

Milliseconds to Decades: Developing a Data Life Cycle Cyberinfrastructure for Scientific Data

Dr. Thomas Hacker

Co-Leader for Information Technology, The George E. Brown, Jr.

Network for Earthquake Engineering Simulation

Assistant Professor, Computer and Information Technology

Purdue University, West Lafayette, IN

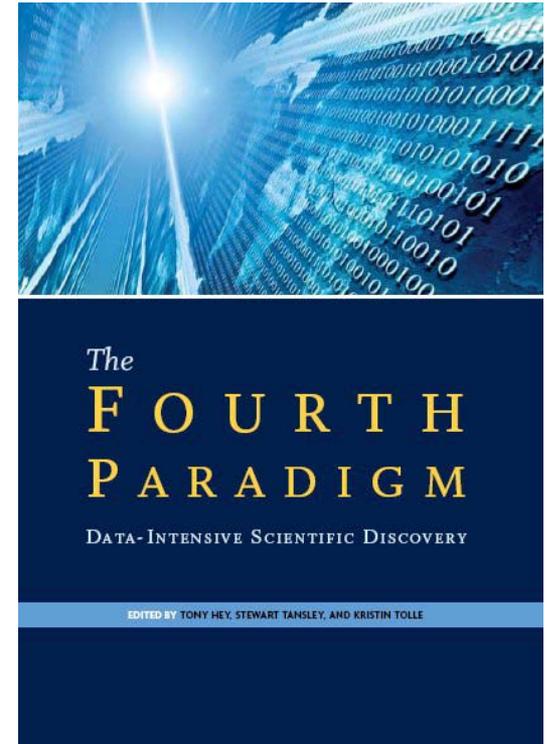
Fermi National Laboratory

January 7, 2011



21st Century Science

- ▶ Theory and Experiment
- ▶ Computational Simulation
 - Third leg of science
 - Past 50 years or so
- ▶ Data
 - Fourth “leg” of science
 - Researchers are flooded with data
 - Tremendous quantity and multiple scales of data
 - Difficult to collect, store, and manage
 - How can we distill meaningful knowledge from data?



Data is the 4th Paradigm

- ▶ Producing an avalanche of high resolution digital data
- ▶ All (or most) of the data needs to be accessible over a long period of time
 - Much of the data can is not reproducible
- ▶ Example – NEES project
 - Structure or sample destroyed through testing
 - Very expensive to rebuild for more tests



NSF Network for Earthquake Engineering Simulation (NEES)

- ▶ Safer buildings and civil infrastructure are needed to reduce damage and loss from earthquakes and tsunamis
- ▶ To facilitate research to improve seismic design of buildings and civil infrastructure, the National Science Foundation established NEES
- ▶ NEES Objectives
 - Develop a national, multi-user, research infrastructure to support research and innovation in earthquake and tsunami loss reduction
 - Create an educated workforce in hazard mitigation
 - Conduct broader outreach and lifelong learning activities

NEES Research Facilities

- ▶ NEES has a broad set of experimental facilities
 - Each type of equipment produces unique data
 - Located at 14 sites across the United States
- ▶ Shake Table
- ▶ Tsunami Wave Basin
- ▶ Large-Scale Testing Facilities
- ▶ Centrifuge
- ▶ Field and Mobile Facilities
- ▶ Large-Displacement Facility
- ▶ Cyberinfrastructure

Oregon State University

University of Minnesota

University of Illinois

University of California Berkeley

<https://www.nees.org>

University of California Davis

University of Buffalo

University of California Santa Barbara

Cornell University

University of California Los Angeles

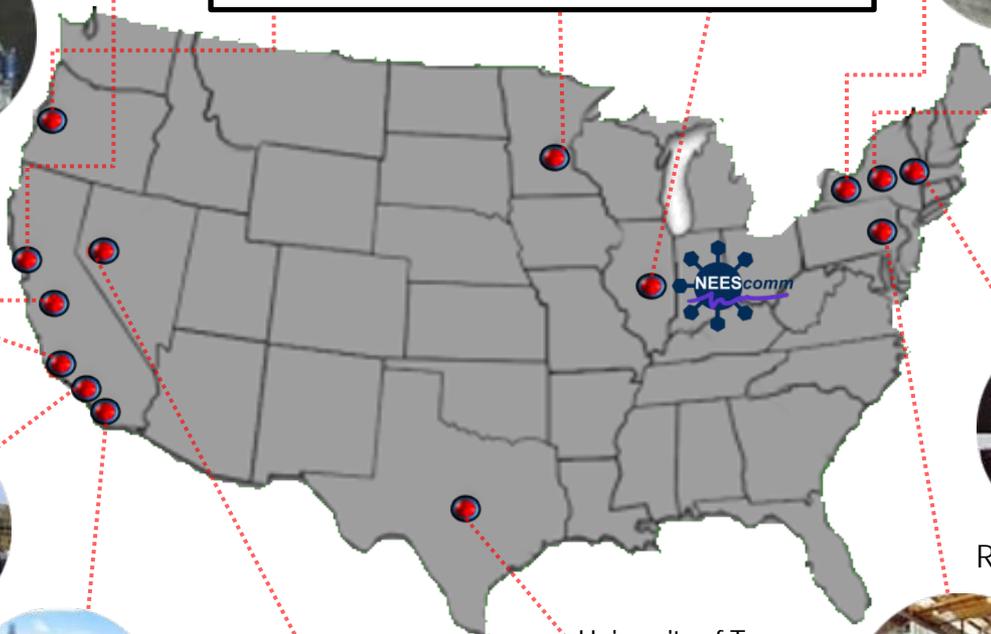
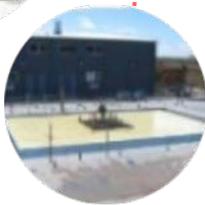
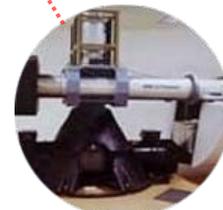
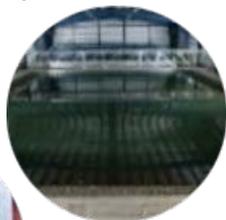
Rensselaer Polytechnic Institute

