NSF Software Programs and Convergence of Architectures for BDA and HPC 2016 BDEC Workshop

Rajiv Ramnath

Program Director
Software Cluster
Division of Advanced Cyberinfrastructure
Directorate for Computer and Information Science and Engineering

rramnath@nsf.gov

Version: 6/17/16

Outline

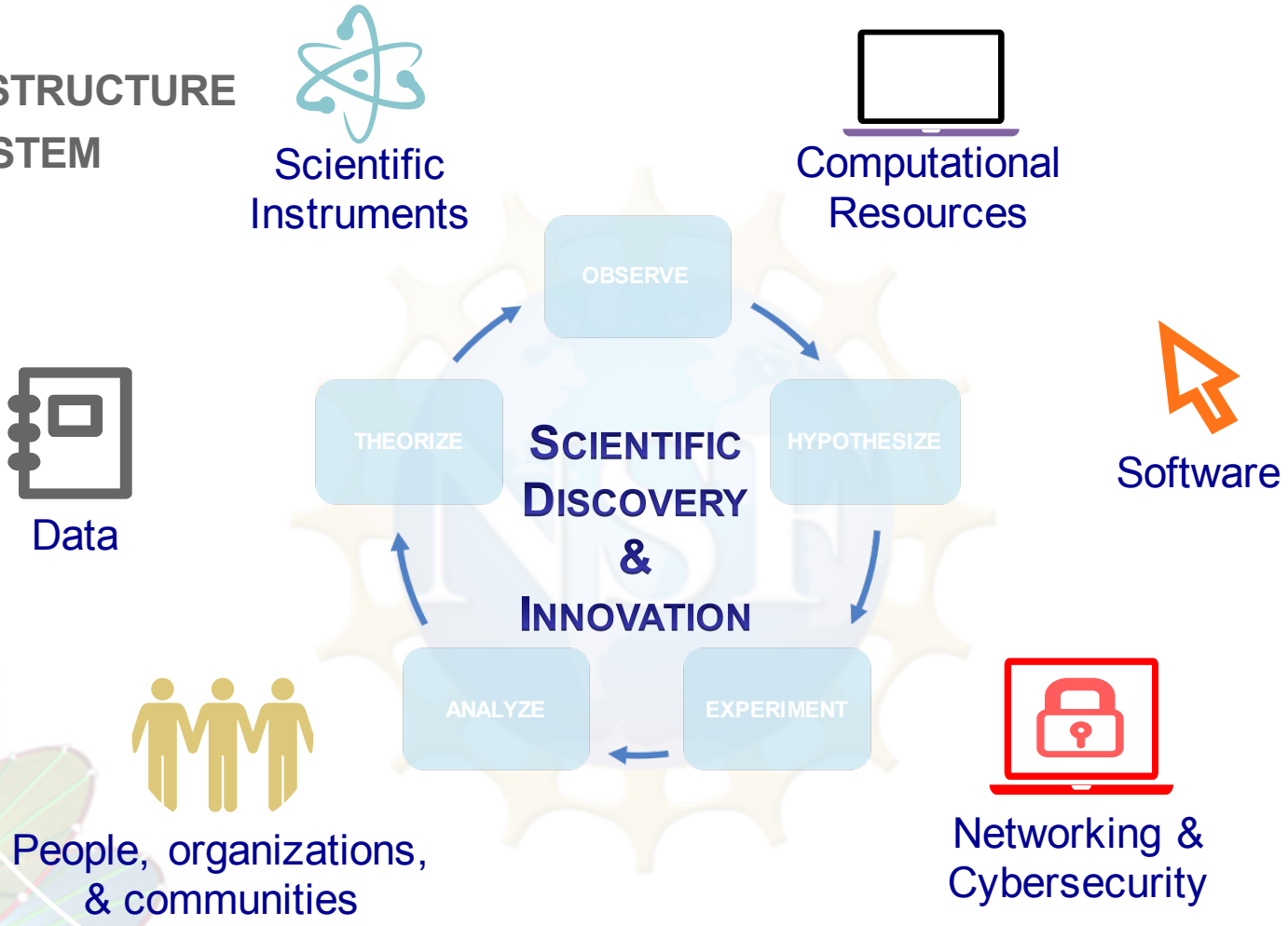
- The NSF CI Vision and ACI
- Software CI Programs
- Future Directions
- Big Data Analytics and HPC

