Cyberinfrastructure at NSF/ACI

Daniel S. Katz

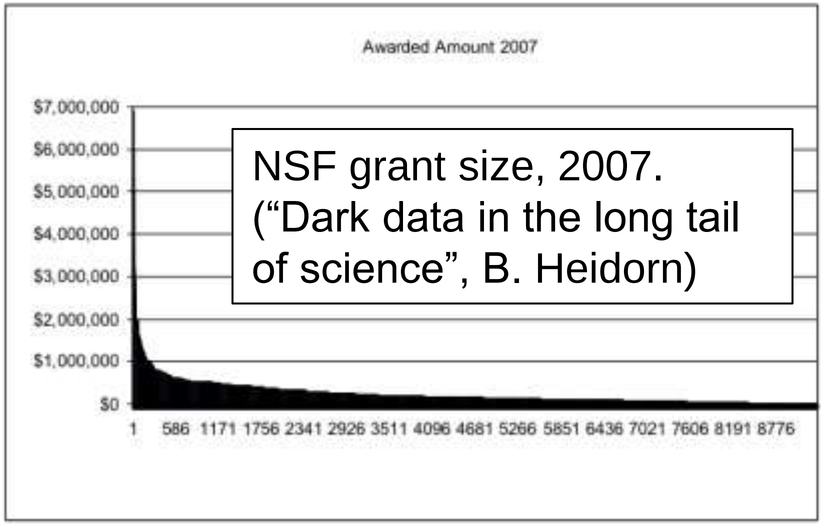
Program Director, Division of Advanced Cyberinfrastructure

Big Science and Infrastructure

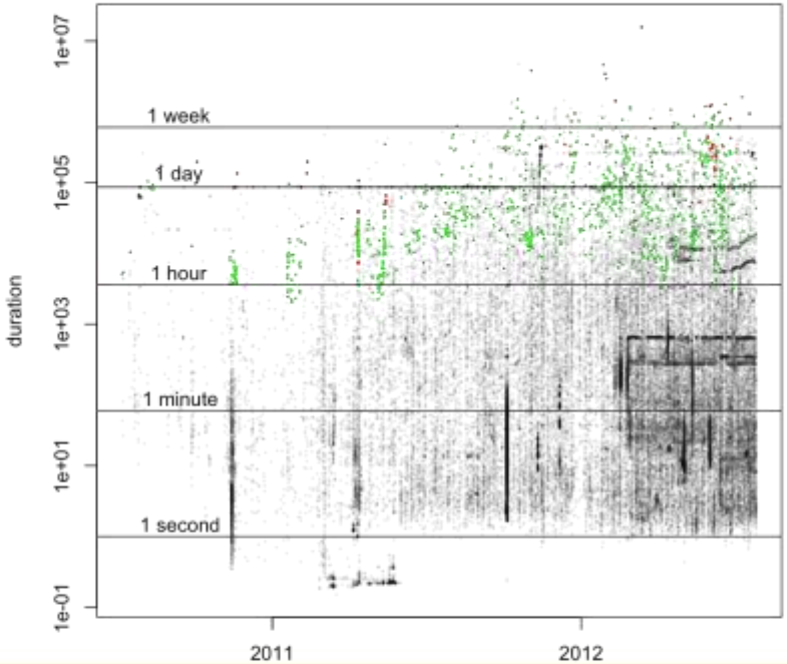
- Hurricanes affect humans
- Multi-physics: atmosphere, ocean, coast, vegetation, soil
  - Sensors and data as inputs
- Humans: what have they built, where are they, what will they do
  - Data and models as inputs
- Infrastructure:
  - Urgent/scheduled processing, workflows
  - Software applications, HPC systems
  - Networks
  - Decision-support systems, visualization
  - Data storage, interoperability

Long-tail Science and Infrastructure

- Exploding data volumes & powerful simulation methods mean that more researchers need advanced infrastructure
- Such "long-tail" researchers cannot afford expensive expertise and unique infrastructure
- Challenge: Outsource and/or automate time-consuming common processes
  - Tools, e.g., Globus Online and data management
    - Note: much LHC data is moved by Globus GridFTP, e.g., May/June 2012, >20 PB, >20M files
  - Gateways, e.g., nanoHUB, CIPRES, access to scientific simulation software



Duration of runs, in seconds, over time.
Red: >10 TB transfer; green: >1 TB transfer.



