



# Towards A National Cyberinfrastructure Ecosystem: NSF Software Programs and Software Institutes

Rajiv Ramnath

Program Director  
Software Cluster

Division of Advanced Cyberinfrastructure  
Directorate for Computer and Information Science and Engineering

[rramnath@nsf.gov](mailto:rramnath@nsf.gov)

Version: 12/6/16 1:12 PM

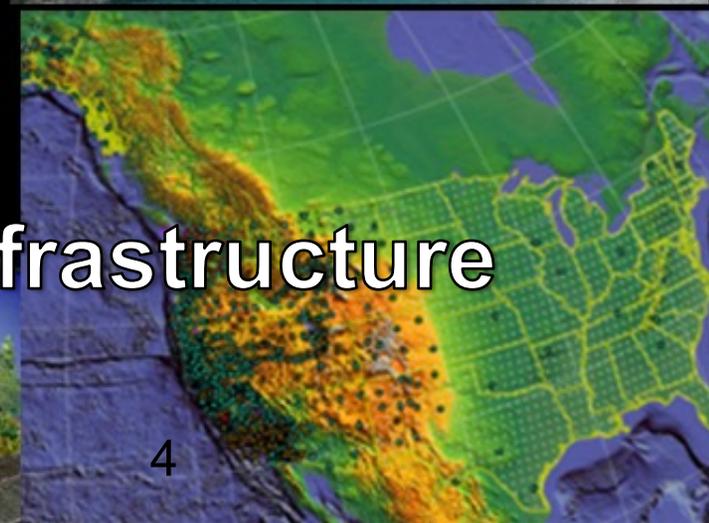
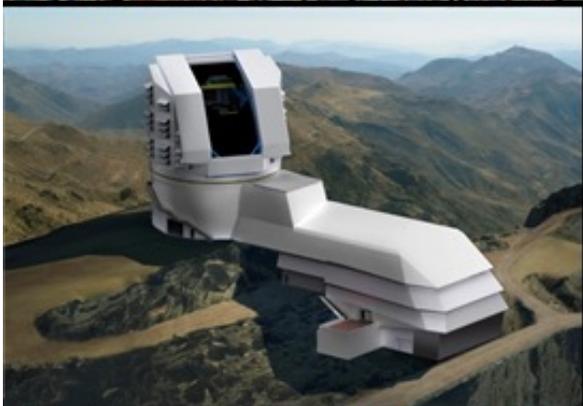
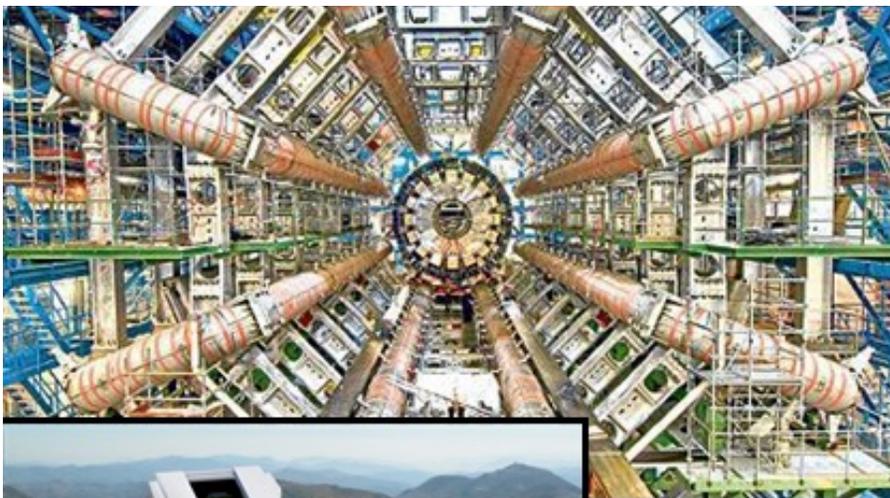


## Outline of Talk

- NSF CI Vision and ACI
- Software CI Programs
- Software Institutes
- Other ACI Programs
- Upcoming Solicitations



