

NEES Cyberinfrastructure: Meeting the Needs of the Civil Engineering Community

Thomas Hacker

Purdue University

West Lafayette, IN

October 2, 2013



NEES

My background

- Co-Leader for Information Technology for the NSF George E. Brown Network for Earthquake Engineering Simulation
- Co-PI for Purdue U.S. CMS Tier-2 Center
- Purdue University - Associate Professor, Computer & Information Technology
- University of Stavanger – Visiting Professor of Electrical Engineering and Computer Science
- Interested in research and education in large-scale cyberinfrastructure systems, reliability, high performance computing and networking

Challenges in operating cyberinfrastructure

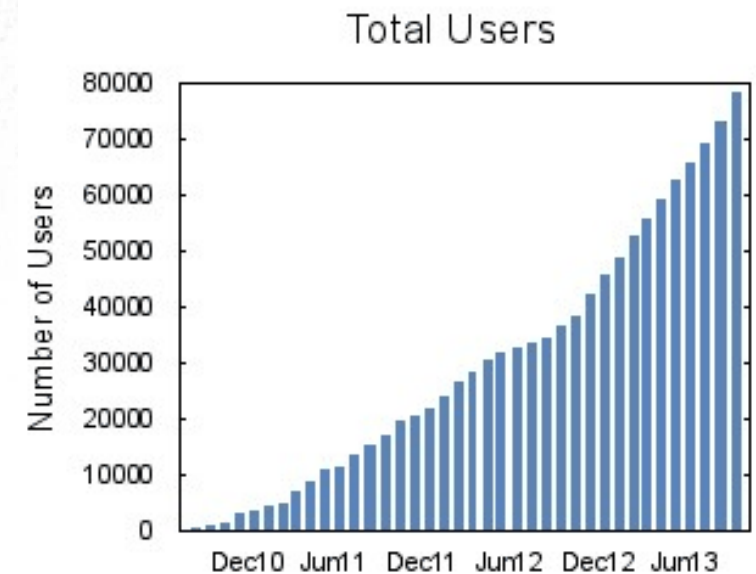
- **Technological**
- **Sociological**

Technological Challenges – Integrating new technologies

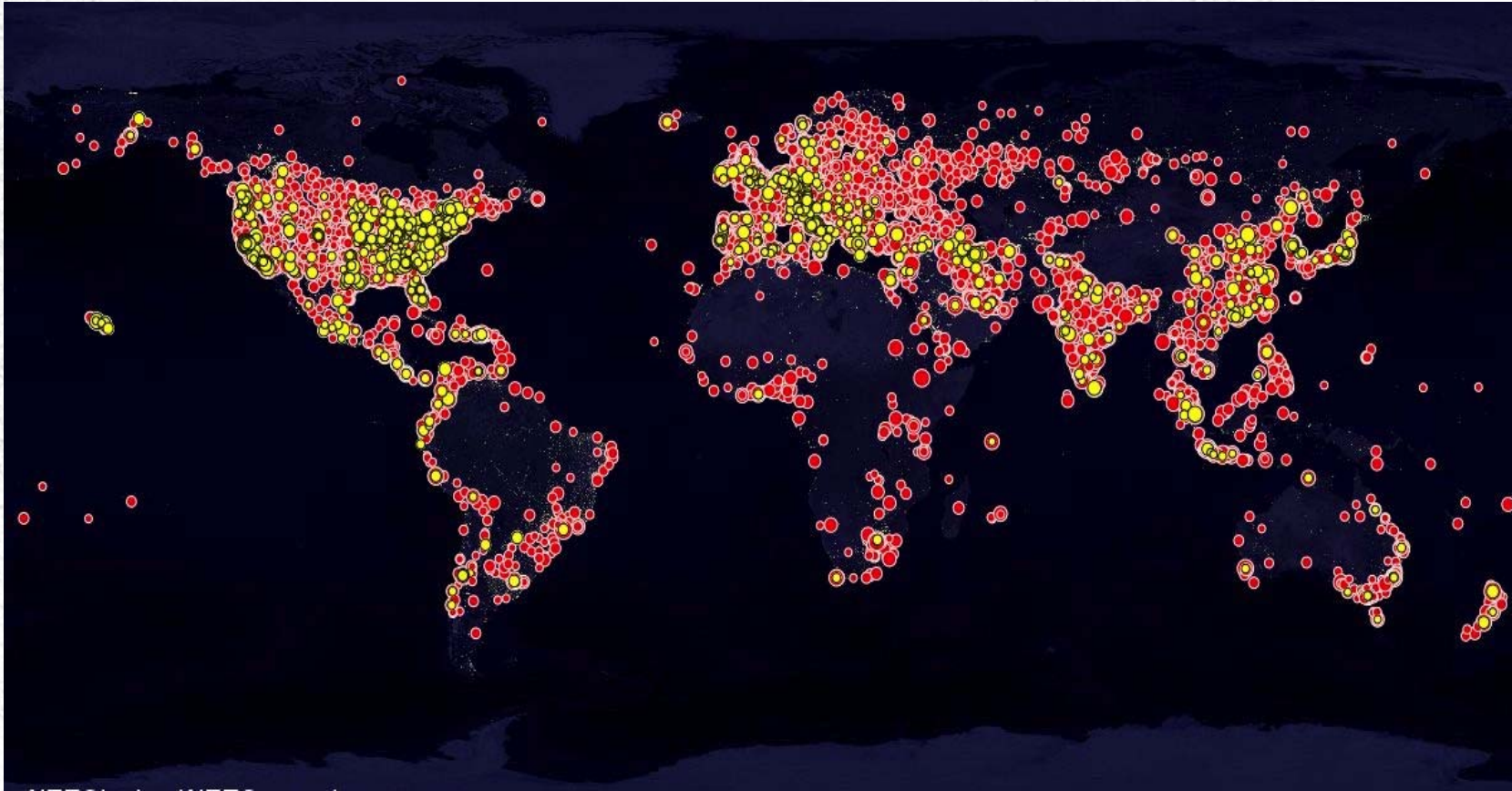
- Reviewing and integrating new technologies
 - New software and services always emerging
 - Infrastructure level
 - Examples: VMware, 10 GBASE-T networking, SSD, OpenStack
 - Are mostly invisible to the users
 - Visible services for users
 - Capability to run Windows tools within the NEEShub
 - New software tools for the community
 - Push and pull of technologies in the campus and national CI community
 - Local: HUBzero community source, Purdue campus clusters
 - National: deciding how and when to integrate new technologies – what are the criteria for selecting?
- Decision to evaluate and use new technologies
 - ***Driven by the documented and prioritized needs of the community***
 - New feature must help to fulfill a requirement
 - Usability and adoption by users is a significant factor
 - Adoption rate slowed if the service is unreliable or requires significant effort to learn

Technological Challenges - Scaling

- Size of the community
 - NEEShub used by 78,177 users over the past 12 months
- Global reach
 - Growing use from 211 countries since 2010



