



George E. Brown Jr.  
Network for Earthquake Engineering Simulation

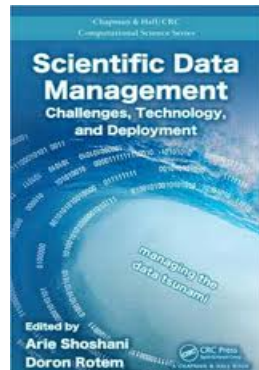
Facility Access Challenges:

# Data Management

Rudi Eigenmann  
*NEES Operations Headquarters*  
*NEEScomm Center*  
*Purdue University*



# Why Data Management ?



LFW 2011: Breakout

- The Data
- Best Practices, Lessons, Challenges
  - Is Synergy Possible?

op

NSF Data Initiatives  
BigData (NSF 12-499)

9 NSF Directorates  
& Offices and 7 NIH  
offices involved



# NEES Physical Infrastructure



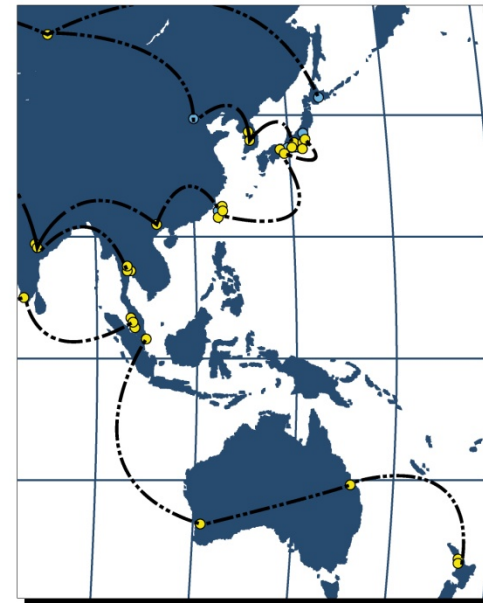


# NEES Cyberinfrastructure

## NEEShub - nees.org



NEES Sites



### NEEScomm

- Data Repository
- Computational Simulation
- Community Support



Cornell University



University of California, Davis



University of Texas, Austin



Rensselaer Polytechnic Institute



University of California, Los Angeles

**Legend**

- NEES Equipment Site
- NEES Partnership
- NEES Researcher



University of Minnesota



University of Illinois at Urbana - Champaign



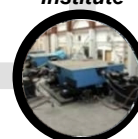
University at Buffalo



Lehigh University



University of California, Berkeley



University of Nevada, Reno



Oregon State University



University of California, Santa Barbara



University of California, San Diego



# NEEShub Project Warehouse

Goal: Data re-use

The screenshot shows the NEEShub Project Warehouse website. The header includes the NEEShub logo and navigation links like 'Tools & Resources', 'Learning & Outreach', 'Project Warehouse', 'Stats', 'Collaborate', and 'Explore'. A search bar is present. The main content area is titled 'NEES Project Warehouse' and contains a search bar with tabs for 'Search', 'Enhanced Projects', and 'User Guide'. Below this, there is a description of the warehouse as a centralized data repository. A 'Search Project Warehouse' section includes a keyword input field and a 'GO' button. To the right, there are 'Popular Tags' (earthquake, tsunami, steel frame) and 'Popular Searches' (earthquake (1126), Variable Rate Cone Penetration Test (737), Shake Table (547), steel frame (506), concrete (503), Soil Resistance (440), punching shear (414), wall (375)). A highlighted section titled 'Semiactive Control of Nonlinear Structures' lists four experiments with their dates and descriptions, each accompanied by a 'Launch Data File' link and a small image.



The screenshot displays a software interface for data visualization and simulation. The interface is titled 'NEEShub - Resources: Tools: inDEED: Session: 9244 "inDEED"'. It features a 'Dataset' panel on the left with a tree view of data sources, including 'Column curvature', 'Displacement tran...', 'Longitudinal SGs...', and 'Shake table'. The main area shows a 3D model of a structural system, likely a bridge or a large frame, with various sensors and actuators indicated. A 'Plot' window is open, showing a graph of data. The interface also includes a 'Find sensors' section and a 'Plot selected sensors' section.