# **NSF CI VISION AND ACI'S ROLE IN IT**

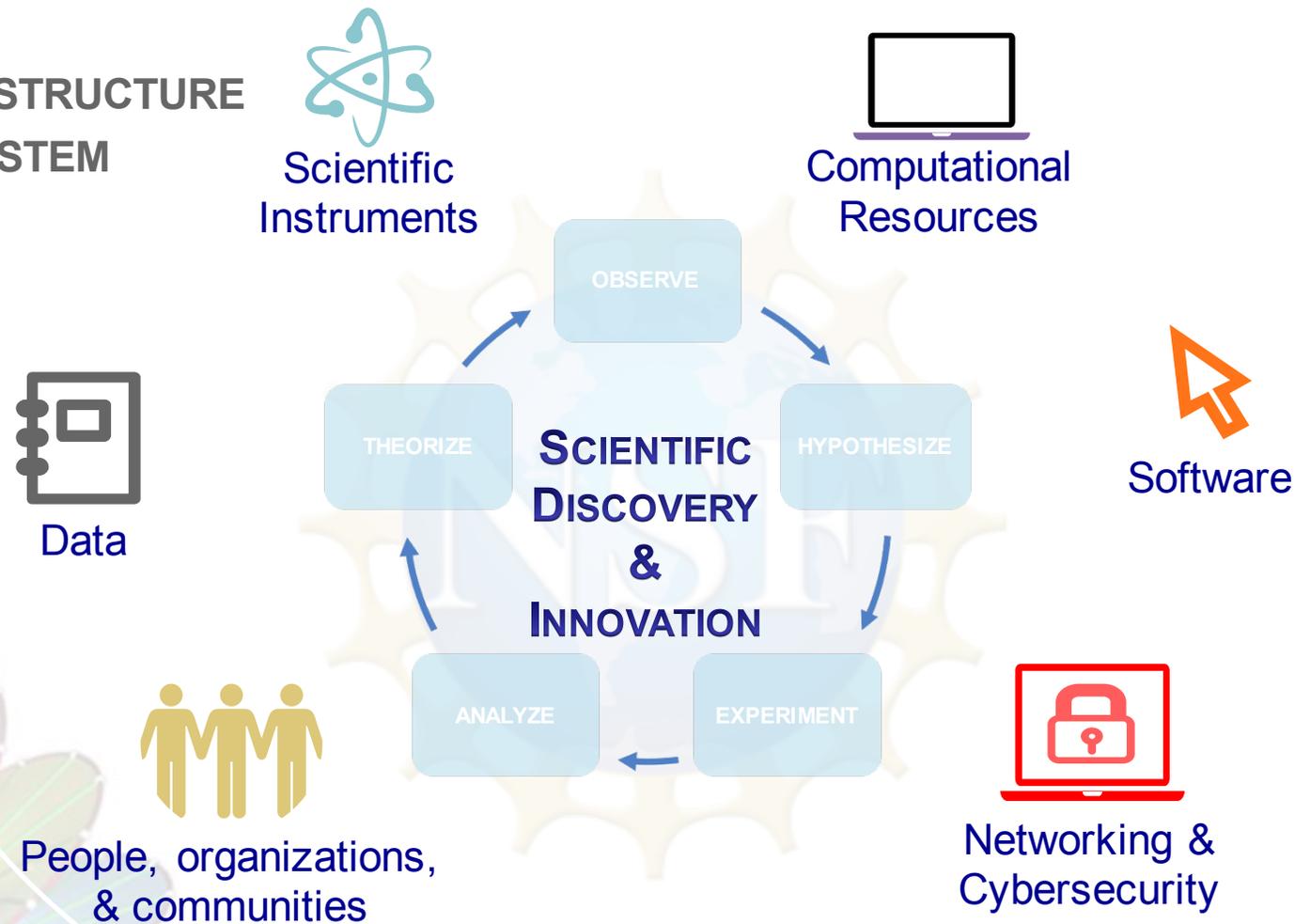


# NSF Supported Science Infrastructure

# NSF views cyberinfrastructure as driven by research priorities and evolving with the scientific process



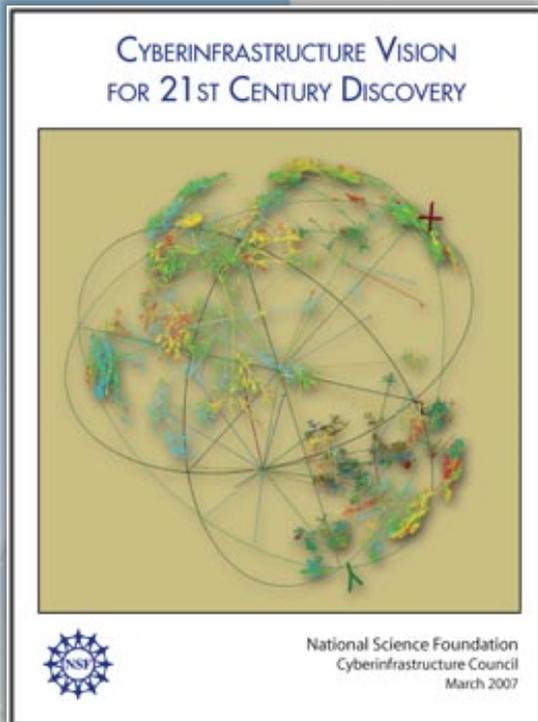
CYBERINFRASTRUCTURE  
ECOSYSTEM





# And informed by community input and experiences

## Initial Vision (2007-2010)



## NSF-Wide Task Force Reports (2009-2011)



## National Academies Study (On going)



Interim Report, Oct 2014  
Final Report expected Fall 2015



# CI Challenge: User-Centric Viewpoint

*Revolution in the scientific workflow: many interfaces to shared services*



Large Facilities



Collaboration Networks

Software



Researcher

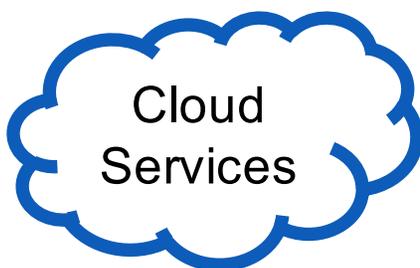
Data



Shared Data/Software Gateway Resources



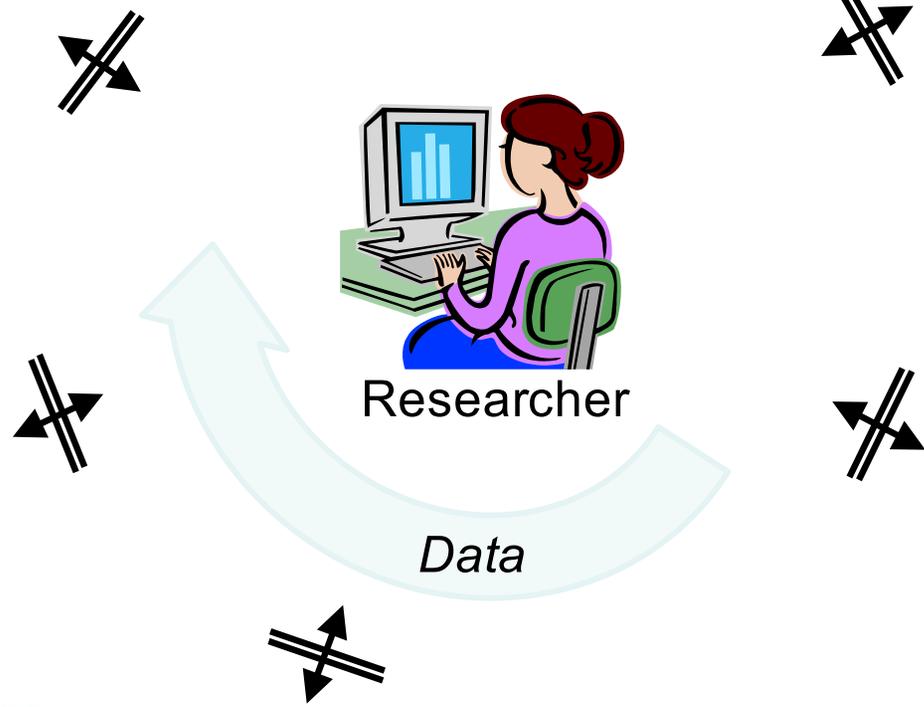
National Computing Resources



Cloud Services

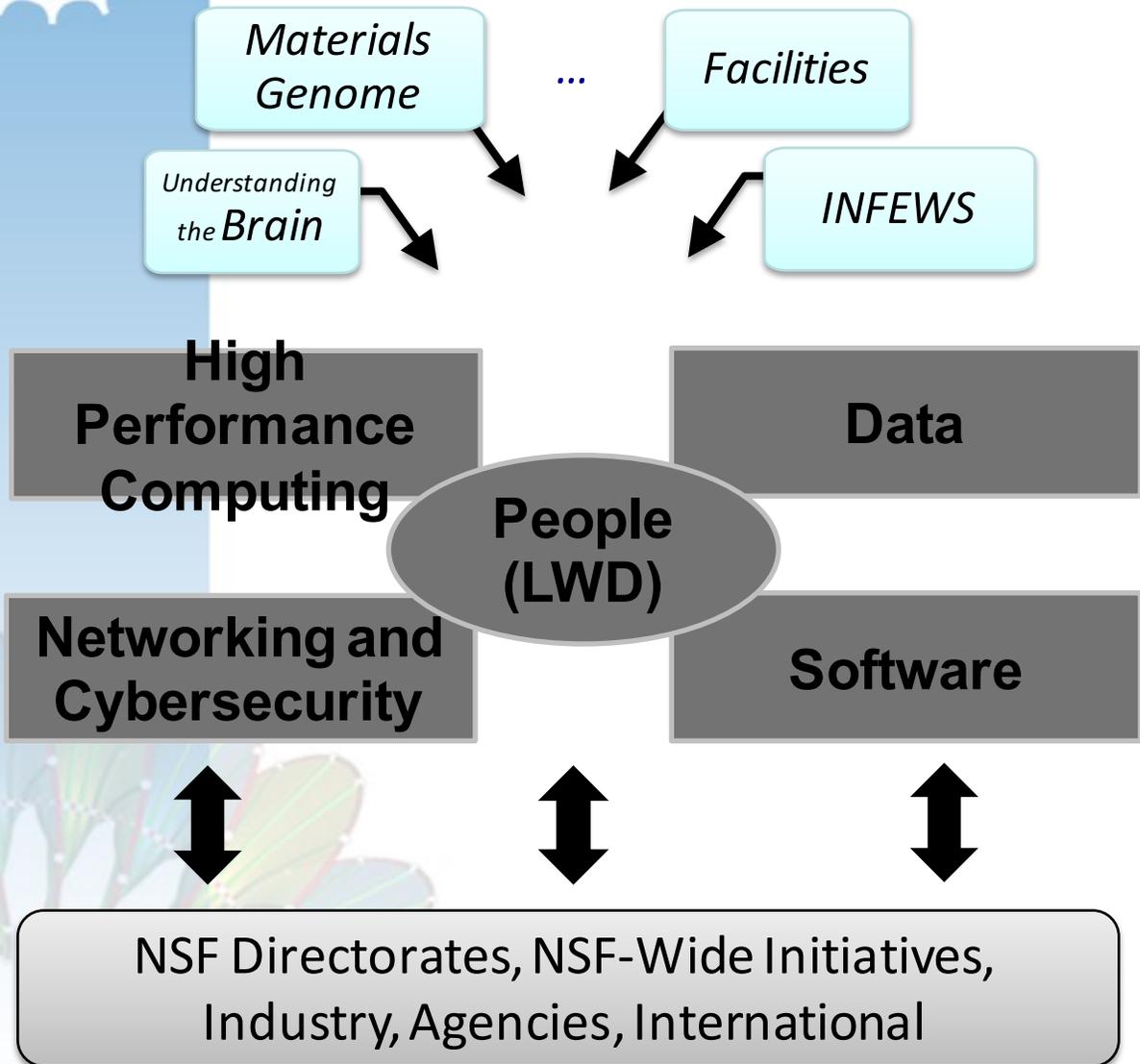


Identities?  
Resources?  
Persistence?



# ACI: Operational View

*Supporting advance CI to accelerate discovery and innovation*



## **Science Drivers**

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

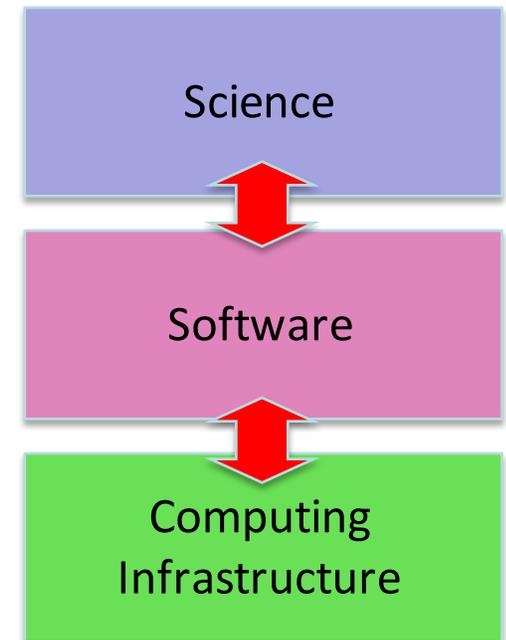
## **ACI investments**

*Convergent investments in technologies and communities to maximize impact*

**Leadership,  
Coordination,  
Partnership**

# Software

- Software (including services) essential for the bulk of science
  - Anecdotally (i.e. Dan Katz says!), ~half the papers in recent issues of Science were software-intensive
  - Research becoming dependent upon advances in software
  - Wide range of software types: system, applications, modeling, gateways, analysis, algorithms, middleware, libraries
  - Significant software-intensive projects across NSF: e.g. NEON, OOI, NEES, NCN, iPlant, etc
- Software is not a one-time effort, it must be sustained
  - Development, production, and **maintenance** are people intensive
  - Software life-times are long vs hardware
  - Software has under-appreciated value





# OVERVIEW OF SOFTWARE PROGRAMS



# ACI's Software Mission

ACI's software programs will

- *seed*
- innovative software-intensive science ecosystems, and
- *shepherd* the ecosystems
- to *sustainably*
- *advancing* science and engineering
- in a *reliable*, and *robust* manner
- through *solicitations* and other *funding* mechanisms, and
- *community engagement* and *outreach*.