Infrastructure Challenges

- Science
  - Larger teams, more disciplines, more countries
- Data
  - Size, complexity, rates all increasing rapidly
  - Need for interoperability (systems and policies)
- Systems
  - More cores, more architectures (GPUs), more memory hierarchy
  - Changing balances (latency vs bandwidth)
  - Changing limits (power, funds)
  - System architecture and business models changing (clouds)
  - Network capacity growing; increase networks -> increased security
- Software
  - Multiphysics algorithms, frameworks
  - Programing models and abstractions for science, data, and hardware
  - V&V, reproducibility, fault tolerance
- People
  - Education and training
  - Career paths
  - Credit and attribution

Cyberinfrastructure

- *"Cyberinfrastructure consists of computing systems, data storage systems, advanced instruments and data repositories, visualization environments, and people, all linked together by software and high performance networks to improve research productivity and enable breakthroughs not otherwise possible."*

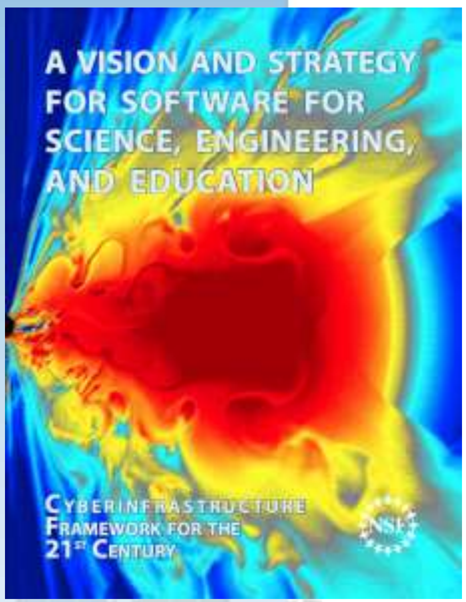
  -- Craig Stewart

- Infrastructure elements:
  - parts of an infrastructure,
  - developed by individuals and groups,
  - international,
  - developed for a purpose,
  - used by a community

Cyberinfrastructure Framework for 21st Century Science and Engineering (CIF21)

- Cross-NSF portfolio of activities to provide integrated cyber resources that will enable new multidisciplinary research opportunities in all science and engineering fields by leveraging ongoing investments and using common approaches and components (http://www.nsf.gov/cif21)
- ACCI task force reports (http://www.nsf.gov/od/oci/taskforces/index.jsp)
  - Campus Bridging, Cyberlearning & Workforce Development, Data & Visualization, Grand Challenges, HPC, Software for Science & Engineering
  - Included recommendation for NSF-wide CDS&E program
- Vision and Strategy Reports
  - ACI - http://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf12051
  - Software - http://www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf12113
  - Data - http://www.nsf.gov/od/oci/cif21/DataVision2012.pdf
- Implementation
  - Implementation of Software Vision
    http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504817

Software Vision



NSF will take a leadership role in **providing software as enabling infrastructure** for science and engineering research and education, and in **promoting software** as a principal component of its comprehensive CIF21 vision

...

Reducing the complexity of software will be a unifying theme across the CIF21 vision, **advancing** both the **use and development of new software** and **promoting the ubiquitous integration of scientific software across all disciplines**, in education, and in industry

- A Vision and Strategy for Software for Science, Engineering, and Education – NSF 12-113

Infrastructure Role & Lifecycle

Create and maintain a CI ecosystem providing new ***capabilities*** that advance and accelerate scientific inquiry at unprecedented complexity and scale

Support the foundational ***research*** necessary to continue to efficiently advance CI

Enable transformative, interdisciplinary, collaborative, ***science and engineering*** research and education through the use of CI

Transform practice through new ***policies*** for CI addressing challenges of academic culture, open dissemination and use, reproducibility and trust, curation, sustainability, governance, citation, stewardship, and attribution of CI authorship

Develop a next generation diverse workforce of scientists and engineers equipped with essential skills to use and develop CI, with CI used in both the research and ***education*** process

