



NSF support for Cyberinfrastructure: Opportunities for Researchers

Evelyn M. Goldfield, NSF/CHE
Ohio Supercomputer Center
Dec 4, 2014





Cyberinfrastructure for the 21st Century (CIF2) : a framework for CI

<http://www.nsf.gov/cise/aci/taskforces/index.jsp>

- Developed in 2009 by OCI, now ACI with a great deal of input from the community
- Forms the basis for much activity at NSF on issues of CI.

**Campus
Bridging**

**Data and
Visualization**

**High
Performance
Computing**

**Workforce
Development**

**Scientific
Software**

**Grand
Challenges**

CDS&E



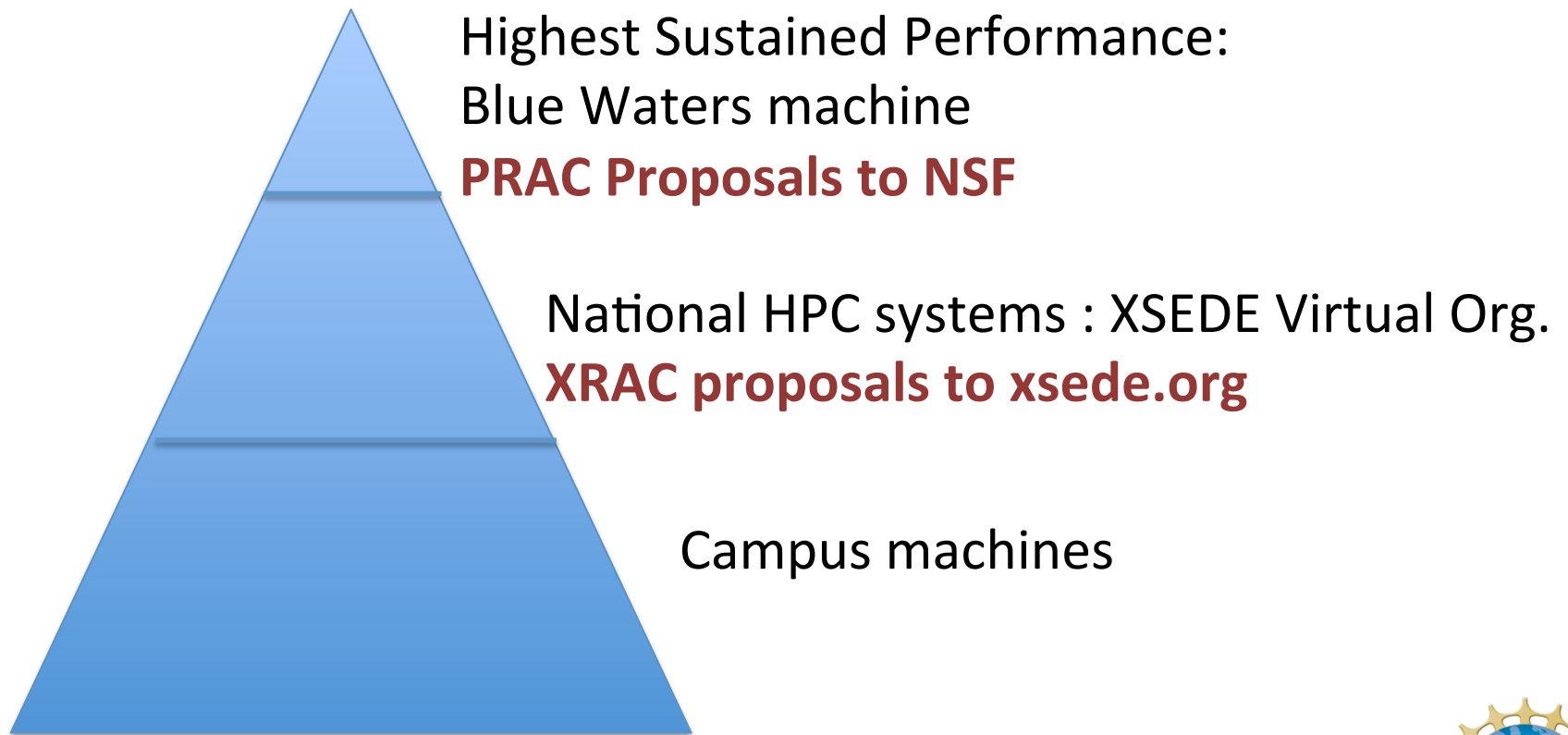
High Performance Computing

NSF supported HPC resources

XSEDE

Blue Waters

Key Enabler of Science and Engineering: Community Access to Computational Capabilities



