role.

Recommendation 1. NSF should sustain and seek to grow its investments in advanced computing—to include hardware and services, software and algorithms, and expertise—to ensure that the nation's researchers can continue to work at frontiers of science and engineering.

Recommendation 1.1. NSF should ensure that adequate advanced computing resources are focused on systems and services that support scientific research. In the future, these requirements will be captured in its roadmaps.

On Page 1 of the Summary it says:

Increased advanced computing capability has historically enabled new science, and many fields today rely on high-throughput computing for discovery.

We could not have asked for a more prominent mention of HTC in this report !!!

Aside on NAS Report (II)

Recommendation 2.1. NSF should integrate support for the revolution in data-driven science into NSF's strategy for advanced computing by (a) requiring most future systems and services and all those that are intended to be general purpose to be more data-capable in both hardware and software and (b) expanding the portfolio of facilities and services optimized for data-intensive as well as numerically-intensive computing, and (c) carefully evaluating inclusion of facilities and services optimized for data-intensive computing in its portfolio of advanced computing services.

Recommendation 2.2. NSF should (a) provide one or more systems for applications that require a single, large, tightly coupled parallel computer and (b) broaden the accessibility and utility of these large-scale platforms by allocating high-throughput as well as high-performance workflows to them.

It's hard too imagine language that is more directly advocating for what we do best.

I.Q. Last Slide

Draft OAC Principles

- Promote Science Excellence
  - Enable fundamentally new scientific advances
  - Attend to the current trends in research
    - Multidisciplinary, geo/institutionally-agnostic research and research teams
    - Complex problems; dynamic workflows; data-rich
    - Robust and reliable science
- Focus on Unique NSF contributions to CI support for Research
  - Holistic view:
    - Build capability, capacity, and cohesiveness of national CI Ecosystem
    - Consider both human and technical aspects of CI
  - "Lean forward" to new approaches and technologies
- Act as a model steward
  - Encourage reuse of investments in CI
    - from industry, federal agencies, academic institutions, etc.
  - Foster partnerships and community development
  - Incent measurement and sharing of results

I.Q. Last Slide

Draft OAC Principles

- Promote Science Excellence
  - Enable fundamentally new scientific advances
  - Attend to the current trends in research
    - Multidisciplinary, geo/institutionally-agnostic research and research teams
    - Complex problems; dynamic workflows; data-rich
    - Robust and reliable science
- Focus on Unique NSF contributions to CI support for Research
  - Holistic view:
    - Build capability, capacity, and cohesiveness of national CI Ecosystem
    - Consider both human and technical aspects of CI
  - "Lean forward" to new approaches and technologies
- Act as a model steward
  - Encourage reuse of investments in CI
    - from industry, federal agencies, academic institutions, etc.
  - Foster partnerships and community development
  - Incent measurement and sharing of results

Take Home Messages

- OSG is as relevant today as it has ever been.
  - HTC is totally mainstream, and nobody does it better than us.
- We have two obvious slam dunks to focus on in our pitch to the NSF:
  - Large Facilities
    - LIGO, IceCube, DES, LSST, SPTPol, Xenon1T, … are our natural allies in addition to the LHC.
  - University CI
    - especially when it comes to inter institutional integration.