NSF CI VISION AND ACI

NSF Supported Science Infrastructure

Cyberinfrastructure - Computing, Instruments, Data Management, Communications and People

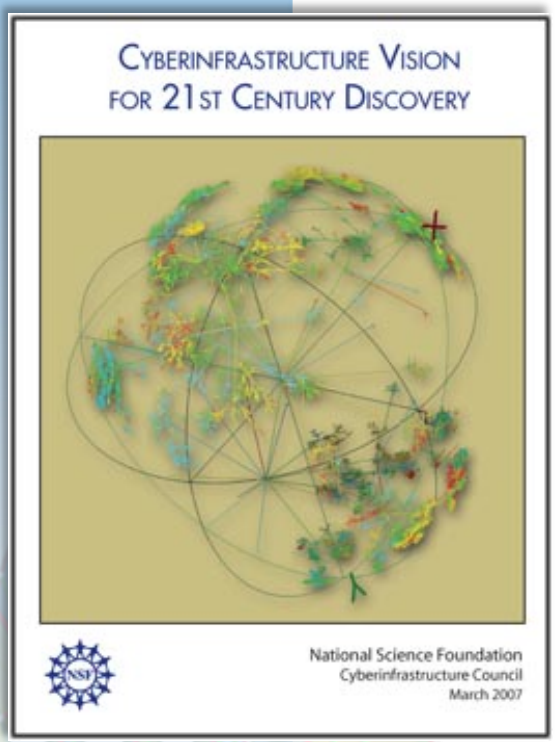
CYBERINFRASTRUCTURE ECOSYSTEM



Driven by research priorities and evolving with the scientific process

Informed by community input and experience

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

Software

Researcher

Shared Data/Software Gateway Resources

Collaboration Networks

Data

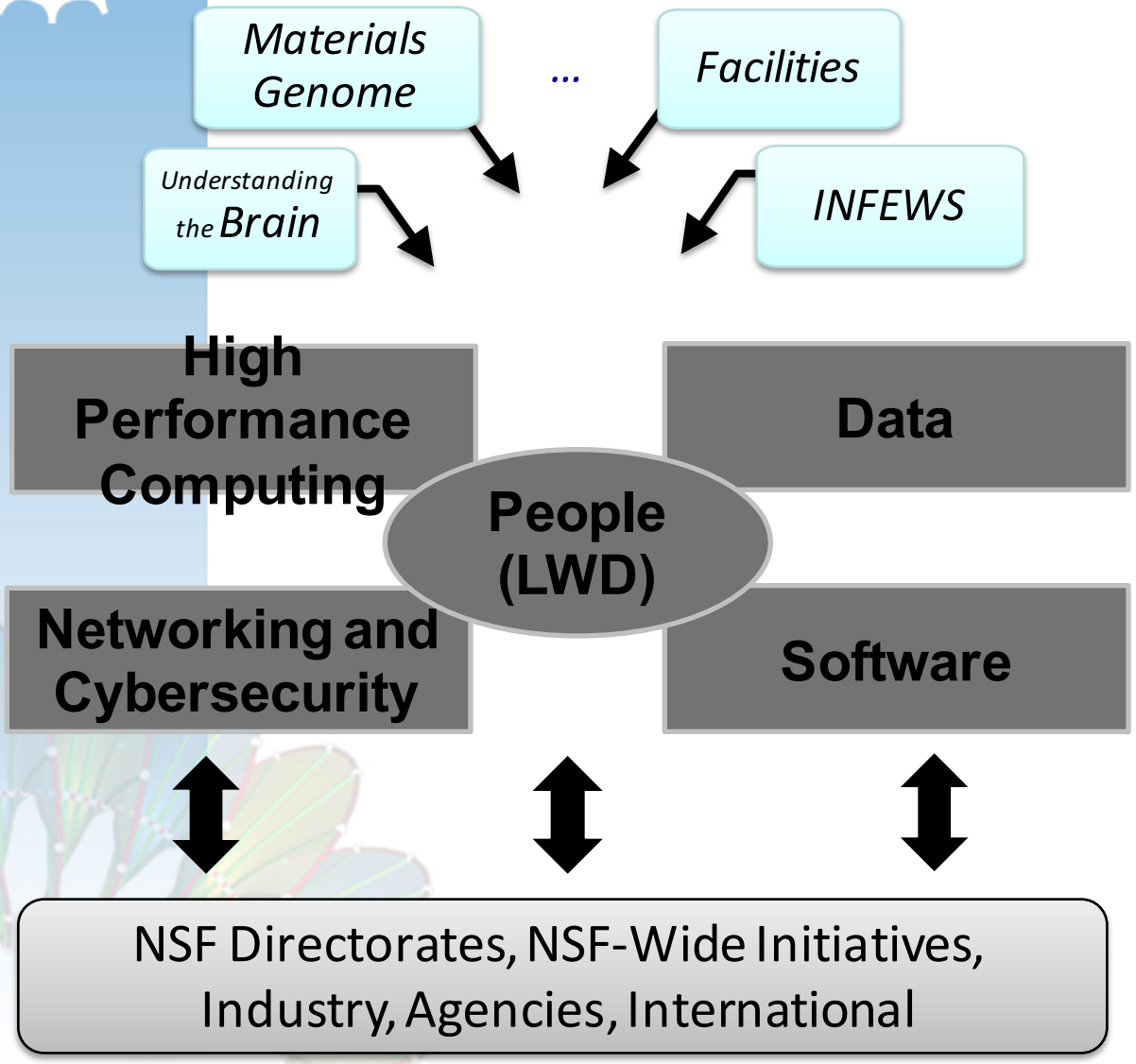
National Computing Resources

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

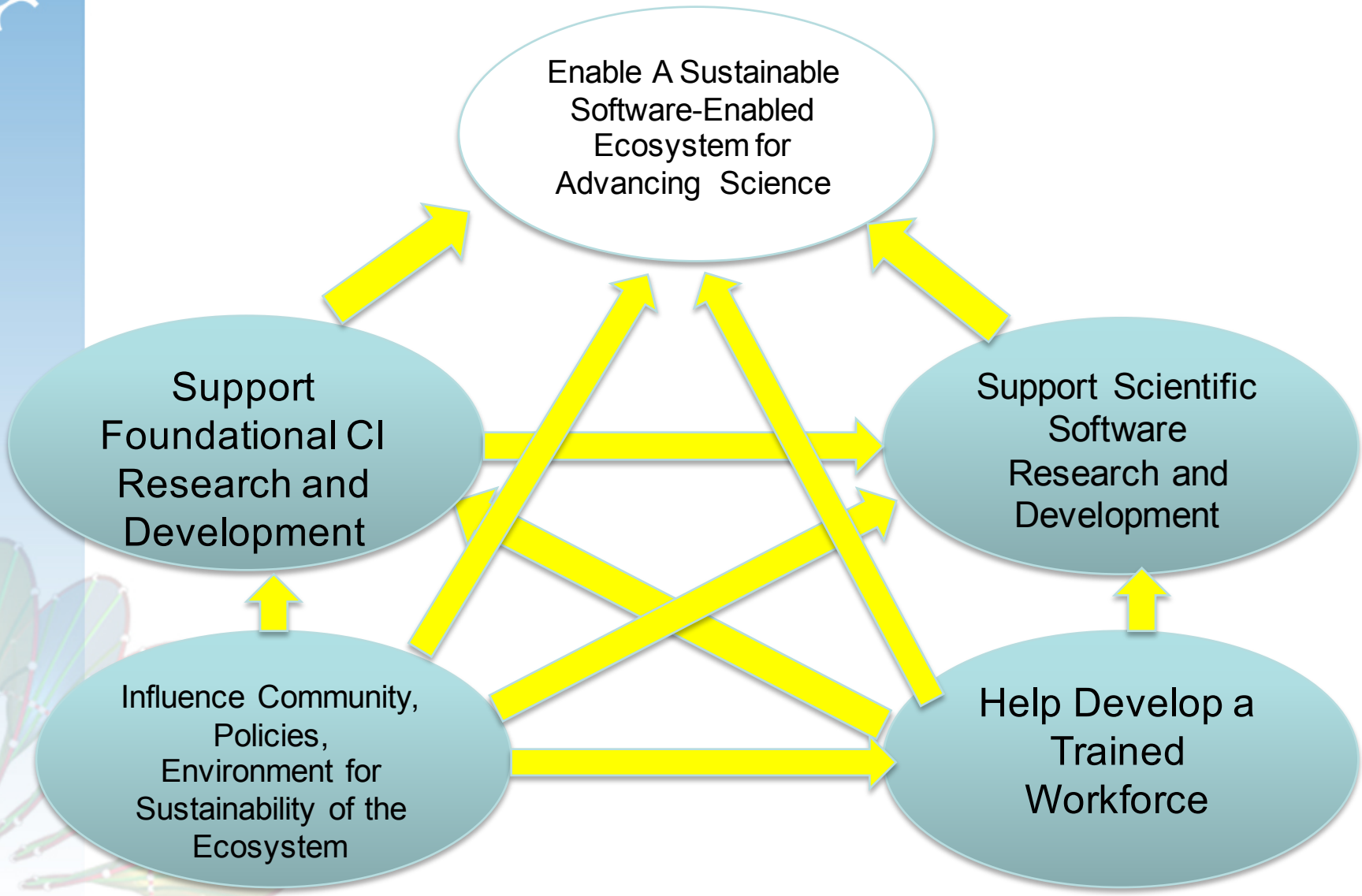