Online Interactive Project Warehouse  
Data Repository

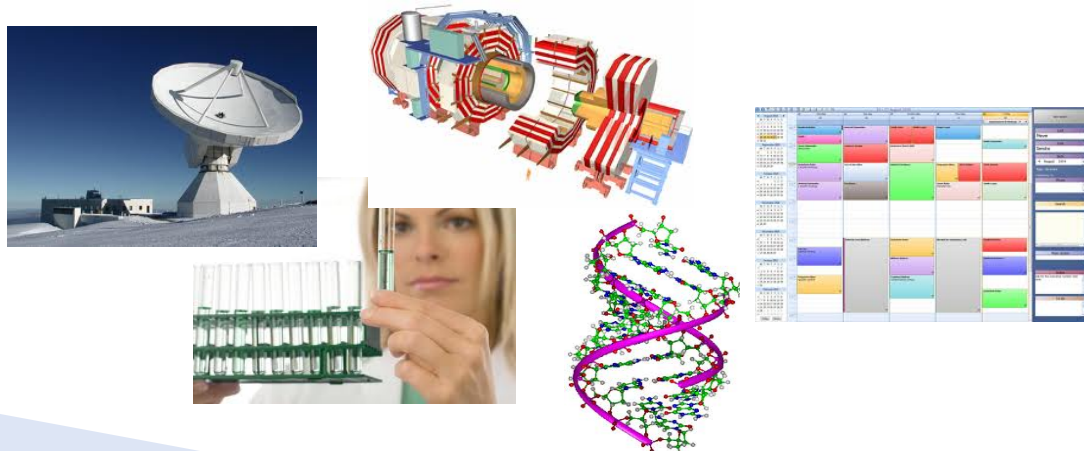
Hosted Online Simulation and Computation 5

# NEES Data in Comparison: Scope & Diversity of data

NEES has 14 sites,  
engineering data

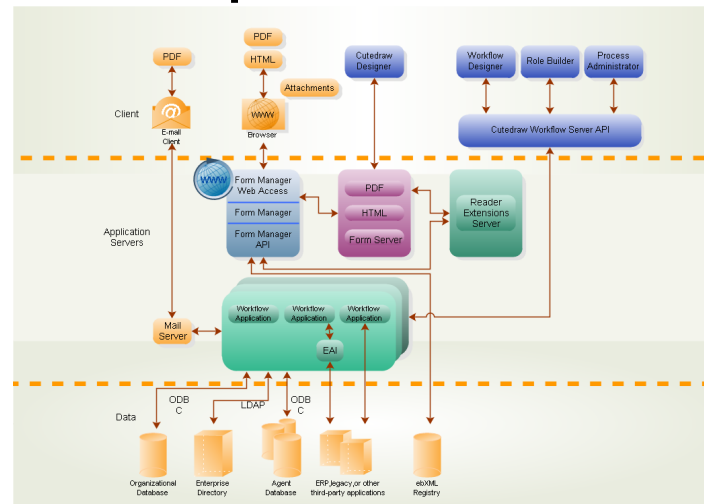


Others: data from earth & universe  
observation, biology, accelerators, medical,  
administrative



