My Goals from this Workshop

- Listen!
- Share OAC’s view on CI and software.
- What we learn here should inform our programs
GATEWAYS
Science Gateways Abound!
These are some that use XSEDE supercomputers
Gateways are used for a variety of purposes:

- Educational tools: 18%
- Computational tools: 16%
- Collaboration tools: 8%
- Data collections: 15%
- Data analysis tools, including visualization and mining: 16%
- Workflows: 6%
- Citizen science resources: 5%
- Interfaces to scientific instruments: 4%
- Interfaces to sensor data: 4%
- Other: 2%
Cyberinfrastructure for Phylogenetic Research (CIPRES)

- 210 US research universities
  - Harvard, Yale, UC Berkeley, Stanford, etc.
  - Non-PhD granting colleges (including one all-women’s college, community colleges, and Hispanic-serving institutions)
- 3 K-12 school systems
- 43 non-governmental organizations,
  - Museums including the Smithsonian Institution, the American Museum of Natural History, and the Field Museum,
  - Botanical gardens, (e.g. Chicago, Rancho Santa Ana, and New York)
  - Institutes (e.g. JCVI and Broad)
- 10 US governmental agencies
  - Including NIH, USDA, NOAA, US Forest Service
- Curriculum delivery (76)
- 2000+ publications since 2010
- 47% of all XSEDE users in Q4 2015
Gateway Usage Growth

Gateway usage surpasses command line users in XSEDE
Gateways enable reproducibility

The Whole Tale
Merging Science and Cyberinfrastructure Pathways

Whole Tale will enable researchers to examine, transform, and then seamlessly republish research data that was used in an article. As a result, these "living articles" enable new discovery by allowing researchers to construct representations and syntheses of data.

Gateways with ties to publishing!
Gateways Enable Ecosystems!

1. Provide critical tools to scientists that accelerate time to science
2. Make complex techniques and technologies usable, while allowing for evolution and adaption
3. Also enable: reproducibility, security, open science, citizen science etc.
4. Support collaborations – may even serve as the community water-cooler

GATEWAYS ARE KEY FOR ACCELERATING ROBUST AND RELIABLE SCIENCE
CYBER-INFRASTRUCTURE PROGRAMS AT OAC
NSF views cyberinfrastructure as driven by research priorities and evolving with the scientific process.
And informed by community input and experiences

Initial Vision (2007-2010)

NSF-Wide Task Force Reports (2009-2011)

National Academies Study

- CI 2030

Prepare

- DCL: Cleared
- Website: finalizing
- Engagement plan

Collect

- Dissemination
- Monitor responses
- More outreach as needed.

Analyze

- First peek, ACCI Meeting April 19-20
- Analyze insights
- Post responses on NSF CI 2030 Website

Report, Next Steps

- Finalize RFI Report and disseminate
- NSB Meeting? (Aug 15-16)
- Incorporate findings into NSF planning.
OAC Operational View
Supporting advanced CI to accelerate discovery and innovation

Science Drivers
Constant exchange within ACI clusters, and with NSF Directorates, Divisions and Programs

ACI investments
Convergent investments in technologies and communities to maximize impact

Leadership, Coordination, Partnership
CI Challenge: User-Centric Viewpoint

Revolution in the scientific workflow: many interfaces to shared services

Large Facilities

Shared Data/Software Gateway Resources

Collaboration Networks

National Computing Resources

Identities? Resources? Persistence?

Cloud Services

Researcher

Data

Software
Software in a Research Cyberinfrastructure Ecosystem

CIF 21 Vision: Integrated advanced computing, networks, data, software instruments, facilities to advance science and engineering

A national research CI architecture for reuse and agility

Software: A primary modality for innovation and discovery permeating all layers of the Cyberinfrastructure
Goal of the Software Programs

Catalyze and support unique, innovative software-intensive science ecosystems in order to advance science and engineering.
Driver - NSF Big Ideas

- Understanding the Rules of Life: Predicting Phenotype
- Work at the Human-Technology Frontier: Shaping the Future
- Mid-scale Research Infrastructure
- Windows on the Universe: The Era of Multi-messenger Astrophysics
- Navigating the New Arctic
- Harnessing Data for 21st Century Science and Engineering
- The Quantum Leap: Leading the Next Quantum Revolution
- Growing Convergent Research
Drivers – A National CI Ecosystem for Robust and Reliable Science

• A National CI Ecosystem
  – Community establishment (directive to leverage Institutes)
  – Sustainability
  – Building on existing assets
  – Towards an infrastructure “platform”

• Enabling Robust and Reliable Science
  – Repeatability -> Replicability -> Reproducibility
  – Uncertainty quantification
  – Software publication, citation
  – Education
Driver: Future of Computing – the 5 NSCI Objectives

1. Exascale computing system…
   – Foundational work on science, algorithms, programming environments, system software, architecture, and performance evaluation

2. Increasing coherence between … simulation and data analytics…
   – Science and technology that use and enable applications involving both computational simulation and data analysis.

3. A viable path forward … [in] the ‘post-Moore’s Law era
   – Foundational work on new device technology, fabrication methods, computer architectures, software techniques.

4. An enduring National HPC ecosystem…
   – Develop, integrate, and deploy building blocks of an HPC ecosystem.
   – Advance the organization, architecture applications of such a system,
   – Enhance user productivity, broaden participation, skilled workforce.