ACI Program Clusters

Systems | Networks | Data | Software

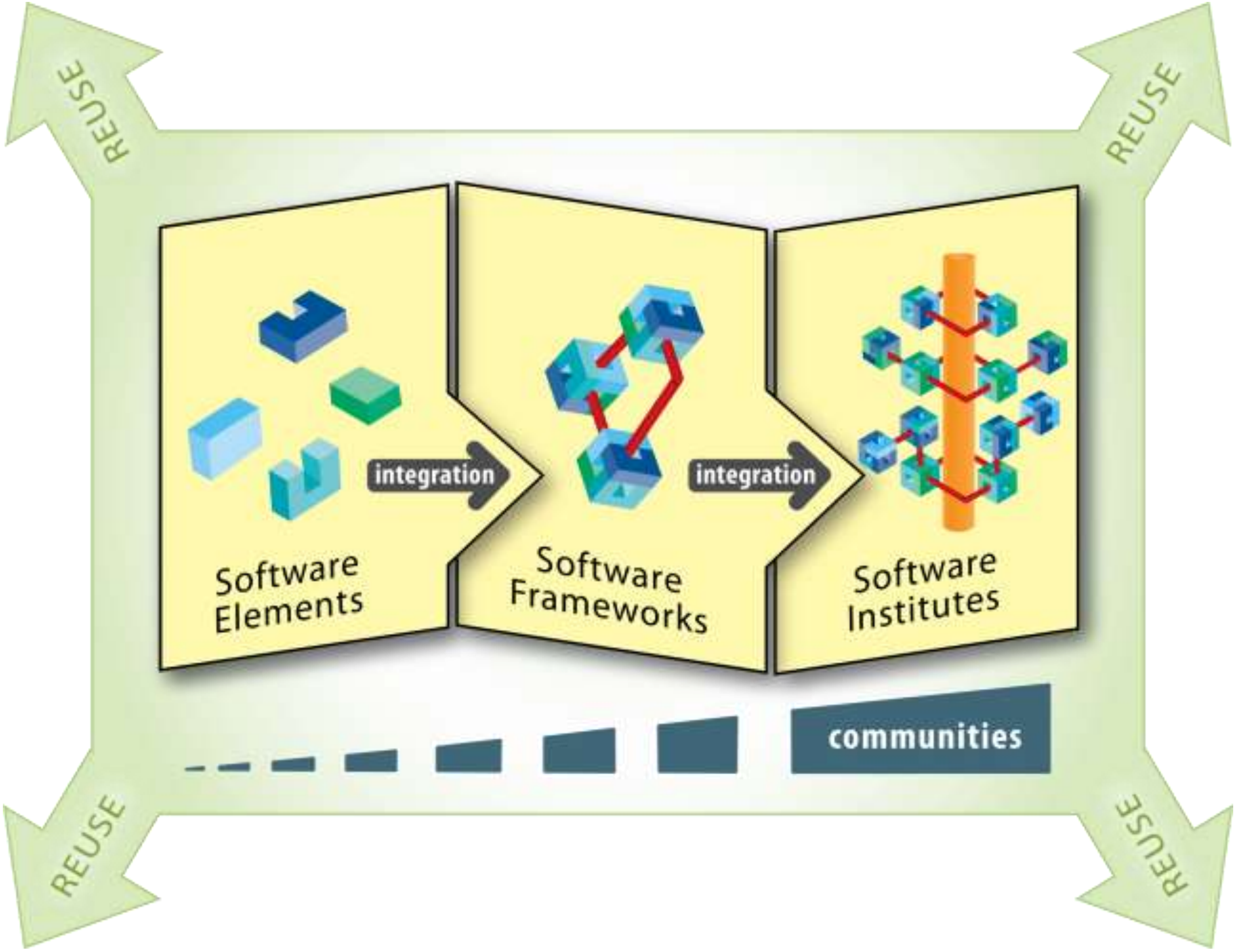
ACI Objectives

- Build and sustain current generation
- Use current generation for science and engineering
- Research and develop next generation
- Learning and workforce development to use current generation and build next generation

ACI Software Cluster Programs

- Exploiting Parallelism and Scalability (XPS)
  - CISE (including ACI) program for foundational groundbreaking research leading to a new era of parallel (and distributed) computing
  - First set of proposals submitted in Feb. 2013, awards in progress
- Computational and Data-Enabled Science & Engineering (CDS&E)
  - Virtual program for science-specific proofing of algorithms and tools
    - ENG, MPS, ACI now; BIO, GEO, IIS in FY14?
  - Identify and capitalize on opportunities for major scientific and engineering breakthroughs through new computational and data analysis approaches
- Software Infrastructure for Sustained Innovation (SI^2)
  - Transform innovations in research and education into sustained software resources that are an integral part of the cyberinfrastructure
  - Develop and maintain sustainable software infrastructure that can enhance productivity and accelerate innovation in science and engineering

Software Infrastructure Projects



SI^2 Software Activities

- Elements (SSE) & Frameworks (SSI)
  - Past general solicitations, with most of NSF (BIO, CISE, EHR, ENG, MPS, SBE): NSF 10-551 (2011), NSF 11-539 (2012)
    - About 27 SSE and 20 SSI projects (19 SSE & 13 SSI in FY12)
  - Focused solicitation, with MPS/CHE and EPSRC: US/UK collaborations in computational chemistry, NSF 12-576 (2012)
    - 4 SSI projects
  - Recent solicitation (NSF 13-525), continues in future years
    - 14 SSE & ~11 SSI projects being funded
- Institutes (S2I2)
  - Solicitation for conceptualization awards, NSF 11-589 (2012)
    - 13 projects (co-funded with BIO, CISE, ENG, MPS)
  - Full institute solicitation in late FY14
- See http://bit.ly/sw-ci for current projects

General Infrastructure Questions

- Elements that are intended to be infrastructure has challenges
  - Unlike in business, more users means more work
  - The last 20% takes 80% of the effort
  - What can NSF do to make these things easier?
- What fraction of funds should be spent of support of existing infrastructure vs. development of new infrastructure?
  - How do we decide when to stop supporting a element?
- How do we encourage reuse and discourage duplication?
- How do we more effectively support career paths for software developers (with universities, labs, etc.)

Questions?

- Now, or later

  dkatz@nsf.gov

  or

  d.katz@ieee.org