NSF Supported National Computational Resources

increasingly diverse and collaborative

Trestles

IO-intensive
10k cores
160 GB SSD/Flash

Gordon

Data intensive
64 TB memory
300 TB Flash Mem



Comet

"Long Tail Science"
47k cores/2 PF
High throughput

To be deployed in FY2015
with XSEDE integration and
user services

Yellowstone Geosciences



Blue Waters Leadership Class



Wrangler

Date Analytics

Stampede

460K cores
w. Xeon Phi
>1000 users
Upgrade in 2015

Maverick

Visualization
Data Analytics

Lonestar*

Large memory

Open Science Grid

High throughput

FutureGrid*

CS testbed

Blacklight

Shared Memory
4k Xeon cores



Darter

24k cores

Nautilus

Visualization
Data Analytics

Keeneland

CPU/GPGPU



SuperMIC

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

XSEDE integration

** To be retired in 2014*

XSEDE: Supporting collaborative expertise, services, education and outreach in an overall ecosystem



XSEDE is a single virtual system that scientists can use to interactively share computing resources, data and expertise.



XSEDE

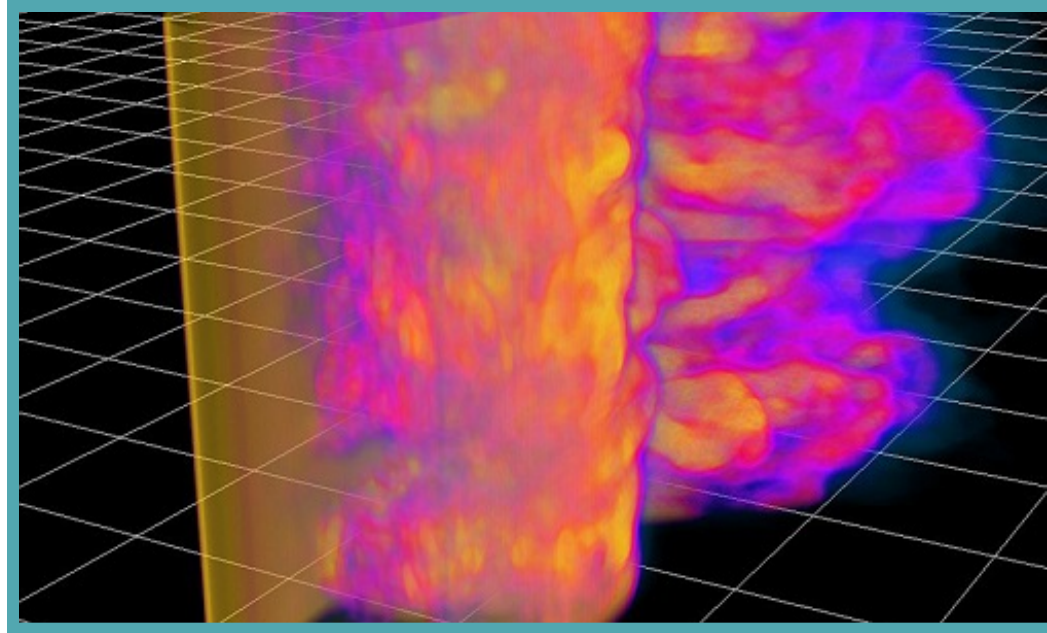
- User services (extensive)
- Resources
- Education and Outreach
- Technology Data Base
- Science Gateways
- It is easy to obtain a start-up account, you do not have to have an NSF award



We would like to make you aware of computing and data resources and services that are available to you at no cost via the NSF-funded XSEDE project. One of the services offered is Extended Collaborative Support, including help with code porting, computational improvements, and data analysis. An initial allocation of resources and/or services is available by submitting the Project Summary for your award along with a 1-2 paragraph description of your needs. More information is available at <https://www.xsede.org/using-xsede>



Blue Waters: one of the most power supercomputers in the world



- At NCSA, supported by NSF and the University of Illinois



Blue Waters

- Designed for sustained petascale performance on several 100,000 cores for a broad class of problems
- Supports visualization of large datasets produced by computations
- More info:
bluewaters.ncsa.illinois.edu/hardware-summary



PRAC: proposal to NSF required to obtain BW allocation

- Next Deadline : 11/18/2015
- Several years of preparation may be needed to get ready to exploit a sustained petaflop system.
- Present a development plan, such as:
 - Code scalability today
 - Year 1: Optimizations for scalability to part of machine
 - Year 2: Further optimizations for full scalability

The Blue Waters team will provide consultation and support.