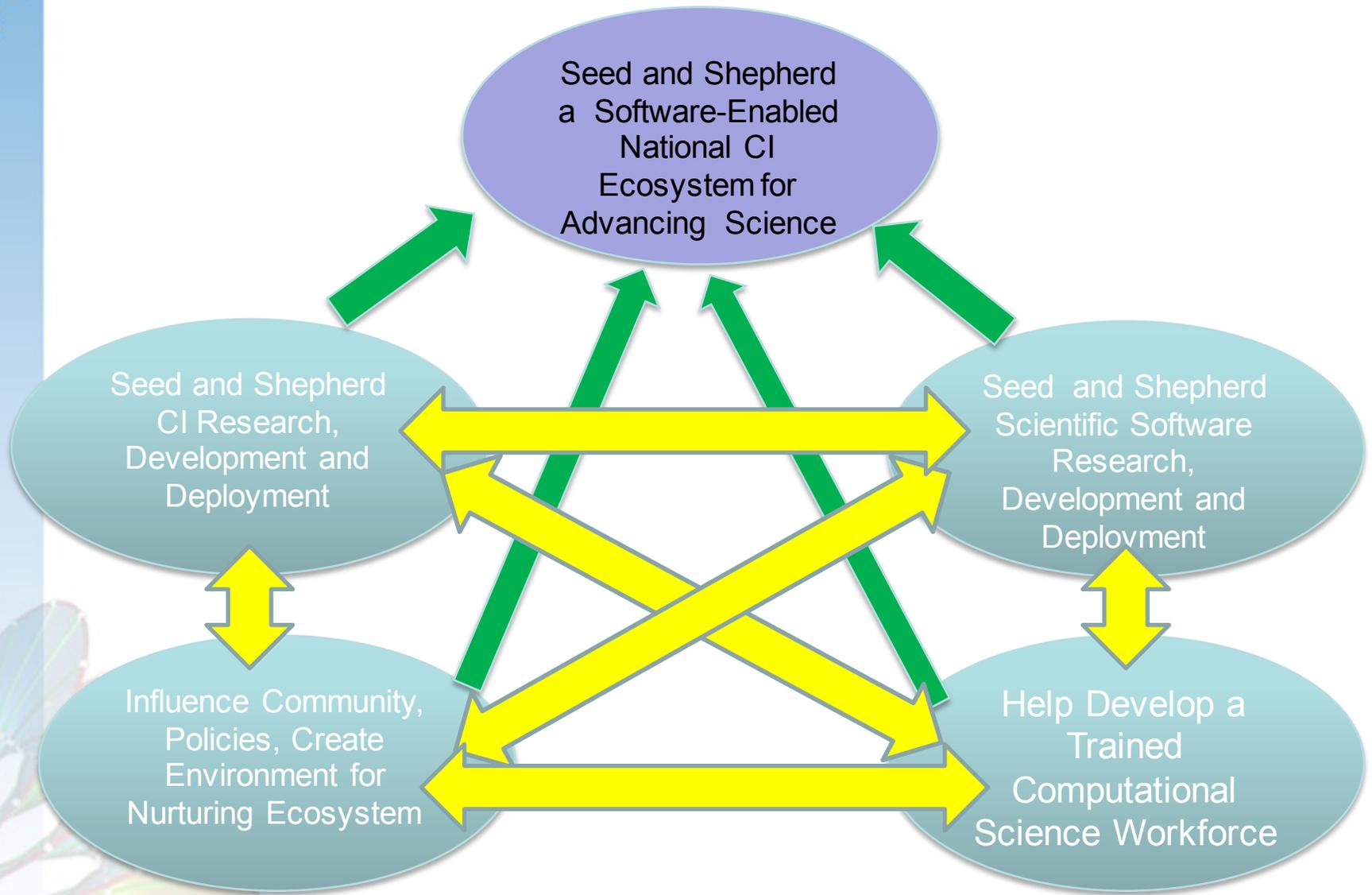
## Vision

ACI's software programs will be the

- *primary*
- *catalyst* of
- the **software components** of  
a
- *national, science cyber  
infrastructure ecosystem.*



# ACI Software Program Strategy





# Overview of ACI-relevant Software Programs - I

## • (Use-Inspired) Research in Software for Science Domains

- Computational and Data-Enabled Science & Engineering (CDS&E)
  - Virtual program for science-specific proofing of algorithms and tools
  - ENG, MPS, ACI; GEO (& BIO informally)
- Critical Resilient Infrastructure Systems and Processes (CRISP)
  - New interdisciplinary approaches for modeling, design, and operation of interdependent critical infrastructures for resilience and adaptability
  - CISE CPS, CNS, ACI. ENG, SBE
- Designing Material to Revolutionize and Engineer our Future (DMREF)
  - MGI, proposals must create iterative process of theory -> simulation -> experiments -> theory.
  - MPS, ENG, ACI

# Overview of ACI Software Programs - 2

- 
- Use-inspired Research in CI (non-lead)
    - Exploiting Parallelism and Scalability (XPS)
    - Scalable Parallelism in the Extreme (SPX)
      - “Out there” ideas for extreme scale
  - Embedded CI Research, Development and Deployment (lead)
    - Software Infrastructure for Sustained Innovation (SI2)
      - Transform innovations into sustained software CI resources
      - **Embedded innovation and empirical research in “practice”**
      - Includes all NSF directorates
  - Opportunistic Investments:
    - Venture/Reuse Fund
    - EAGERs
    - Workshops
    - Supplements

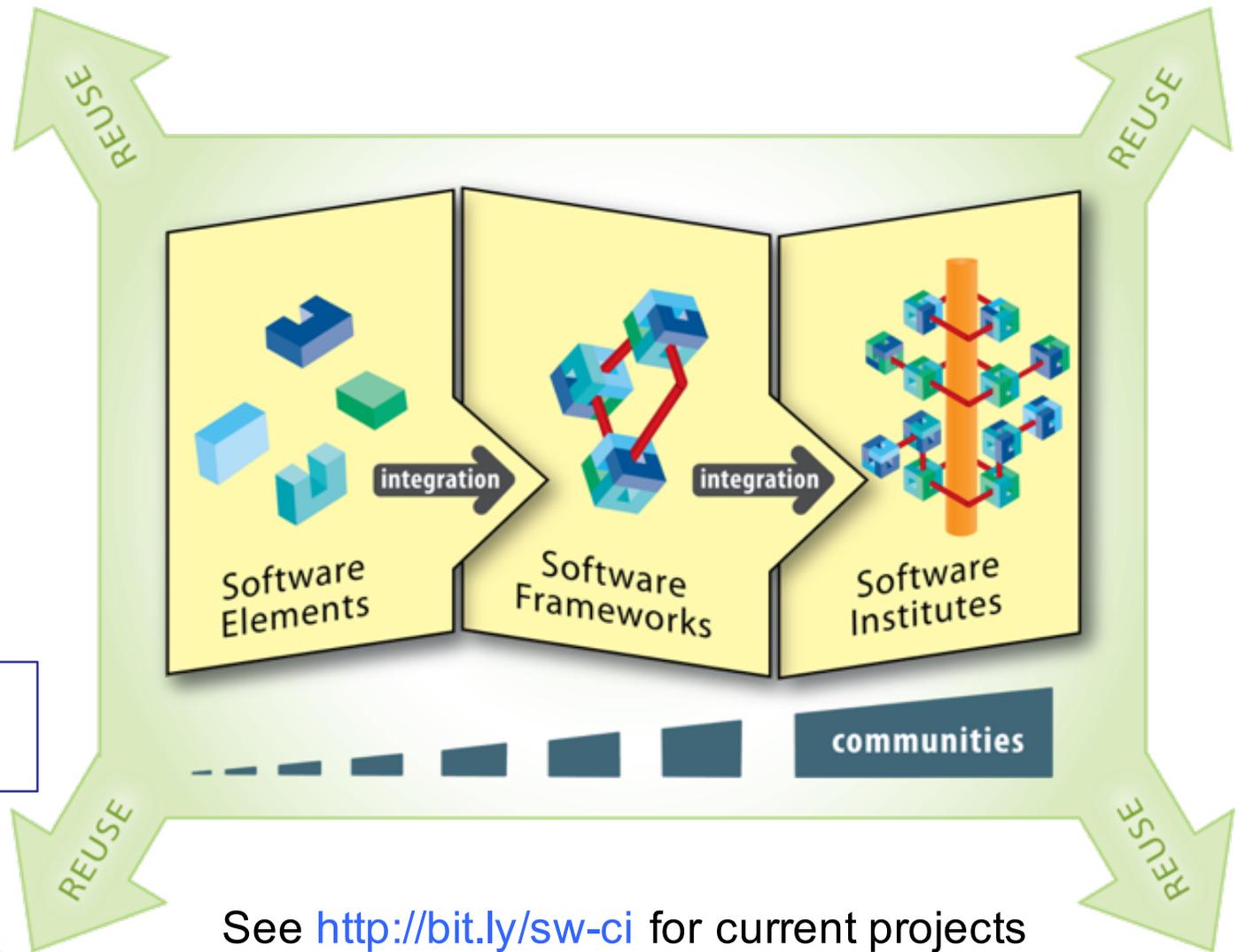
# Flagship: Software Infrastructure for Sustained Innovation (SI2)