# NEES Data in Comparison: Workflow

- ▶ NEES data has simple workflow and provenance relationships

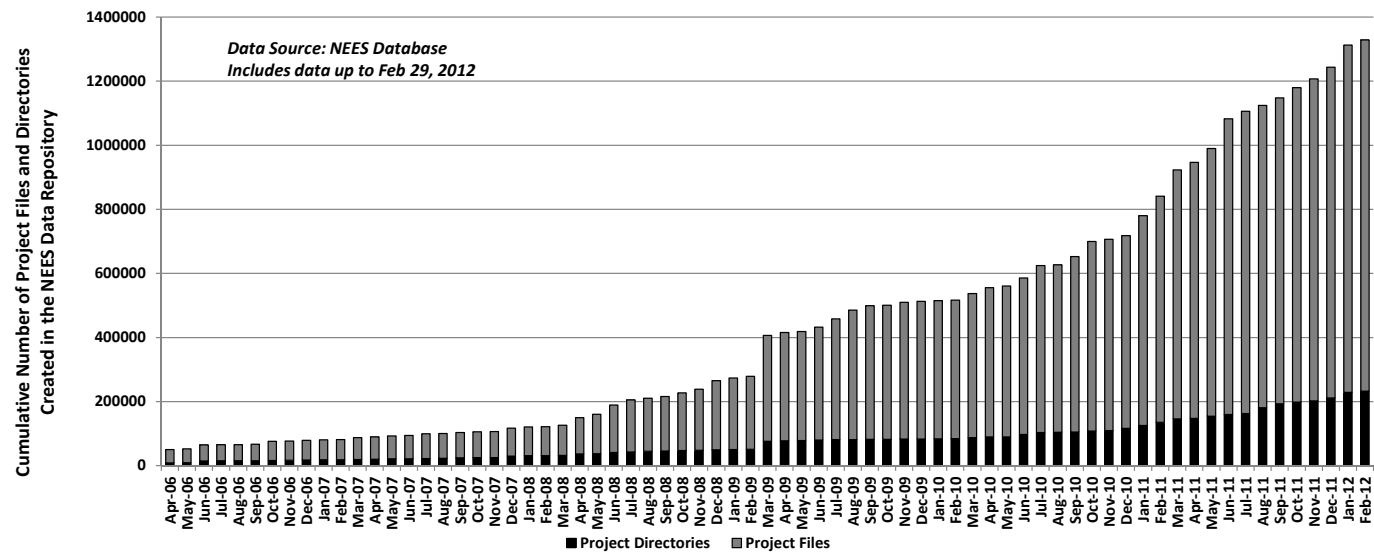


- ▶ Others: complex workflows and provenance, e.g., in medical processes.



# NEES Data in Comparison: Data volume

NEES volumes are moderate (50 TeraBytes):  
sensor logs are small compared to images  
and videos



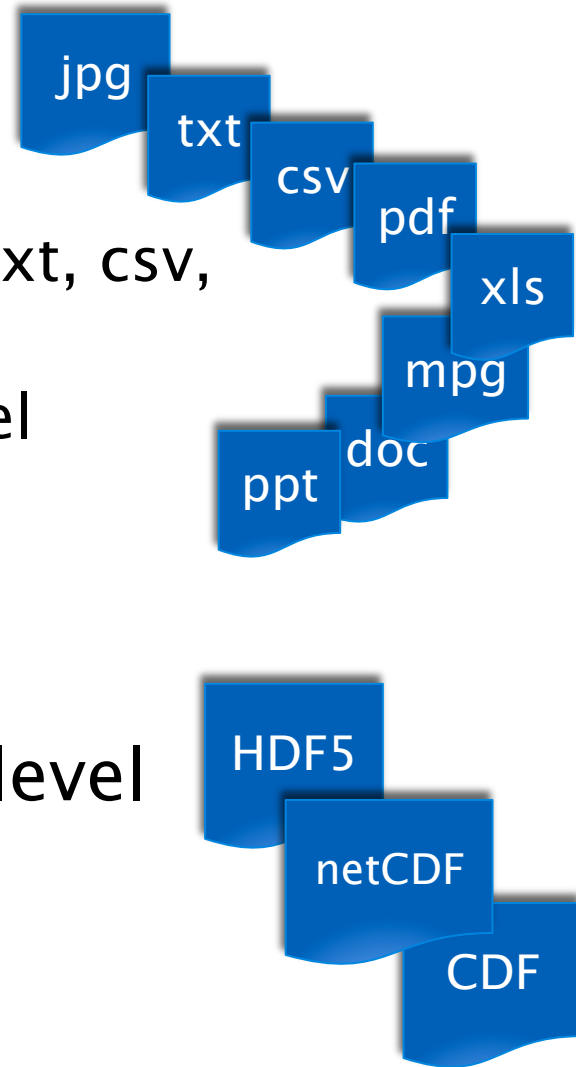
Others: few Terabytes to Petabytes

# NEES Data in Comparison: Data formats

## NEES formats:

- common basic formats (jpeg, text, csv, pdf, xls, mpeg, doc, ppt,...).
- No commonly accepted high-level formats.