Performance Optimization

Optimizing your code on Blue Waters may entail:

- Discovering and exploiting parallelism within codes
- Overlapping different types of operations
- Exploiting multi-level caches, local/remote memory
- Orchestrating intra/inter-node communication
- Performing parallel I/O
- Exploiting heterogeneous processors
- Dealing with petabyte-size memory-resident data

You need to show you have the necessary expertise



SI² Software Infrastructure for Sustained Innovation

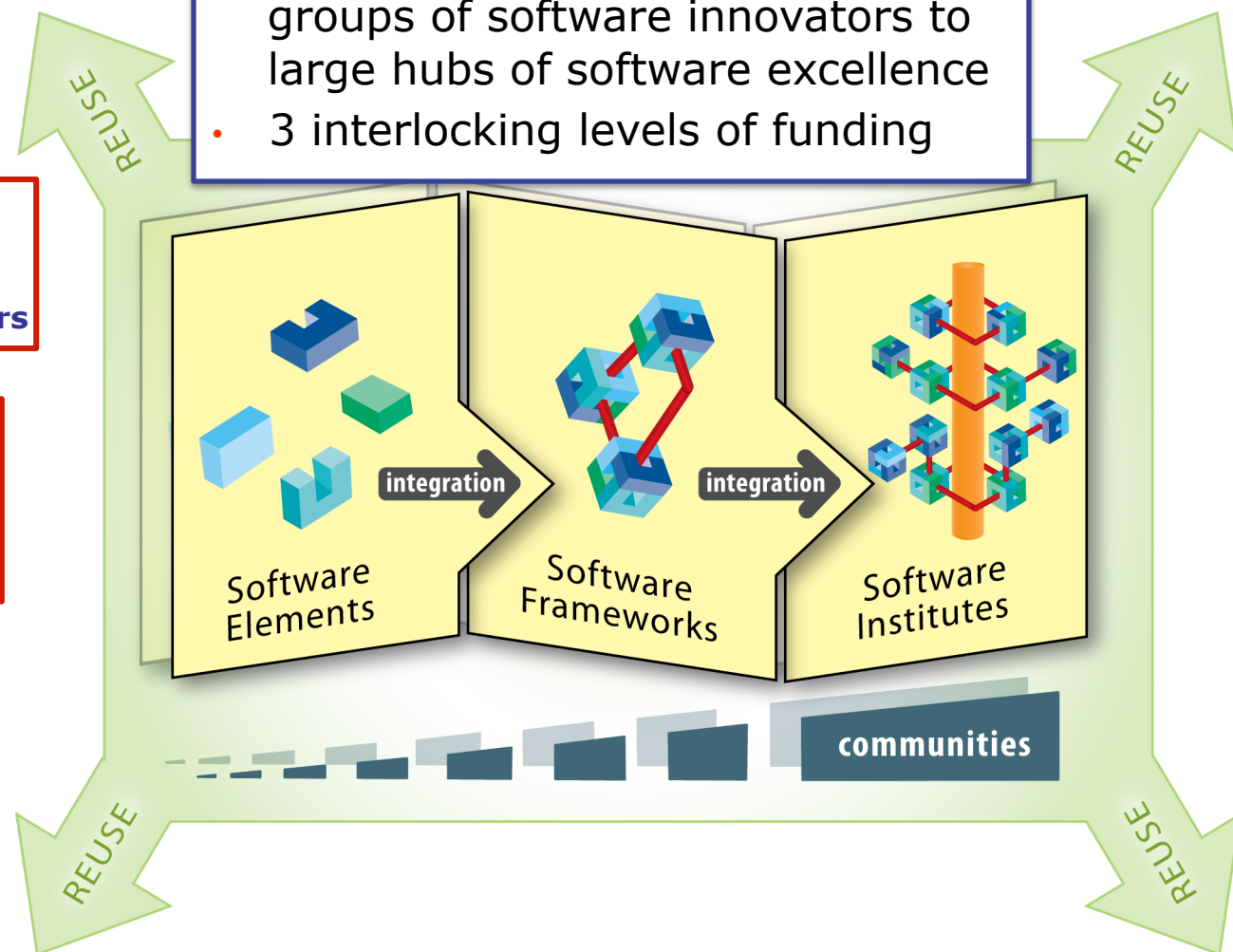
- Create a software ecosystem that scales from individual or small groups of software innovators to large hubs of software excellence
- 3 interlocking levels of funding