Monthly count of the annual number of users. Source: nees.org/usage

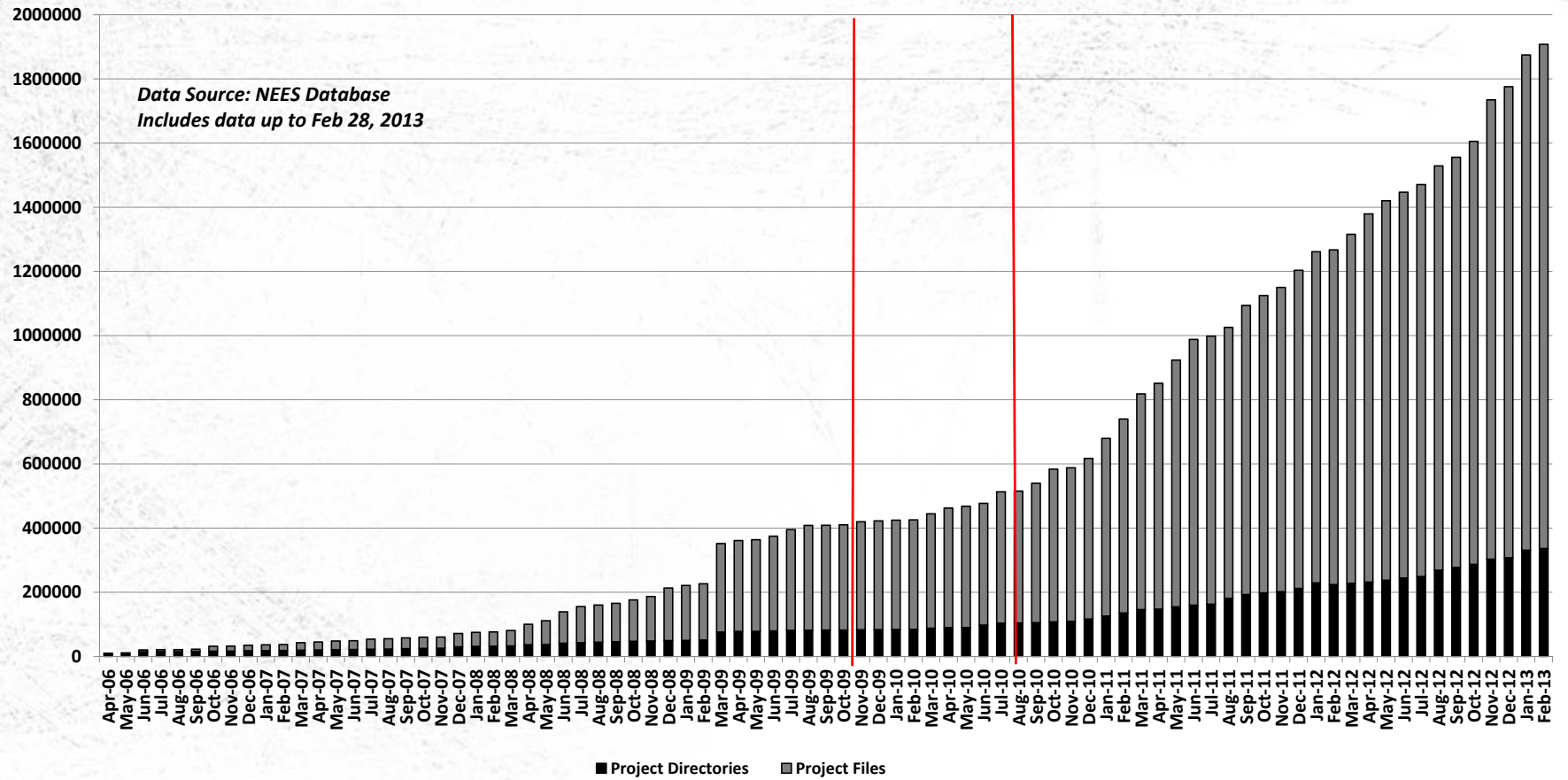


Red dots represent researchers and students browsing NEEShub, watching videos, and taking courses while performing 840,656 web and 38,854 tool sessions between August 2010 and April 2013. Yellow dots represent users who are running simulations. Dot size corresponds to the number of users at a location.