Elements: \$500K/3 years

Frameworks: \$1M/year 3-5 years

Institutes: \$3-\$5m/year 5-10 years



See <http://bit.ly/sw-ci> for current projects



# Unique Criteria

- Fill a **recognized need** (in the science community)
- Create **innovative**, robust and reliable research **capabilities** in science and engineering for typically **NSF-funded researchers**
- **Embed research** and innovation into the project activities
- Use a **comprehensive** user-engaged, dynamic, software process, **management**, w/ operational and successes metrics
- Resourced by **teams** with **credibility** in **software** practice, **and science**
- **Build community** through direct engagement
- Progress towards **sustainability** after NSF funding has ended
- Further a **national CI ecosystem** (reuse, integrate, adopt)



# 2017 Priorities

- A National CI Ecosystem
  - Community establishment (directive to leverage Institutes)
  - Sustainability
  - Leveraging existing assets
  - Towards an infrastructure “platform”
- Enabling Robust and Reliable Science
  - Repeatability -> Replicability -> Reproducibility
  - Uncertainty quantification
  - Software publication, citation
  - Education
- Internal operational improvement
  - Portfolio analysis
  - Expansion of software programs to new areas
  - Process improvement
  - Alignment with strategic goals; less ad hoc funding
  - Systematic community outreach – talks, newsletter, PI meetings



# 2017 – NSCI Executive Order 13702

## EXECUTIVE ORDER

-----

### CREATING A NATIONAL STRATEGIC COMPUTING INITIATIVE

By the authority vested in me as President by the Constitution and the laws of the United States of America, and to maximize benefits of high-performance computing (HPC) research, development, and deployment, it is hereby ordered as follows:

The NSCI is a whole-of-government effort designed to create a cohesive, multi-agency strategic vision and Federal investment strategy, executed in collaboration with industry and academia, to maximize the benefits of HPC for the United States.



# NSCI Objectives of NSF Interest

1. Computing at exascale (reduce time to science by 100x)
2. Coherence between modeling and simulation and data analytic computing.
3. Systems that go beyond Moore's Law
4. An enduring national ecosystem
5. Public-private collaborations (as appropriate for broader impact and sustainability).



# NSCI at NSF

Driven by Science Frontiers

Identified by  
science  
domain  
across NSF

100x improvement in  
“time to science”

Example of science frontiers:

- Mapping the human brain with synaptic resolution
- Optimal decision making in complex high-dimensional problems
- Understanding multi-phase turbulent flow
- Device and systems technology enabling new compute paradigms
- Extreme computing and data analysis for experimental and observational facilities
- Analyzing social behavior from large-scale data collections
- Understanding high complexity in biological systems



# **SOFTWARE INSTITUTES!**



# What is a Software Institute? What does it do?

1. Enables transformative science through software
2. Long-term community hub of software infrastructure and technologies
3. Serving a research community of substantial size and disciplinary breadth
4. Provides pathways for community involvement, including industry
5. Establishes community processes for software development and sustainability
6. Establishes community structures and mechanisms for support, outreach, and workforce development
7. Helps establish community research directions, standards and models, and collaborative development activities.
8. Enables software elements developed within the community to be transitioned to conform to community software frameworks, standards and processes,
9. Leverages existing infrastructure investments.



# Evaluation Criteria

- Is the Institute enabling new science and broader impacts?
- Is the Institute demonstrating success in addressing the issues it identified to motivate its formation?
- Has the Institute responded to changes in the issues of concern from the community?
- Is the Institute achieving a consensus among the software developers in the Institute's topic area?
- Are the software products available in the Institute's topic area being integrated or made interoperable, including, but not limited to the products from NSF-supported teams working in its area?
- Is the Institute achieving broad buy-in among the software user community in the Institute's topic area?
- Has the Institute created and begun to implement a sustainability plan that is realistic and likely to be successful?

**An institute would specialize these success metrics**



# Software Institute Case Studies

## The Science Gateway Community Institute (SGCI)

- University of California, San Diego/San Diego Supercomputer Center
- Texas Advanced Computing Center at the The University of Texas at Austin
- University of Michigan
- Indiana University.
- Purdue University
- Elizabeth City State University.
- University of Notre Dame

## The Molecular Science Software Institute (MolSSI)

- Virginia Tech
- Rice University
- Stony Brook University.
- University of California, Berkeley
- Stanford University,
- Rutgers University,
- University of Southern California
- Iowa State University

# Why a Science Gateway Community Institute?



NSG Portal

The Neuroscience  
A Portal for Computatio

DiaGrid

The Materials Project

The logo for The Materials Project, featuring a dark blue background with a hexagonal grid pattern and three geometric shapes (a triangle, a diamond, and a cube) in the upper left corner.

System Grid

Structural Prote

Virtual Laboratory  
and Planetary M

134,886 ... to date  
No Form ... access

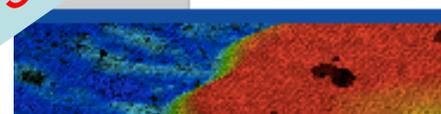
A screenshot of the DiaGrid website showing statistics and a brief description.

Computational Anatomy Portal  
Brain || Heart



CIG Seis

Involving left ventricle  
Generating representative shape models (ES)

A screenshot of the Computational Anatomy Portal interface showing various anatomical models and data.