ACI investments

Convergent investments in technologies and communities to maximize impact

Leadership, Coordination, Partnership

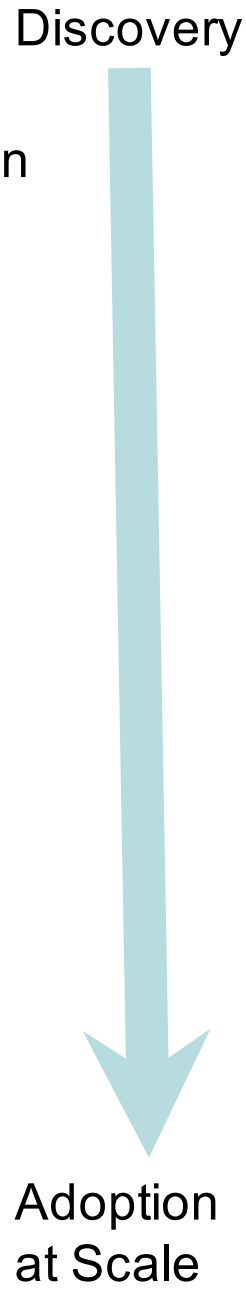
SOFTWARE CI PRIORITIES AND PROGRAMS

ACI Software Program Strategy



Progression of Software CI Programs

- Exploiting Parallelism and Scalability (XPS) $1M-2M
  - Goal: CS research that could lead to CI-relevant innovation
- Computational and Data-Enabled Science and Engineering (CDS&E) $1M-2M
  - Goal: Capture computational techniques in software
- Designing Materials to Revolutionize and Engineer our Future (DMREF) $1M-$2M
- Critical Resilient Infrastructure Systems and Processes (CRISP) $1M-2M
- Software Venture (Reuse) Fund $1M-2M
  - Goal: Encourage reuse
- Software Infrastructure for Sustained Innovation (S2I2) $15M-$20M
  - Goal: Bring innovation to scientific discovery
  - Focus: Robustness, adoption, sustainability
  - Flagship



Other Programs with Software Engagement

- Advances in Biological Informatics (ABI)
- Geoinformatics Program in the Division of Earth Sciences (EAR)
- Polar Cyberinfrastructure Program in Polar Programs
- Critical Techniques, Technologies and Methodologies for Advancing Foundations and Applications of Big Data Sciences and Engineering (BIGDATA)

