

A Brief Overview of the Software Development for CI and Strategic Technology for CI Programs

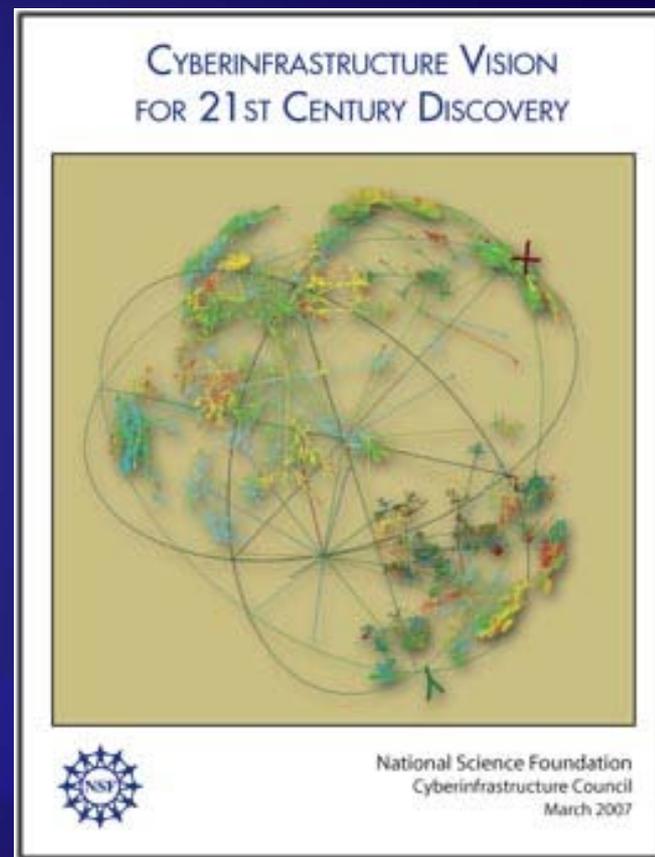
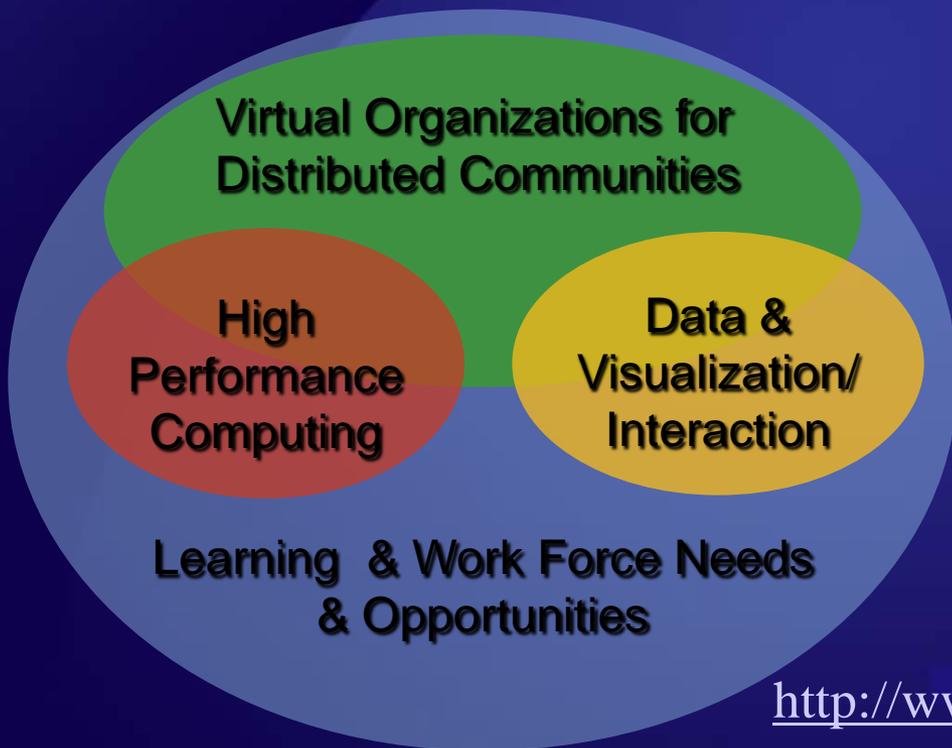
Dr. Jennifer M. Schopf
Office of CyberInfrastructure
National Science Foundation