c Modeling Portal

CIPRES  
GATEWAY

Collaborative™ Empowering A New Plant Biology

Integrat

ics &  
e collab  
ersity

Welcome to ROSIE  
Rosetta Online Server that Includes Everyone

Science Gateways Abound  
These are some that use XSEDE supercomputers

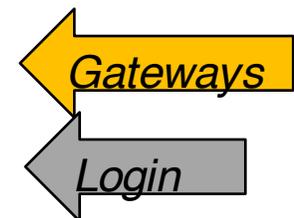
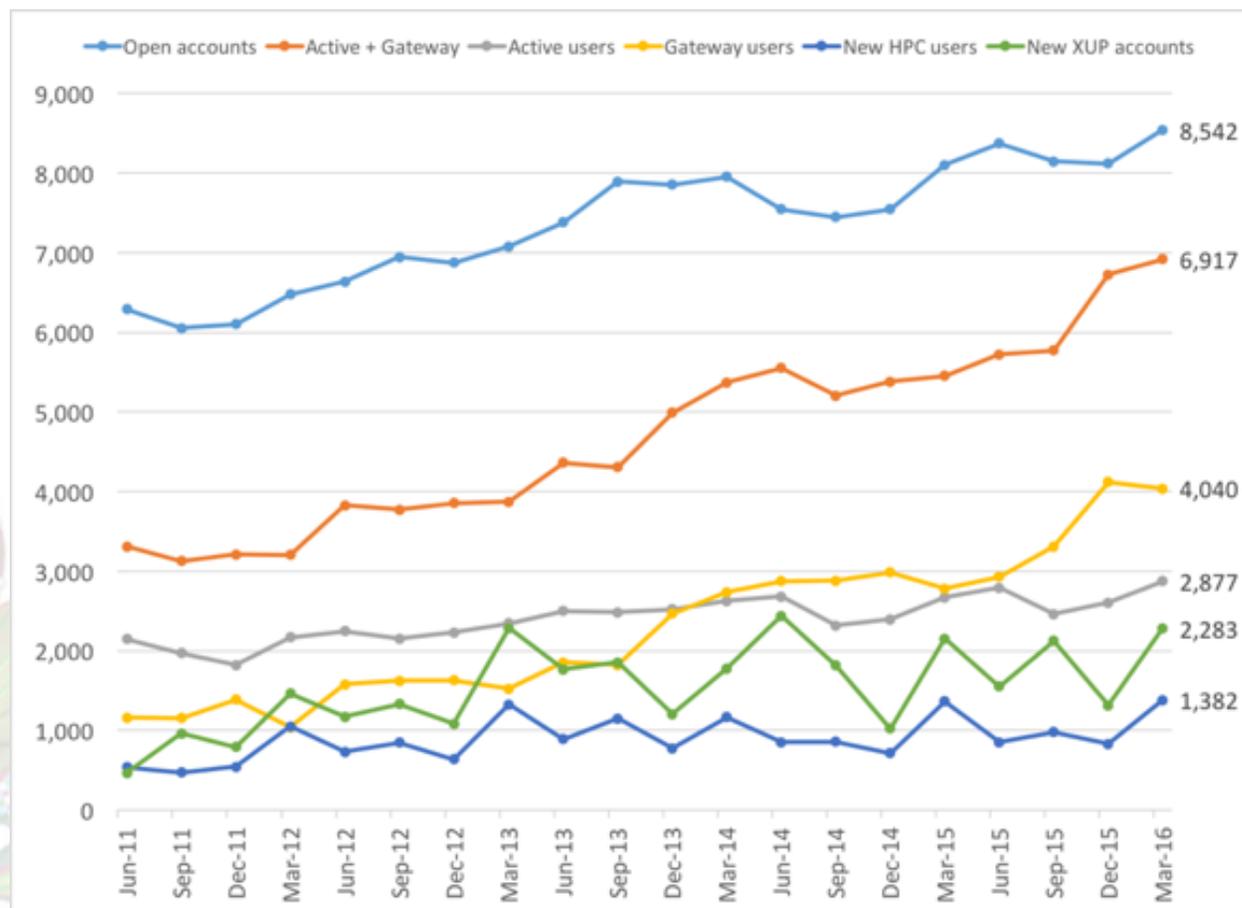
A large, diagonal banner with a light blue background and red text, stating 'Science Gateways Abound' and 'These are some that use XSEDE supercomputers'.

Acknowledgement: Nancy Wilkins-Diehr, SGCI



# Why? Gateway Usage Growth

Gateway usage surpasses command line users in XSEDE

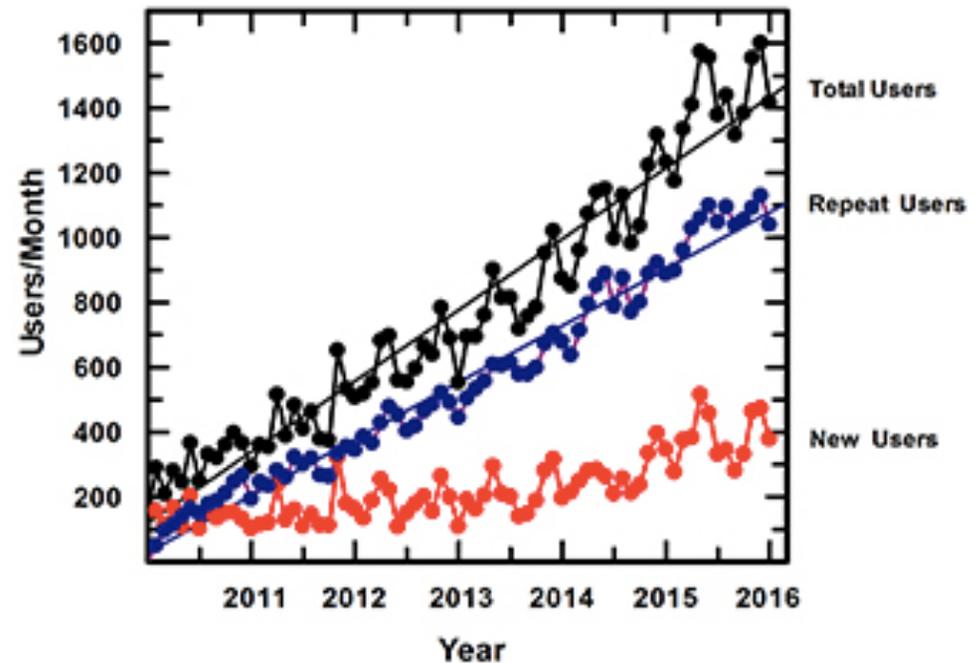


Acknowledgement: Nancy Wilkins-Diehr, SGCI



# Vision – Prototypical Example (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



PI Mark Miller, SDSC, [www.phylo.org](http://www.phylo.org)



# Science Gateways Software Institute

Press Release 16-088

## NSF commits \$35 million to improve scientific software

### **Science Gateways** Community Institute

The second award, led by the University of California, San Diego, establishes the **Science Gateways** Community Institute, a multi-institutional consortium that will increase the capabilities, number and sustainability of **science gateways**. Gateways are mobile or web-based applications that provide broad access to the nation's shared cyberinfrastructure to scientists and citizens alike.

"Gateways foster collaborations and the exchange of ideas among researchers and can democratize access, providing broad access to resources sometimes unavailable to those who are not at leading research institutions," said Nancy Wilkins-Diehr, associate director of the San Diego Supercomputer Center and principal investigator for the project. "Sharing expertise about basic infrastructure allows developers to concentrate on the novel, the challenging, and the cutting-edge development needed by their specific user community."



