Sustained Investment in Software

Manish Parashar
Office of Cyberinfrastructure,
National Science Foundation
http://www.nsf.gov/oci
Science is Revolutionized by CI

- Modern science
  - Data- and compute-intensive
  - Integrative

- Multiscale Collaborations for Complexity
  - Individuals, groups, teams, communities
  - Great challenges of 21st century

- Must Transition NSF CI to support
What is needed?

**CF21: NSF-wide CI Framework for 21\textsuperscript{st} Century Discovery**

- Comprehensive, *balanced, integrated*, national high performance CI
- Support for a broad vision of cyber-science
  - All components needed for 21st century science
  - People who use/develop CI: Cyber-workforce development
- Architect carefully, collectively across all NSF
  - Dear Colleague Letter by all NSF units in December 2009
  - Six national Task Forces underway to inform
    - All NSF offices represented
    - HEC, Campus Bridging, Software, LWD, Data, “GCC/VO”
  - New NSF-wide programs being developed; Campus-to-MREFC integration
Cyberinfrastructure Framework 21st Century Science and Engineering (CF21):

New Paradigms & Practices

• Transformed by CI
  • End-to-end
  • Fundamentally collaborative & data-driven
  • Software is an integral part
• Unprecedented opportunities
• New requirements, challenges
• New thinking in/approaches to computation science!
OCI Special Role in CF21

- Driver for integrative CI activity via CF21
  - Working with all units, community to develop the vision and implementation plan

- Catalyst for *coordinated, linked* investments in
  - CI in all forms: campus, centers, MREFC
    - Leadership in prototypes, pilots and early deployment; establishing best practices
  - People: supporting next generation of researchers
  - Cyber-science: applications of CI to enable science, leading to new science and collaborations

- Coordinator of NSF ACCI Task Force Activity
  - Advisory Committee on CI (ACCI) has members representing all NSF units
Software is Critical

- CI – Unprecedented complexity, challenges

- Software is essential to every aspect of CI – “the glue”
  - Drivers, middleware, runtime, programming systems/tools, applications, ...

- This software is different .... ?
  - In its natures, who builds it, how is it built, where it runs, its lifetime, etc.

- Software crisis?
  - Software complexity is impeding the use of CI
    - Science apps have $10^3$ to $10^{6+}$ lines, have bugs
    - Developed over decades – long lifecycles (~35 years)
  - Software/systems design/engineering issues
    - Emergent rather than by design
  - Quality of science in question
Approaches to SW

- Funding for software development fails to support needs of the full life cycle
- Current models typically ad-hoc
  - Software as a byproduct
    - Scientific competence ≠ Software competence
  - Software in the small (prototyping, proof of concept)
  - Software in “isolation” – reinvention is common
- Some successes + many not very good
  - Unpredictable, unreliable
  - Expensive & requiring unrealistic skill base
  - Support, scaling, maintenance, porting
- No overarching vision, no coordination .... lost investments!!!
Software Grand Challenge

- **SW as the modality for CF21 and Computational Science in the 21st Century**
- **Sustainable SW as a CI resource**
  - What SW to sustain?
  - How to sustain it?
- **Fundamental Grand Challenge: Robust, Sustainable and Manageable Software at CI-Scale**
  - Repeatability, Reliability, Performance, Usability, Energy efficiency, ....
- **Sustainability, manageability, etc., are NOT add-ons – it has to be integrated into the design**
Many complex aspects....

- **Building the right software** – application involvement, understanding requirements
  - scales, types of software, target user communities

- **Building software right** – teams, reward structures, processes, metrics, verification/testing

- **Protecting investments** – active management, sustainability, leverage/reuse, ownership, business models

- **Building trust** – user community must be able to depend on the availability of a robust and reliable software infrastructure!
Sustained Long-Term Investment in Software

- Transform innovations into sustainable software that is an integral part of a comprehensive cyberinfrastructure
  - robust, efficient, resilient, repeatable, manageable, sustainable, community-based, etc.

- Catalyze and nurture multidisciplinary software as a symbiotic “process” with ongoing evolution
  - Domain and computational scientists, software technologists

- Address all aspects, layers and phases of software
  - Systematic approaches
    - Theory validated by empirical trials
  - Tools that embody and support processes
  - Metrics, validation mechanisms, governance structures
  - Amortised over large (global) user communities
  - Support for maintenance and user support
Sustained Long-Term Investment in Software: Mechanisms

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

Focus on innovation  Focus on sustainability
Sensor Nets

Experiments/Instruments

Data Archives (DataNet)

Scientific Software Innovation Institutes (S2I2)

Visualization/Analytics

Infrastructure (XD)
Sustained Long-Term Investment in Software: Misc Issues

- Scale and complexity beyond community experience
  - Many unknowns: models, modes, scales, ....
    - domain, community specific aspects...
    - crosscutting aspects and many links...
- Must be grown bottom up in a coordinated way
  - smaller group evolving into community wide teams and institutes
- Must leverage existing investments, expertise
- Collaborations across communities, disciplines and directorates critical
Sustained Long-Term Investment in Software: Metrics of Success (Beyond LOC)

- Buy-in from the broader community
- Demonstrated leverage and reuse
- Emergence of successful models, processes, architectures, metrics for S&E software – empirically validated
- Emergence of models and mechanisms for community sustainability of software institutes
- Accepted research agenda by academic community
Summary

- Science is being revolutionized through CI
  - Compute, data, networking advance suddenly 9-12 orders of magnitude after 4 centuries
  - All forms of CI—integrated—needed for complex science

- NSF responsive: developing much more comprehensive, integrated CF21 initiative
  - Community deeply engaged in planning
  - Activities begin FY10, ramp up FY11-12 and beyond

- Focus on sustainability, people, innovation
  - Longer term programs, better linked, hubs of innovation
  - Support development of computational scientists who develop and/or use advanced CI

- Robust, reliable, sustainable software is critical!
Thank You!

Sustainable System
“meets the needs of the present without compromising the
ability of future generations to meet their own needs”
[UN Brundtland Report 1987, of sustainable development]
More Information

- Manish Parashar
  mparasha@nsf.gov
  www.nsf.gov/oci