5. Public-private collaboration…”
   – Existing programs, such as GOALI, SBIR/SBTT, and IUCRC
   – Technology transition to and from practice
   – Advance the use of HPC technology in the commercial sector
Driver: Encouraging Convergence and Co-design Builds the Future Research Infrastructures

Access
Visualization
Data Quality
Collaboration
Tools
Exploratory Analysis

DATA

HYPOTHESES
INFORMATION
DATA IN CONTEXT

Science domains
Systems, algorithms, Foundations, Cyber infrastructure

Workforce

THERORY

DISCOVERY

Analytics
High Performance Computing
Computational-Mathematical - Statistical Methods/Models

Interpretation
Model Validation
Redesign

Experiments
Data Collection
Benchmark
Data Sets
Driver: OAC Specific Priorities

• Multidisciplinary and omni-disciplinary software cyberinfrastructure
• Meaningfully integrate, leverage or build on other ongoing OAC-supported programs
• Integrated innovation and research
• Proposals that consider security, trustworthiness and reproducibility
• Proposals with objectives that align with and contribute to the NSCI.
Driver: Directorate Specific Priorities

- CISE: CI for CISE research. Advancing SE.
- BIO: Of need to BIO + Other Directorates.
- EHR: STEM learning and learning environments, workforce development, and broadening participation.
- ENG: Computational tools that enable in its research areas
- GEO: Serve the geosciences end-users. Integrate with prior investments – EarthCube, integrate CS and GEO researchers.
- MPS: Core research areas, MGI, science at the boundaries, broad application (DMS), Education, community Development
- SBE: SBE 2020, SBE + one other directorate
Driver: Embedded Research in CI Projects Creates Robust Insights that help Build Academic Reputations

Systems Integration:
- New integration techniques - auto-generation of integration code from interface specifications
- Studies of software engineering methods for s/w integration – DevOps, continuous deployment
- Studies of integrative methods for data science
- Empirical studies on software reuse in science
- Analytical models for understanding/evaluating performance, scalability, security during integration
- Service-based integration of data analytics and HPC system architectures

HCI:
- Search based composition of services
- Human-computer interfaces and interaction design and evaluation during integration - e.g. when “surprise” is a given

SBE:
- Ethnographic studies on how scientists actually work
- Economic and social aspects of reuse
- Economic and social aspects of integration
- Science of team science in dynamic situations

Education:
- Learning theories for "just-in-time" application (novice vs. expert learning)

Domain science:
- End-to-end composition of models across scales (neuron->cognition, chip->data center)
Structure: Software Cyberinfrastructure Pipeline

**Research and Development**
- SPX (CISE)
- CDS&E (Cross)
- DMREF (Materials)
- CRISP (Infrastructures)
- EAGER
- Venture
- Other programs and avenues

**Development and Deployment**

**Sustainability**
- Open Source Community
- Institutional research support
- Foundations
- Integration into Education
- SAAS
- IP Licensing
- Commercial Products

**Outside NSF**
Flagship Program – SI2

- Create a software ecosystem that scales from individual or small groups of software innovators to large hubs of software excellence
- 3 interlocking levels of funding

**Scientific Software Elements (SSE)**
1–2 PIs, <$500k, 3 years

**Scientific Software Integration (SSI)**
For focused groups
$200k - $1M per year, 3–5 years

**S²I² Conceptualization Awards**
Planning awards aimed at organizing an interdisciplinary community and understanding its software requirements and challenges.
Unique Criteria For SI2

• Fill a recognized need in the science community
• Create innovative, robust and reliable research capabilities in science and engineering for researchers
• Embed research and innovation into the project activities
• Use a comprehensive user-engaged management plan
• Resourced by teams with credibility in engineering, and science
• Build community through direct engagement
• Progress towards sustainability after NSF funding has ended
• Further a national CI ecosystem (reuse, integrate, adopt)
Showcase – SI2 Projects Across the Stack

Scalable Infrastructure for Multiscale and Multi-Physics Applications

A. Elastic and muscular fibers that model the heart motion of blood in the left (red) and right (blue) motion.

Figure 2. 3D FSI simulation of cardiac dynamics.

BLAS-like Library Instantiation Software (BLIS)

Swift/E: Integrating Parallel Scripted Workflow

Figure 1: Overview of the Proposed MPI Profiling and Performance Engineering Framework
Showcase – SI2 Software Institutes

NSF commits $35 million to improve scientific software

Awards will support long-term hubs dedicated to strengthening scientific software ecosystem

Acknowledgement: Daniel Crawford, MolSSI
Research Workforce is also Infrastructure - We Support The Career Pipeline

- **Goal:** Build robust careers paths in Cyber-Infrastructure (CI) and Computational and Data-enabled Science and Engineering (CDSE)
- **Techniques:** Leverage existing programs for early-stage researchers. Develop new programs in areas of need/challenge

**CAMPUS CLIMATE**

**CURRICULA, Educational Resources**

**RESEARCHERS**

**Cyber Scientists** to develop new capabilities

**Professional Staff** to support new capabilities

**Area Scientists** to exploit new capabilities

**NRT/IGERT**

**REU Sites**

**CI-TraCS**

**CRII**

**CAREER**
State of the Software Programs

- Software is a critical part of the research cyberinfrastructure

- Demand for the software programs stays strong with several successful outcomes to report

- The software programs continue to be well-aligned with OAC, Directorate and NSF priorities
BACK TO GATEWAYS!
Opportunities, Issues, Solutions (?)

1. Opportunities:
   • Global science
   • Broader funding and sustainability base
   • Accelerated learning
   • Embedded research
   • Shared technology

2. Issues:
   • Software licenses
   • Intellectual property (laws)
   • Are there tradeoffs with respect to openness?

3. Solutions?
   • Contributed labor.
   • Other quid pro quos?
Actions

• Co-run this workshop with the SGCI and AU gateways conferences
• Propose a key science or education problem that would be transformed by a gateway with international requirements
Thanks!