# A successful gateway institute will provide leadership to

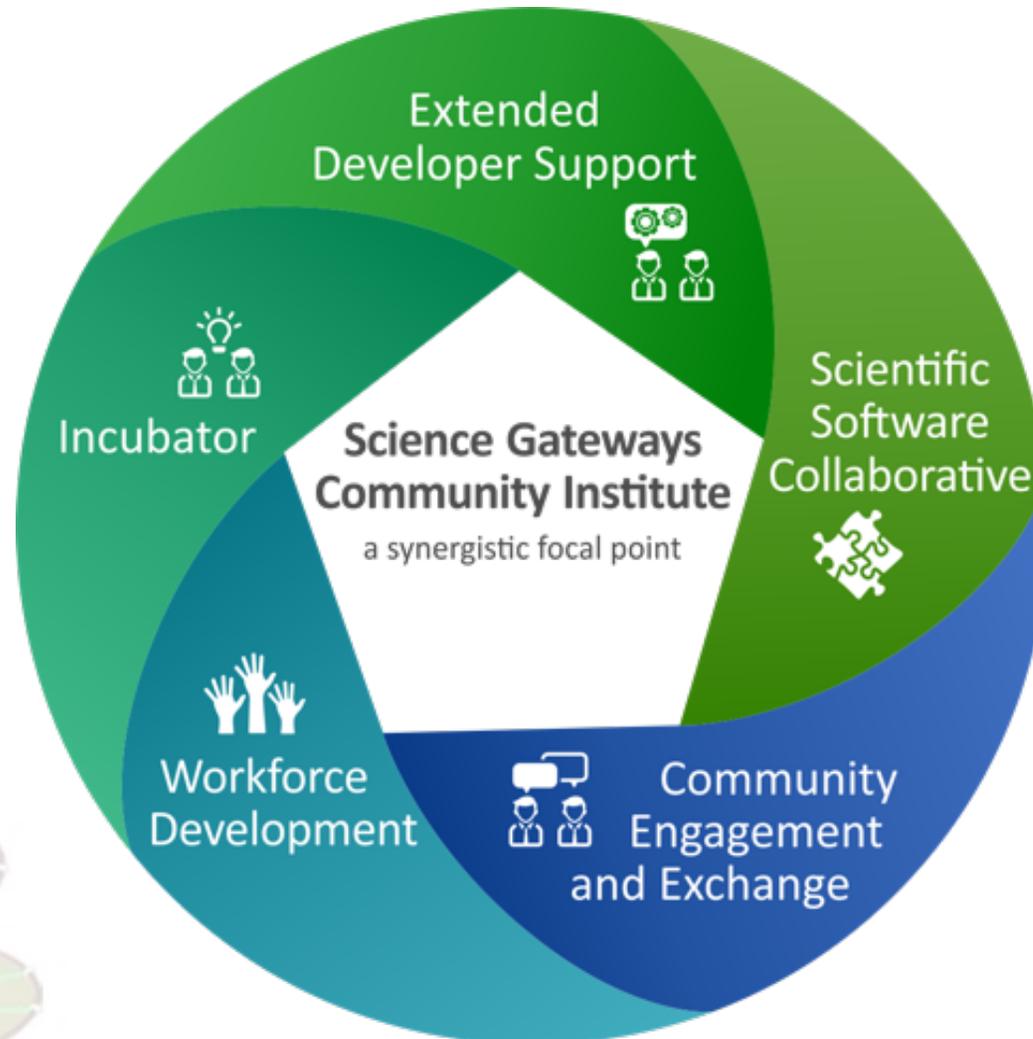
- 1) *bring science gateway developers together with each other and with the developers and operators of existing and potential cyberinfrastructure elements that science gateways integrate and enable the use of*
  - *in order to promote the efficient, effective, and sustainable development of scientific web and mobile interfaces*
- 2) *educate developers and the next generation of investigators to effectively use the gateway software ecosystem to solve real research problems; and*
- 3) *educate the next generation of researchers to enable them to create the software cyberinfrastructure required to both advance fundamental understanding of science gateways and enable researchers to address the grand challenge problems of the future*

Acknowledgement: Nancy Wilkins-Diehr, SGCI



# Science Gateways Community Institute

Est. Aug, 2016



Acknowledgement: Nancy Wilkins-Diehr, SGCI



# Vision for SGCI success

5-10 years from now

- Science gateways form a vibrant community
  - Inter-agency, **international**, collegial
- Creating gateways is easier
  - Created with more thoughtfulness, so they are more sustainable
- Gateway developers have stable career paths
  - More efficient environments on campuses
- Students are excited to stay in the sciences
- Transformative changes in how science is done

Acknowledgement: Nancy Wilkins-Diehr, SGCI



# The Molecular Sciences Software Institute (MoSSI) – Why?

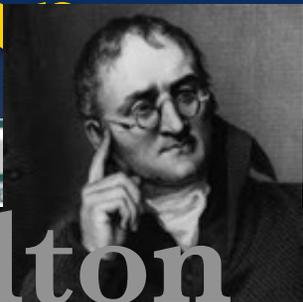
- Community interest in serving and enhancing the software development efforts of the broad field of computational molecular science.
- Collaborative effort by a consortium of universities
- Joint support from numerous NSF domain science divisions: Chemistry (CHE), Division of Materials Research (DMR), Office of Multidisciplinary Activities (OMA)