NSF Vision for Cyberinfrastructure

- ❖ “National-level, integrated system of hardware, software, data resources & services... to enable new paradigms of science”





Office of CyberInfrastructure Focus Areas

- ❖ Software
- ❖ Campus Bridging and Networking
- ❖ Data and Visualization
- ❖ High Performance Computing
 - including Clouds, Grids, etc.
- ❖ Grand Challenges and VOs
- ❖ Education/Workforce development



STCI/SDCI in a Nutshell

- ❖ Pragmatic deployable infrastructure
- ❖ Real infrastructure to support real end users
 - End users in NSF domains
 - Show a community need
 - Avoid “build it and they will come”
- ❖ To a smaller degree, applied research
 - Scalability, self healing approaches, etc.



SDCI: Software Development for Cyberinfrastructure (FY07)

- ❖ Three focus areas
 - HPC, Data, and Middleware
- ❖ Two types of proposals
 - New development
 - Show compelling case for new software development
 - Improvement and Support
 - Original software must have a track record of use and impact



Required Characteristics for SDCI Proposals

- ❖ Multiple application areas and expected usage
- ❖ Awareness/distinction among alternatives
- ❖ Project plan with proof-of-concept and metrics
- ❖ Open source
- ❖ Use of NMI Build and Test facility
- ❖ Demonstration in first 2 years
 - Check in with PO halfway through
- ❖ PI meetings written into budget



How Was SDCI Different

- ❖ Built on the history of ITR and NMI
- ❖ Emphasized production quality results
 - Build and Test
 - Open Source
- ❖ Post award management



SDCI 2007

- ❖ 127 submitted proposals requesting > \$145M
- ❖ 26 awards
 - 5 HPC – All Improvements
 - 6 Data – 5 New, 1 Improvement
 - 15 Middleware – 2 New, 13 Improvement
- ❖ Over \$28M in total award funding
- ❖ Included co-funding support from BIO, ENG, MPS, and EPSCoR



SDCI 2010, Briefly

- ❖ Due Feb 26, 2010; \$15M in funding
- ❖ Similar in spirit to SDCI 2007
- ❖ 5 software focus areas
 - HPC (Debug, Fault Tol, Perf Tuning, Migration)
 - Data (Doc/metadata, security/protection, data transport/mgmt, data analysis and viz)
 - Network (e2e perf, nw tools, testing, co-alloc.)
 - Middleware (instr access, monitoring/mngt/test, user interfaces/accessibility, bridging)
 - CyberSecurity
- ❖ 3 themes: Sustainably, self manageability, power/energy management



Strategic Technologies for Cyberinfrastructure: STCI

- ❖ Support work leading to the development and/or demonstration of innovative cyberinfrastructure services
- ❖ Broadly defined
- ❖ Fill gaps left by other programs and solicitations



Proposals should:

- ❖ Be activities that include a demonstration of the potential impact on science or engineering research or education
- ❖ Generate outcomes not currently under development elsewhere
- ❖ Meet a clearly described cyberinfrastructure need not met elsewhere fill gaps
- ❖ Generate outcomes that will be of interest to a range of science and engineering communities



STCI

- ❖ Successful proposals
 - Emphasize the broad applicability of the work
 - Identify current and prospective end users
 - Include supporting material
 - workshop reports, NSB recommendations, etc
 - Explain why your proposed work cannot be funded under other current NSF programs
- ❖ Target dates twice a year
 - August and February, since 2007
- ❖ Funding varies year to year
 - Usually \$4-5M



What comes into STCI? Everything!