Scientific Software Elements (SSE)

1-2 PIs, <\$0.5M, 3 years

Scientific Software Integration (SSI) Focused Groups

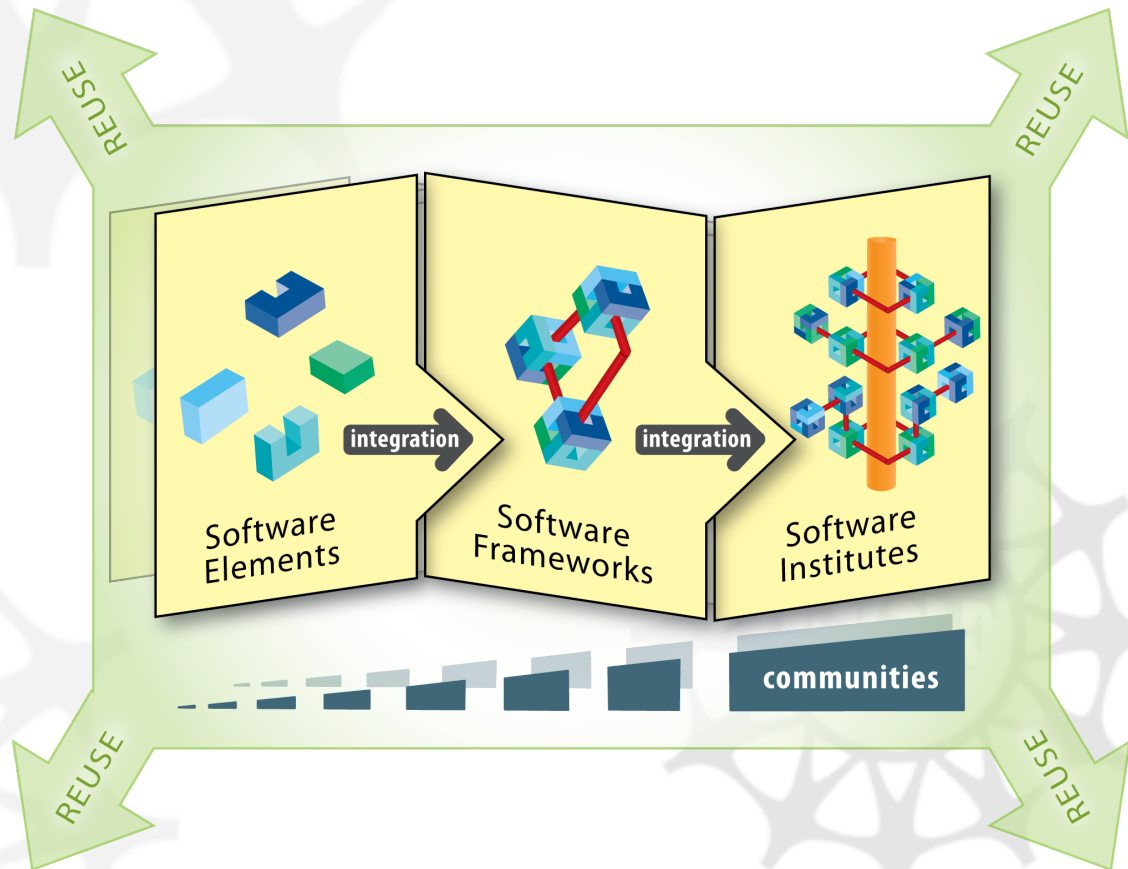
\$200K-\$1M per year,
3-5 years





Software Infrastructure for Sustained Innovation (SI²)

- SI² recognizes that software is the integral enabler of computation, experiment and theory, and a key part of the cyberinfrastructure that enables science and engineering research to achieve new discoveries.
- SI² encourages PIs to provide software that
 - Has a purpose (enables specific science and engineering research)
 - Has a user community that will be engaged in the project
 - May be used by a wider audience
 - Fits with other infrastructure, including other software