# Why? Large Community Software Base



Gaussian



Dalton



Amber

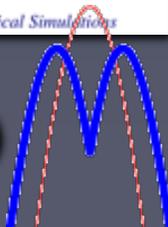


ORCA



TURBOMOLE

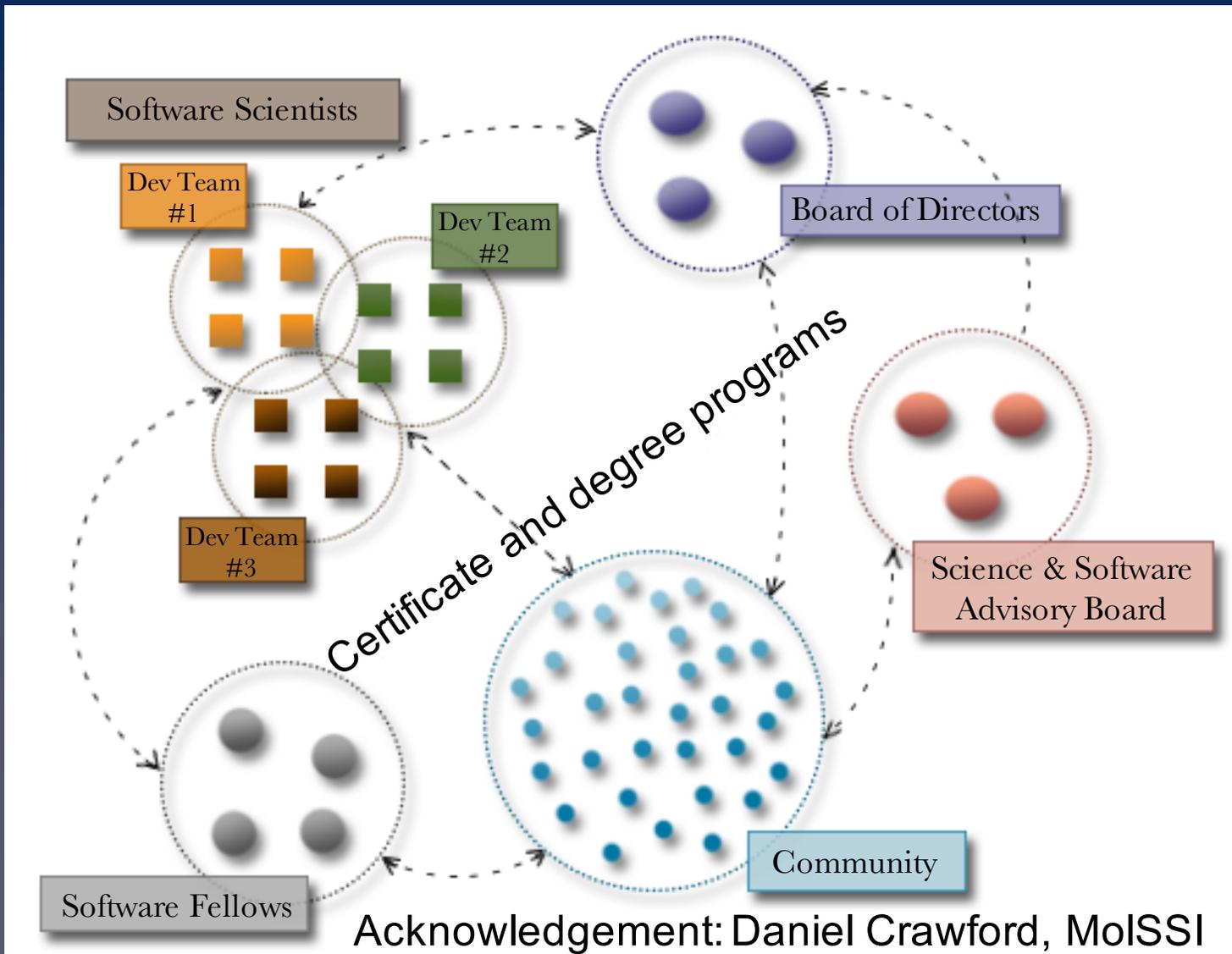
MOLPRO



Acknowledgement: Daniel Crawford, MolSSI



# MolSSI Structure





# Guidance

- Science Drivers
- Leadership, management, advisory
- Community
- Software
- Structure and Processes
  - SDLC, Engagement, Learning and Workforce, Diversity, Change Management
- Sustainability
- Metrics

**Overarching: Baseline and progression**



# OTHER ACI PROGRAMS



# Upcoming Solicitations

- Software Infrastructure for Sustained Innovation (SSE, SSI, S2I2) Software Elements, Frameworks and Institute Conceptualizations - NSF 17-526
  - Webinar- 2017 SI2 Updates December 13, 2016 3:00 PM to 4:00 PM
- CyberTraining - Training-based Workforce Development for Advanced Cyberinfrastructure - NSF 17-507
- Cybersecurity Innovation for Cyberinfrastructure (CICI) - NSF 17-528
- Data Infrastructure Building Blocks (DIBBs) - NSF 17-500

## The DIBBs Solicitation [awalton@nsf.gov](mailto:awalton@nsf.gov)



- Seeks proposals that develop robust, scalable, well-designed cyberinfrastructure (the 'building blocks') contributing to future discovery and innovation across disciplines
  - Guided by science and engineering research priorities
  - Built upon recognized community data collections
  - Result in clear, tangible cyberinfrastructure products
  - Implemented through collaborations between cyberinfrastructure experts and specific science and engineering research communities
- Focus areas:
  - Early Implementation Awards: up to 6 awards, each up to \$4M total for up to 5 years
  - Pilot Demonstrations: up to 5 awards, each up to \$500K total for up to 3 years



# Cybersecurity Innovation for Cyberinfrastructure (CICI) NSF 16-533 - ANIKOLIC@nsf.gov

Activities that impact the security of science, engineering and education environments

Target community is operational cyberinfrastructure

- \$7M available. Usually 7 – 9 awards.
- 2 Areas:
  - Secure and Resilient Architecture - \$1M awards
  - Regional Cybersecurity Collaboration - \$500K awards
- Sample Topic Areas of Interest:
  - Secure interoperability of cloud/campus
  - Security Metrics
  - DDoS Defense
  - Shared assessments and security design reviews





# XD Network of Computational Resources and Services - REIGENMA@nsf.gov



**Yellowstone Geosciences**

**Blue Waters Leadership Class**



**Open Science Grid**  
High throughput

**Blacklight**  
Shared Memory  
4k Xeon cores

**Bridges (2016)**  
large, coherent shared-memory



**Comet (new)**  
"Long Tail Science"  
47k cores/2 PF  
High throughput

**Gordon**  
Data intensive  
64 TB memory  
300 TB Flash Mem



**Stampede**  
460K cores  
w. Xeon Phi  
>1000 users

**Wrangler (new)**  
Data Analytics

**Maverick**  
Visualization  
Data Analytics

**Jetstream (2016)**  
Cloud-based

**SuperMIC**  
380 nodes – 1PF  
(Ivy bridge, Xeon Phi, GPU)



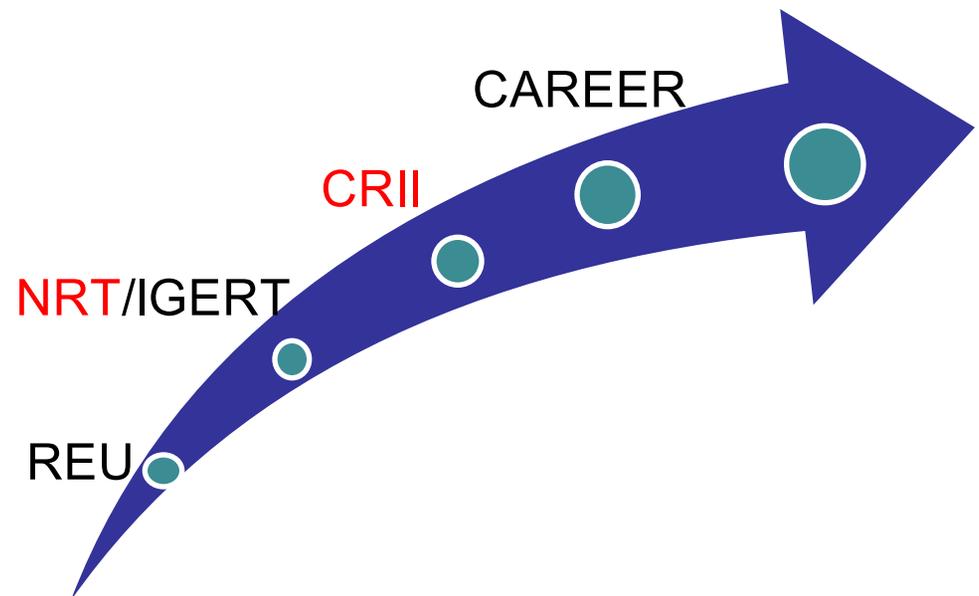
**Darter**  
24k cores

- Coordination through XSEDE**
- Resource Allocation
  - Advanced User Support
  - Education and Outreach
  - Common Software



## LWD: The Career Pipeline – [sprasad@nsf.gov](mailto:sprasad@nsf.gov)

- 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





# QUESTIONS?

Rajiv Ramnath [rramnath@nsf.gov](mailto:rramnath@nsf.gov) Vipin Chaudhary [vipchaud@nsf.gov](mailto:vipchaud@nsf.gov)