- ❖ Networking
- ❖ Education and Outreach
- ❖ SW library development
- ❖ Portals for science domains
- ❖ CI Training
- ❖ Middleware
- ❖ Debuggers, compilers, schedulers
- ❖ Visualization and data analysis
- ❖ Security tools
- ❖ Etc etc etc



Some Stats

- ❖ Usually about 25-30 per target date
 - However, in the last 12 months over 100
- ❖ How many get funded
 - Usually 2-3 per target date
- ❖ What kind of funding requests?
 - Very small to very large
- ❖ What gets funded
 - Before stimulus funds, average was ~\$800K total over 3 years, largest was \$1.2M over 4
 - Stimulus changed the game
 - Please note: All stimulus funds were spent in FY09

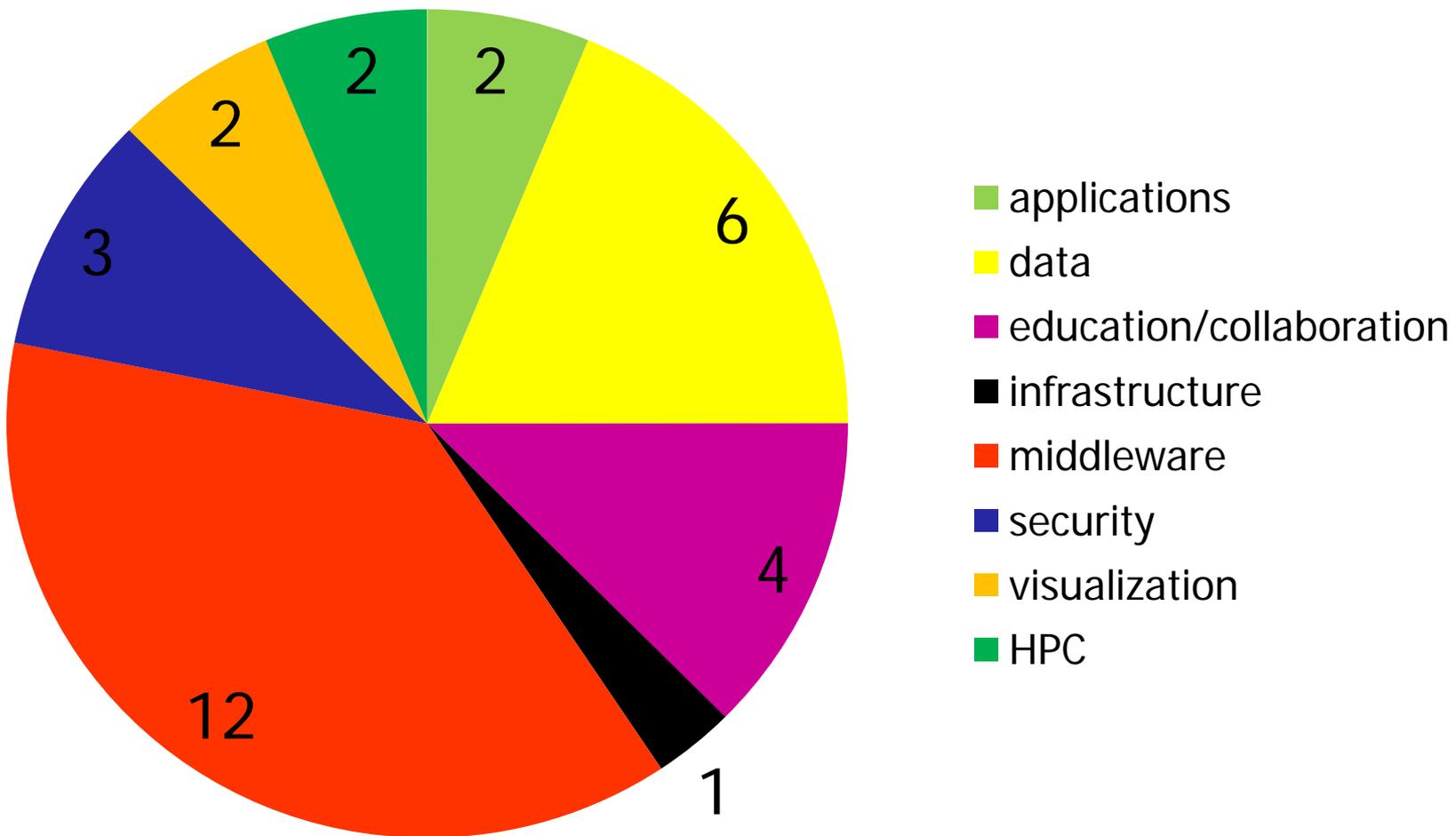


STCI Funded Awards

Type	2007-8	2009 with ARRA funds	Total
Applications		2	2
Data	2	4	6
Education/ Collaboration	1	3	4
Infrastructure	1		1
Middleware	4	8	12
Security	1	2	3
Visualization	1	1	2
HPC libraries		2	2
Totals	10	22	32
Funding	\$8,756,114	\$19,544,682	\$28,300,796



Funded Areas 2007-2009





What are the post-award requirements

- ❖ Close to nothing

HOWEVER

- ❖ All infrastructure meant to be pragmatic, strongly needed, and showing evidence of ongoing support
- ❖ Software proposals, especially, are encouraged to be sustainable



STCI 2010

- ❖ Basic concept the same
- ❖ Review criteria:
 - A clearly identified need not addressed elsewhere
 - A potential for significant impact across multiple disciplines.
 - A potential for generating outcomes of interest to a broad range of communities
 - A list of tangible success metrics
 - Explanation as to why the project is not suitable for other NSF programs
 - A demonstration of strong support from within the science and engineering community



So why are we meeting together?

- ❖ Two programs have similar goals
 - Supporting NSF end user communities
 - Building infrastructure that will last
- ❖ Moving toward production quality infrastructure
- ❖ Thinking about sustainability



What Does Sustainability Mean?

- ❖ “Ability to maintain a certain process or state”
- ❖ In a biological context