SI² Program: Software Infrastructure for Sustained Innovation

- **NSF program provides a framework for software development and support to advance NSF research in science and engineering**
- **Priorities for the program:**
 - **Sustainable, robust, and reliable software**
 - **Pathways to include innovation in software**
 - **Software engineering processes that work for different communities**
- **Due dates:**
 - **SSE: Jan 19-Feb 2, 2015**
- **SSI: June 12 – June 26, 2015**



Data Infrastructure Building Blocks (DIBBs)

Building cyberinfrastructure to enable advances in science and engineering research, “blocks at a time”

- **Emphasizes cyberinfrastructure investments:**
 - **guided by science and engineering research priorities**
 - **built upon recognized community data collections**
- **Promotes ‘pilot’ activities in three categories**
 - **Small collaboration-focused projects, including prototypes (short duration)**
 - **Intermediate-scale pilots (typical duration)**
 - **At-scale early implementation pilots (possibly long duration)**
- **Collaborates with specific science and engineering research communities, to ensure continuing relevance**



Computational and Data-enabled Science and Engineering (CDS&E)

- Science disciplines are both leading consumers and hard drivers of cyber-capability: their needs force, & their research creates, breakthroughs – in algorithms, in simulation & modeling methods.
- **CDS&E is a cross-directorate program involving MPS, ENG, and CISE/ACI “to identify and capitalize on opportunities for major scientific and engineering breakthroughs through new computational and data analysis approaches”**

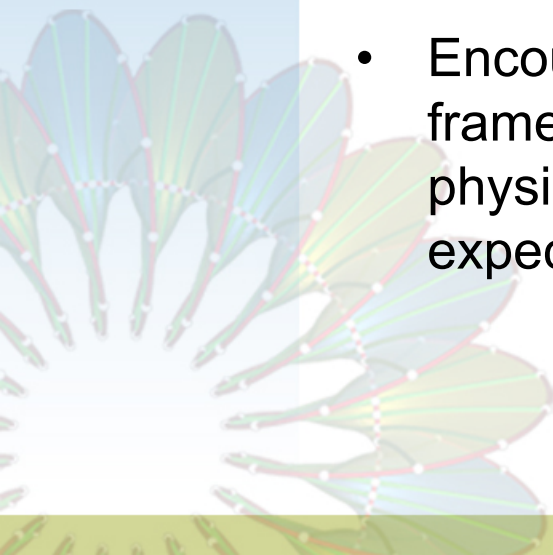


- Support for development, adaptation, or utilization of the capabilities offered by advancing both research and infrastructure in computation and data handling
- A “meta-program” – submit through pre-existing funding opportunities – see announcement PD 12-8084



Computational and Data-Enabled Science and Engineering (CDS&E)

- Promotes the creation, development, and application of the next generation of mathematical, computational and statistical theories and tools.
- Encourages adventurous ideas that generate new paradigms and that create and apply novel techniques, generating and utilizing digital data in innovative ways
- Encourages ideas at the interface between scientific frameworks, computing capability, measurements and physical systems that enable advances well beyond the expected natural progression of individual activities





CDS&E

- Bottoms up, cross-cutting program
- All MPS divisions (AST, CHE, DMR, DMS, PHY), ENG, ACI
- Informal participation: GEO, BIO, CISE not ACI
- Proposals are submitted to identified **programs within divisions**, **not required to be inter-disciplinary**
- All proposals identified as CDS&E are discussed by the CDS&E working group for possible cofunding (**many are inter-disciplinary**)
- About half of the CDS&E funded proposals are co-funded





3 Routes to CDS&E

- Through funding Opportunity: PD 12-8084
- Label your proposal CDS&E and submit to any participating division
- Program Director identified





to find CDS&E awards

- Search the NSF awards data base in a number of ways
- Search on PE: 8084
- Search on PRC: 8084
- Limit the search using the appropriate NSF organizations





The Future????

Future Directions for NSF Advanced Computing Infrastructure to Support U.S. Science and Engineering in 2017-2020

William D. Gropp and Robert J. Harrison
committee co-chairs





Preliminary Report : Issues and Questions

1. Responding to need for integrated scientific discovery.
2. Responding to technical challenges: including disruptive architectures
3. Responding to growing demand





Your Input is desired!!!!

Most useful if submitted by Jan

To submit comments and download/read the interim report:

<http://nas.edu/sciencecomputing>

To follow study activities:

<http://cstb.org>

or email study director Jon Eisenberg,
jeisenbe@nas.edu