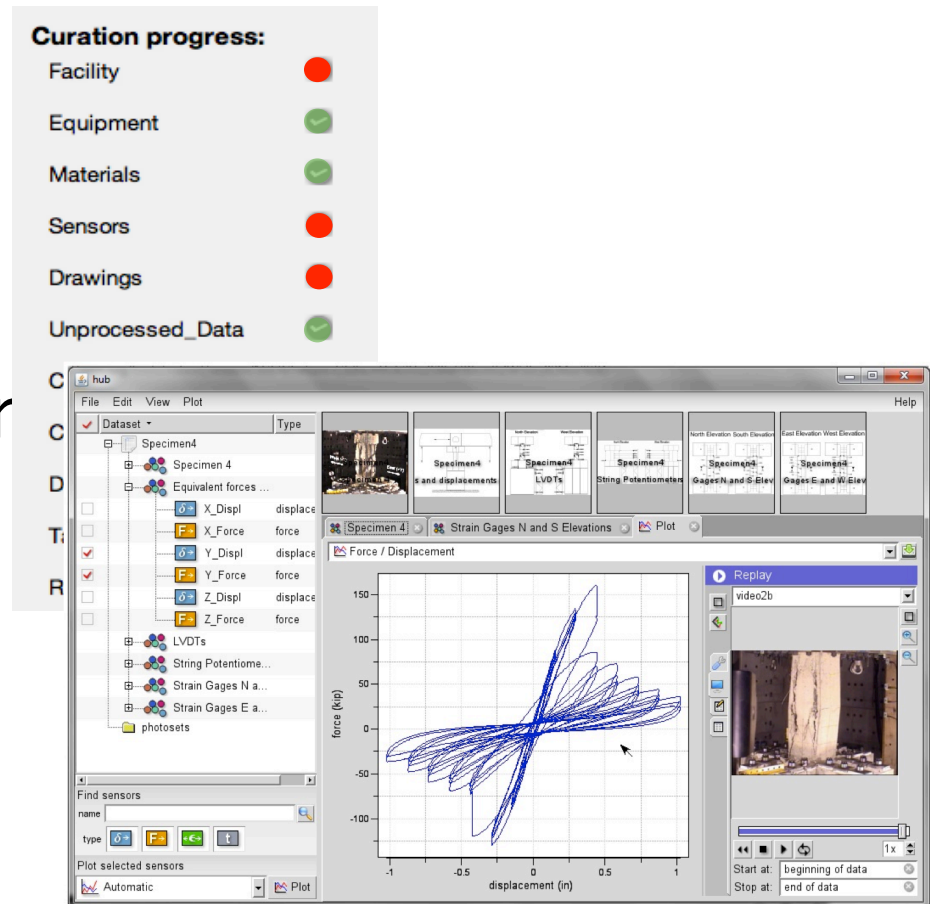
Others: some areas have high-level format agreements



# NEES Data in Comparison: Data curation

NEES Curation goals:

- ▶ quality control
- ▶ long-term preservation
- ▶ maximize reuse



Others: curation is not always an issue



# How does NEES manage data?

## A few key points – by NSF GPG:

- ▶ types of data collected
- ▶ (meta)data standards
- ▶ Policies
  - access and sharing
  - Intellectual property
  - re-use
- ▶ plans for data archiving and preservation

The screenshot shows the NEEShub website interface. At the top, there are navigation options for '6 months' and '12 months'. Below the header, there's a section for 'End of experiment' and a table with columns 'Mark' and 'Pr'. A CD-ROM labeled 'Backup' is shown in the center. To the right, there's a 'Notes' section with a table of project details. At the bottom, there are two blue callout boxes: 'Dublin Core' and 'PREMIS'. A large blue arrow points from the 'Backup' CD towards the 'Preservation' text.

Mark	Pr		
1A	Title		
1B	Short		
1C	Princ		
	Inves		
1D	Team		
1E	Spon:		
	Number		
1F	Start and End Dates	Date	Start and end date of the project.
1G	Project Description	Text	Brief description of the project. Similar to an NSF abstract (limited to 300 words).
1H	Executive Summary	PDF document	Focused summary project results and impact. Significant results are figures and plots, and
1I			

**Preservation:**  
Aiming for 20 years  
but  
uncertainty past 2014

Dublin Core      PREMIS

# How does NEES manage data?

## A few key points – NEES Specifics

Web-based, open

OPEN DATA

**Curation progress:**

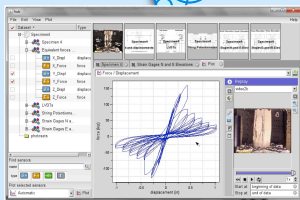
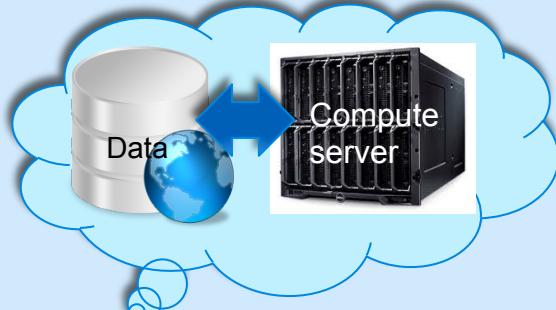
Facility	●
Equipment	●
Materials	●
Sensors	●
Drawings	●
Unprocessed_Data	●
Converted_Data	●
Corrected_Data	●
Derived_Data	●
Tags	●
Report	●

Curation

Central data repository



data-tool co-location



advanced access tools

OAIS Reference Model



serving distinct audiences

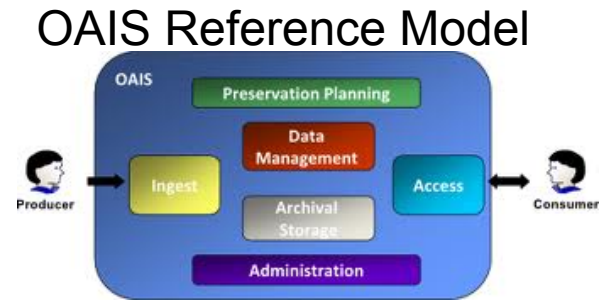


high-quality software architecture



# Lessons & Challenges

- ▶ Speak librarian language
- ▶ Data ingest is hard
  - Reliable transfers, data reorganization, complete metadata
  - Need “carrots and sticks”
- ▶ Data needs to become a “publication”
- ▶ Reuse needs continual encouragement