Technological Challenges - Scaling

- Volume of content
 - Growing amount of projects and project files and directories in the Project Warehouse
 - Over 2M project files and directories
 - 227 projects containing over 1500 experiments
 - Pace of file and directory creation increased by 7X since 2010

Cumulative Number of Project Files and Directories Created in the NEES Data Repository



Technological Challenges – Cybersecurity and Data Curation

- Maintain the security and integrity of NEES cyberinfrastructure
 - Scanning for potentially vulnerabilities
 - Monitoring traffic and user activity
 - Enforcing password change policies
 - Continual security scanning of the Project Warehouse
- Ensuring the long-term viability of data
 - Full-time data curator
 - Maintaining file integrity
- Data reuse by others in the future
 - Sufficiently detailed metadata needed to understand stored data
 - Data curator works with researchers to audit and review project data
 - File types (i.e. PDF, Word, etc.) must be readable in the future
 - Need to save of copy of a file in an open format (e.g. ODF)

Curation status of NEES projects and experiments In the NEES Project Warehouse

Curation Status in the Project Warehouse

Thank you for visiting the NEES Site Schedule and Curation page. To see the multi-year schedule for each site, please click on the schedule link under the site name. To download an excel spreadsheet with filters that contains all projects in the NEES network, please click on the link under "All Sites". If you are interested in the curation status of individual sites, please click on the link 'Curation Status' under each facility. The following tables display curation status for the whole NEES data repository. The last column is updated in real time.

The data from a test are expected to be curated and public within twelve months after the tests are completed. Projects and experiments that are curated and public are considered **COMPLETE**. Projects and experiment that are on-time following per the NEEScomm policies are considered in good standing and **CURRENT**. Projects and experiments that fail to meet the deadlines for archiving and publication of their research data are considered **NON-COMPLIANT** with the [NEEScomm Data Sharing and Archiving Polices](#).

Note: Although the sites have no control over curation status, the data is listed by site for convenience of looking up individual projects.

The curation status as of Wednesday 2nd of October 2013 10:47:55 AM :

Total NEES Curation Status by Projects

	Jul 2012	Jan 2013	Oct 2013	Percent Complete
Complete ?	63	76	98	<div style="width: 60%;"><div style="width: 60%;"></div></div> 60%
Current ?	27	23	39	<div style="width: 24%;"><div style="width: 24%;"></div></div> 24%
Non-compliant ?	43	44	25	<div style="width: 16%;"><div style="width: 16%;"></div></div> 16%
TOTAL	133	143	162	100%