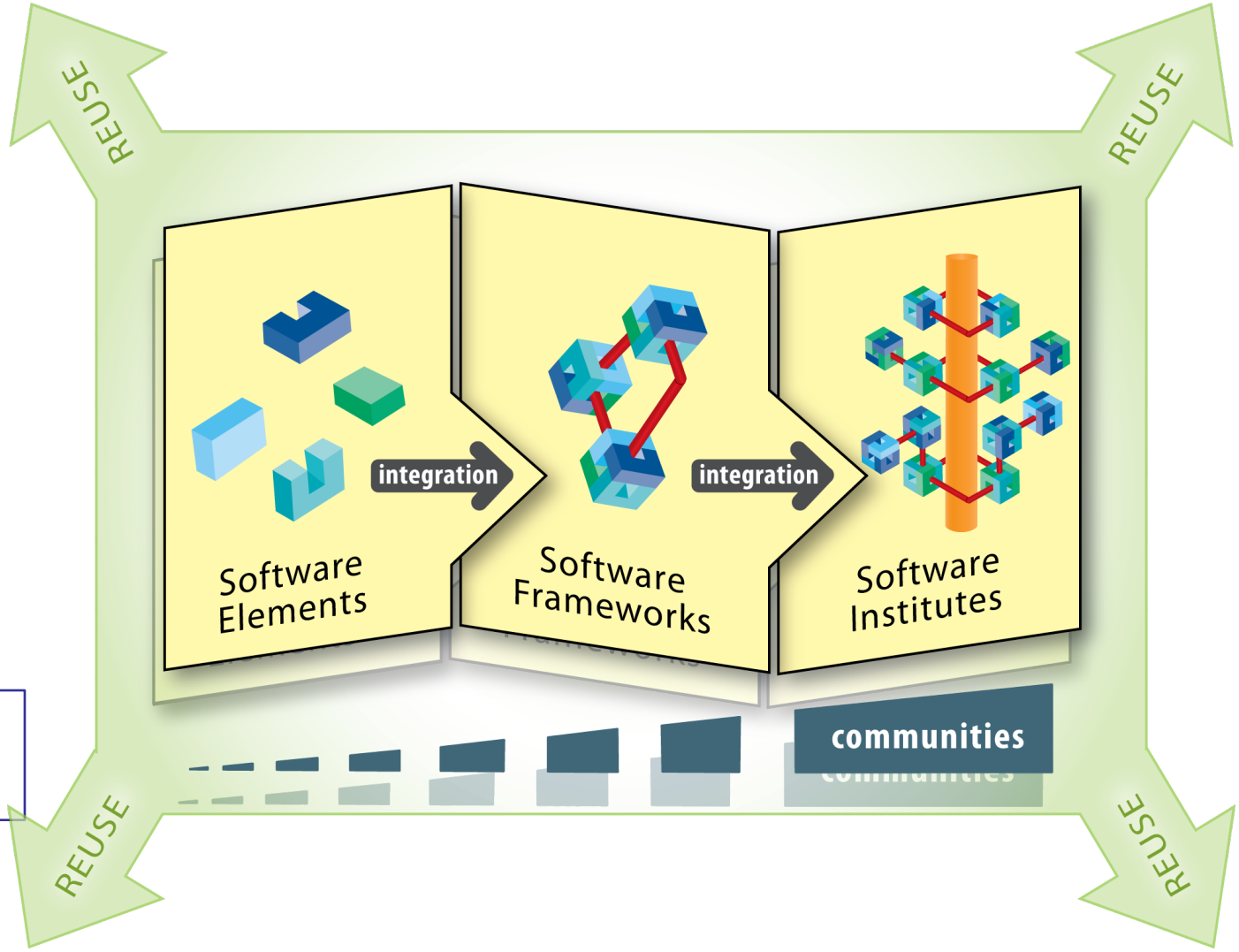
- National Science Foundation Research Traineeship Program (NRT) CIF21 Track
- EAGERS, RAPIDs and Workshops to address areas of opportunity

ACI Flagship - Software Infrastructure for Sustained Innovation (SI2)

Elements: $500K/3 years

Frameworks: $1M/year 3-5 years

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



Gateways Institute **Chemistry and Materials Science Institute**

FUTURE DIRECTIONS

Broader Software Challenges

- **Funding models**. NSF supports projects for up to 5 years; lifetime of software projects can span 20+ years
- **International funding.** Software collaborations span countries, but funding agencies don't
- **Career paths for software-focused researchers**. University structure and academic culture rewards publications; what about researchers whose main products are software? Stop the Google Bus…
- **Incentives, including credit**. How should software be cited? How are all software contributions recognized?
- **Training**. What software engineering practices work in science software?
- **Interdisciplinary work**. Many software contributors work in both computer science and another science or engineering area,or multiple areas, but doing so doesn't fit our siloed system and culture and is often discouraged.
- **Portability**. How to deal with changing hardware, middleware, and languages?
- **Dissemination.** Documenting available software and examples of how it has been used, strengths, weaknesses, and the experience of other users.