 - Resources must be used at a rate at which they can be replenished
- ❖ In a software context
 - Creating software that can be used in broad contexts (reuse)
 - Funding models that encourage long-term support (beyond normal NSF grants)

Note: I’m defining software VERY broadly – everything in your environment, middleware, tools, numerical libraries, application codes, etc.)



One Future: Software As Infrastructure

- ❖ NSF should fund software sustainably the same way it does other infrastructure.
 - Same as telescopes, colliders, or shake tables
 - Line items in the directorate budgets
 - Constant or growing over time, reliably
 - Factor in “maintenance” and “replacement”
 - Eligible for programs like MRI and ARI
- ❖ Software is around even longer than hw
 - Hardware refresh ~3 years
 - Software can grow over decades
 - (what’s the right funding ratio of sw to hw in a large-scale CI project?)



However, if software is viewed as infrastructure by NSF...

- ❖ PIs must also treat it as such
 - Reliable, robust, reproducible, production-quality software
 - Reporting requirements (including uptime, usage statistics, and safety/security reporting)
 - Formal planning approach- including scheduling/estimation, requirements development, deployment plans, risk assessment, etc.
 - Teams with "professional engineering" backgrounds
- ❖ Some people think simply open source is enough – it's not.

Open Source Software is Like a Free Puppy



- ❖ Long term costs
- ❖ Needs love and attention
- ❖ May lose charm after growing up
- ❖ Occasional clean-ups required
- ❖ Many left abandoned by their owners
- ❖ May not be quite what you think





For Sustainability to work

Fundamental Change in culture for both
development groups and funders



This PI Meeting

- ❖ Not your typical PI meeting
- ❖ Exploring together what it means to provide sustainable, pragmatic software infrastructure to support real end user communities
- ❖ Thanks for coming



More Information

More Information

➤ Jennifer Schopf

- jschopf@nsf.gov
- jms@nsf.gov

Office of Cyberinfrastructure

<http://www.nsf.gov/oci>