# Lessons & Challenges

- ▶ Training and assistance for data ingestion
  - I have no time to learn, but
  - I'm frustrated when it doesn't work the way I want it



- ▶ Automated curation



- ▶ Trustworthy repositories

- prior work in digital repositories, library science useful, e.g., TRAC analysis (ISO 16363)



- ▶ Adding preservation data and mechanisms

- ▶ Modest data now, but HD video adds a lot

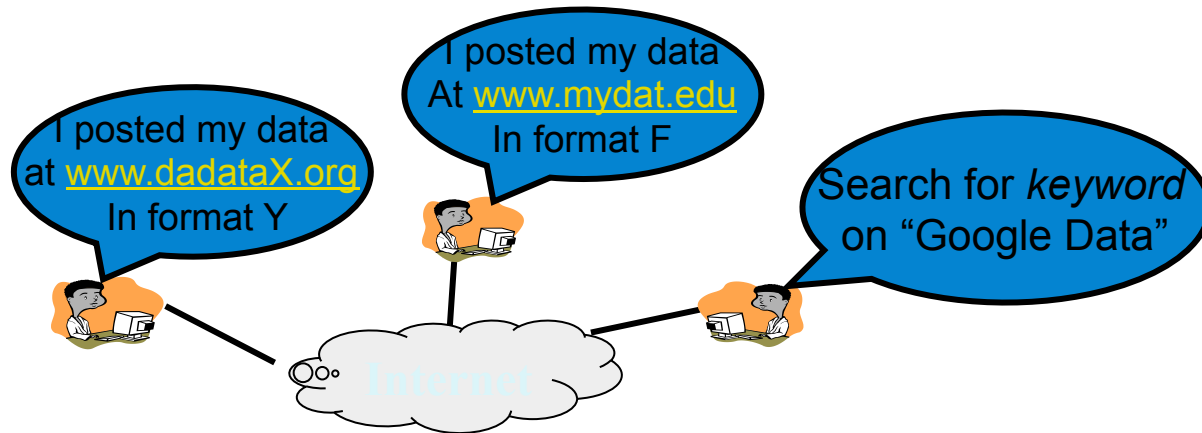
## Data Preservation



# Is Synergy Possible? – A Vision

Data Discovery: Imagine scientists could *discover* the existence of data in diverse fields. Imagine what they might do.

- ▶ Important feature: Force no one to change their formats and best practices.



- Interoperability will be a much later step.
- Given the current “big data” initiatives – there may even be funding for this.

# Conclusions – Take Home points

- ▶ DM is an essential part of facilities access
- ▶ NEES project maintains a data repository for earthquake engineering community
- ▶ Lessons& Challenges:
  - Learn from digital libraries
  - User challenges: ingest, reuse, training
  - Curation
  - Building a trustworthy, long-term data repository
- ▶ Synergy among Large Facilities is possible

# NEESR NEESwood Project

## Wood-frame construction

Full-scale, 3D, 2-story  
townhouse tested at U.  
Buffalo

Full-scale, 3D, 6-story 40 x  
60 ft condominium tested  
at E-defense

Light-frame wood buildings  
can be engineered to  
perform well in large  
earthquakes

NEESR-SG: PI John van de Lindt  
Colorado State U., U. Buffalo, E-  
defense





# NEES: An Essential National Resource



## NEES Cyberinfrastructure

A screenshot of the NEEShub website. The header includes the NEEShub logo, navigation links (Tools &amp; Resources, Learning &amp; Outreach, Project Warehouse, Sites, Collaborate, Explore), and a search bar. The main content area features a large banner for "EDUCATION" and several sections: "In the Spotlight" with links to data turbine and simulation tools; "Use NEEShub to..." with bullet points for accessing projects, running simulators, learning with data, and sharing research; "How-To Videos" and "NEES Videos on YouTube" with thumbnail images. Below this is the "NEES Project Warehouse" section, which includes a search bar, a description of the repository, and a "Search Project Warehouse" form with a "Keywords" field and "Advanced Search" link. Popular tags like "earthquake", "tsunami", and "steel frame" are listed at the bottom.

## NEES Laboratories



## Community