2016- Priorities - Towards a National Cyberinfrastructure Ecosystem

- Multidisciplinary and omni-disciplinary software:
- That builds on other ongoing NSF-supported programs
- Aligns with national priorities such as NSCI
- Techniques, tools and processes for rapid integration of software that reduces cost of custom solutions and custom integrations
- Includes embedded innovation and research on the development, effectiveness, usability, adoption, and organizational aspects of the software and the project.
- Has serious considerations of security, trustworthiness and reproducibility.
- Identifies comprehensive, innovative approaches to sustainability (e.g. SAAS, incorporation into university offerings, commercialization)
- Enables science-inspired education and workforce development
- Is assessed using comprehensive metrics of impact
- Promotes software best practices and standards
- Builds a highly-capable workforce that uses and sustains this software and is sustained by the ecosystem.

In 2026

- Address societal grand challenges:
  - Disaster scenarios imposed by climate change
  - Sustainable provisioning of food, water and energy
  - Economic shifts imposed by a decentralized world order
  - Education for all
- Broader Research Participation and Processes:
  - Distributed, dynamic, multi-disciplinary collaborations
  - Lots of research “in the wild” in situ with the object of research
  - Data-driven, integrative over multiple scales, “dual-use” sources
  - Spectrum of stakeholders, participants –scientists, citizens, industry.
- Dynamic, Integrated Research Processes:
  - Coupled modelling, simulation, experiment, data, informatics
  - Range from explorations to repeatable workflows and back.
  - Research methods – ranging from quantitative to qualitative - will need to be drawn from across disciplines and then integrated.
  - Participants will engage and disengage depending upon the stage of the research.
  - Datasets, instruments and computation will be brought in and utilized as and when needed.

Dynamic CI Processes on Interoperable, Configurable CI for Science at Large Scale

- A continuum of usable, robust, infrastructures
- Toolboxes of composable computational research methods
- Workflows that adapt to data and humans in-the-loop.
- Dynamically configurable systems, software and networks (software defined everything, software injection, parameterized components)
- Needing security, access, reproducibility and trustworthiness techniques for dynamic situations (human back in the loop?).
- HCI that supports a range of users
- Dynamic business models potentially with "on-demand negotiation”
- Learning aimed towards “integrative synthesis” rather than disciplinary depth.

CONVERGENCE OF BIG DATA STACKS AND HPC

ACI Guidance to BDEC

- Work towards system architectures that will “just work”.
- Propose integrative efforts – e.g. workflow-based bridging to evolving underlying system architectures
- Be use driven (tied to clear science drivers, start with pilots)
- Work towards a National Cyberinfrastructure Ecosystem
- Leverage our programs – we will help guide you!

Potential Science Drivers

- Design from First Principles of Materials (e.g. Catalysts)
- Physics-based Multi-Scale Earth System Models
- Quantitative Approaches to Biomedical Big Data
- Advanced Computation for Building Systems
- Others? Please Propose ..

Common Themes

- Linking across multiple scales
- Building from first principles
- Efficient uncertainty quantification (recording and replaying intermediate states).
- Computational steering with in-situ data analysis combined with computation
- “Composable” software and hardware stacks
- Productivity support – compiilers, IDEs

How NSF Programs Will Evolve

- EAGERs, Workshops, DCLs for Exploration and Pilots
- 2017 SI2 solicitation that will provide clearer guidance re: the National Strategic Computing Initiative (NSCI)
- NSCI will become the overarching framework 2018 onwards
- CI-Focused And Domain Driven Institutes
  - A software platforms Institute?
  - Other domain science institutes?

QUESTIONS?