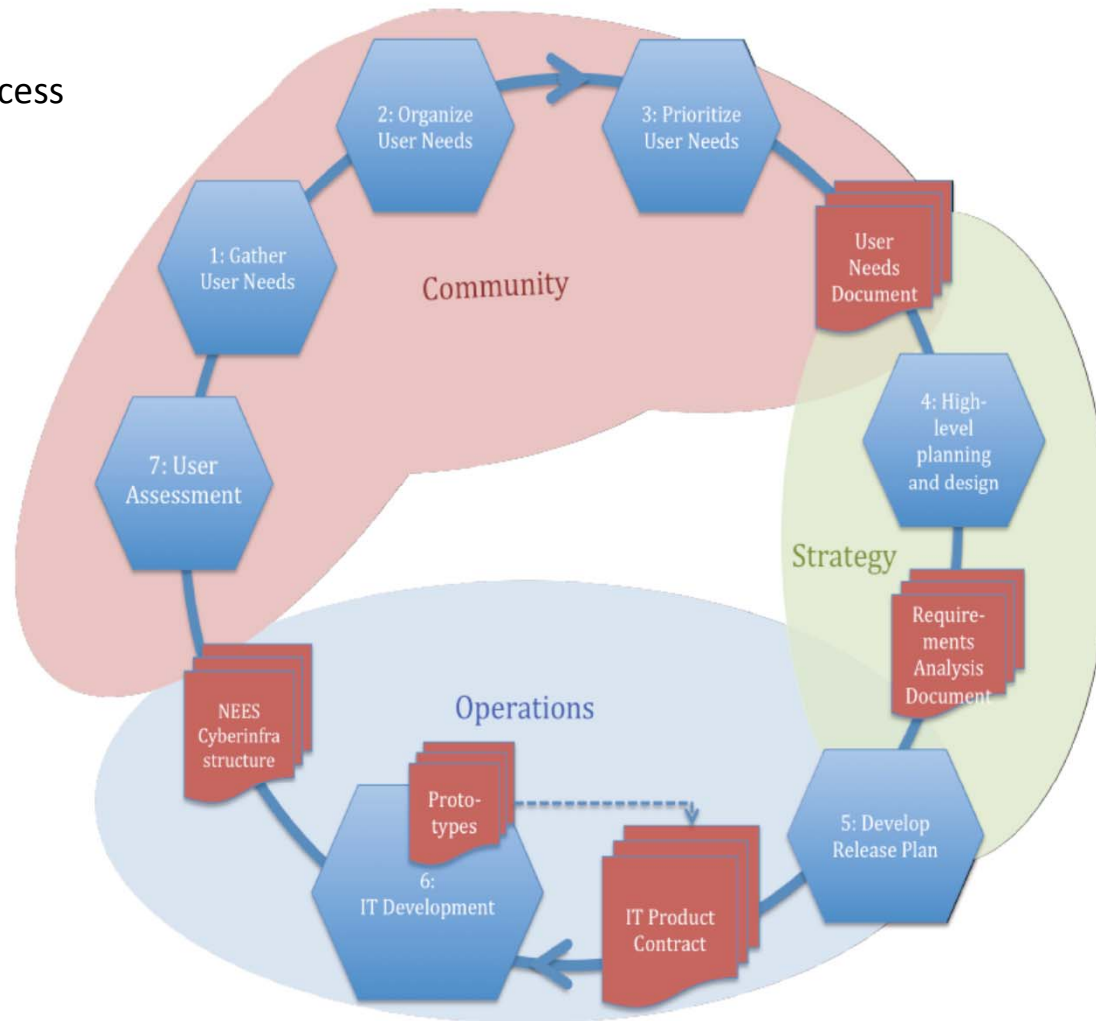
Total NEES Curation Status by Experiments

	Jul 2012	Jan 2013	Oct 2013	Percent Complete
Complete ?	604	649	1246	<div style="width: 86%;"><div style="width: 86%;"></div></div> 86%
Current ?	83	97	98	<div style="width: 7%;"><div style="width: 7%;"></div></div> 7%
Non-compliant ?	152	177	102	<div style="width: 7%;"><div style="width: 7%;"></div></div> 7%
TOTAL	839	923	1446	100%

Sociological Challenges – Community involvement

- Prioritizing CI efforts and ensuring that the CI is usable by the community
 - Requirements Analysis and Assessment Subcommittee (RAAS), Project Advisory Committees
 - Prioritized requirements, core feedback
 - Software development process focused on highest priority requirements with release cycle
- Domain scientists need to be involved
 - Every aspect of cyberinfrastructure design, development, and operation
 - The CI must meet the needs of the community
- Identifying the community
 - Researchers, practitioners, and students involved in improving the response of the built environment to earthquakes and tsunamis.
 - NEES site staff
 - Different communities: faculty, staff, site IT folks
 - Staff at NEES sites who aid the community to design and build specimens and to conduct experiments.
 - NEEScomm staff
 - Recruitment, training – finding highly skilled staff

NEES IT Development Process



Sociological Challenges – Facilitating adoption of cyberinfrastructure

- Rogers *Diffusion of Innovations* model
 - Knowledge -> Persuasion -> Decision -> Implementation -> Confirmation
- Help the community to move along this process to discover, learn, and adopt cyberinfrastructure
- Pedagogy
 - Much of the research work is performed by graduate students
 - NEES has training programs focused on helping the community learn about and learn to use CI facilities (e.g. NEEShub Boot Camp, NEESacademy, and Webinars)
- The key is to provide CI facilities that meets the expressed needs of the community