University of California San Diego

University of Nevada Reno

University of Texas Austin

Lehigh University



Shake Table

- ▶ Each earthquake is unique
 - Ground motion unique for each earthquake
 - Up-and-down, back-and-forth, brief sharp strike
 - Strength varies as well
- ▶ Shake table
 - Recreates the movement of the ground
 - Five degrees of freedom
 - Computer system controls table motion
- ▶ Simulate ground motion
 - Ground motion measurements from real earthquakes
 - Synthetic ground motion from researchers

NEES Shake Tables

- ▶ University of Buffalo, SUNY
 - Two shake tables
- ▶ University of Nevada, Reno
 - Three bi-axial shake tables
 - Can function in unison

Six degree of freedom table at University of Nevada, Reno



Image: Univ of Nevada, Reno

NEES Shake Tables

- ▶ University of California, San Diego
 - Largest outdoor shake table in the U.S



Source: UCSD

Tsunami Wave Basin

- ▶ Tsunamis carry tremendous stored energy and can travel across very long distances
 - Can reach very high vertical heights once they reach shore
- ▶ Oregon State University operates the world's largest facility for studying the effects of large waves – the Tsunami Wave Basin



Image: Oregon State Univ

Tsunami Wave Basin

- ▶ 49m long, 27m wide, 2m deep
- ▶ 30–segment wave generator on basin end
 - Can form periodic and random waves
- ▶ Floor of basin
 - Slopes up
 - Researchers can modify the bathymetry of basin floor
- ▶ High definition and high speed cameras to collect data

Oregon State Tsunami Wave Basin



Large-Scale Testing Facilities

- ▶ Researchers also want to study effects of forces on specific structural components
 - Shake tables emulate ground motion
- ▶ NEES operates six large-scale testing facilities
 - Can test full-scale columns, beams, and other structural components

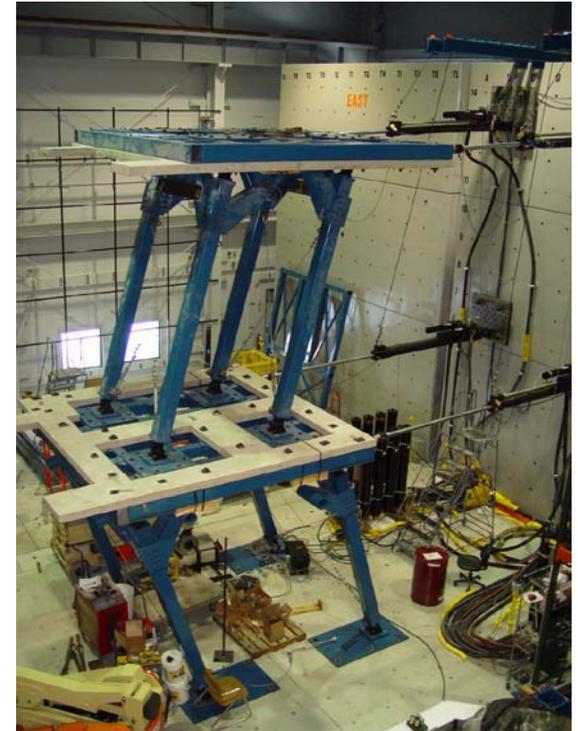
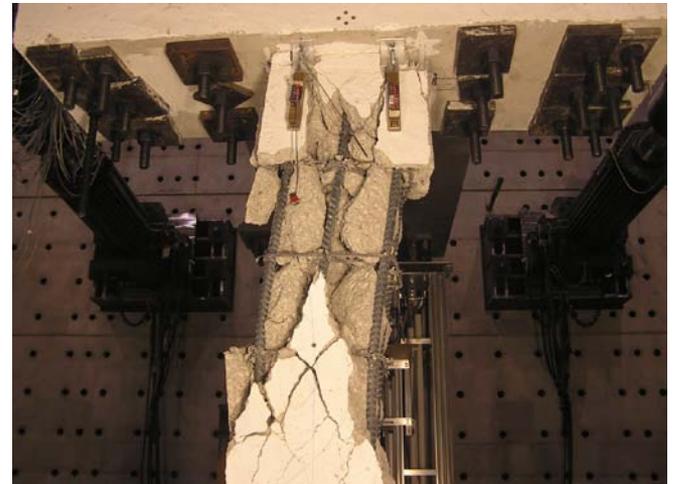


Image: Univ of Buffalo

Large-Scale Testing Facilities

- ▶ Lehigh University
 - Reaction wall, strong floor
 - dynamic actuators
- ▶ UC-Berkeley
 - Reconfigurable Reaction Wall
- ▶ University of Illinois Urbana-Champaign
 - Multi-Axial Full-Scale Sub-Structured Testing & Simulation (MUST-SIM)
- ▶ University of Minnesota
 - Reaction walls
 - Multi-Axial Subassemblage Testing (MAST)



Large-Scale Testing Facilities

- ▶ Cornell University
 - Large Displacement Facility
- ▶ Designed to simulate the lateral displacement and compressive deformation that occurs during fault rupture
 - Type of damage that can disrupt critical services like water and gas pipelines
- ▶ Also can be used to assess seismic performance of highly ductile above-ground structures
- ▶ More information: <http://nees.cornell.edu>

Cornell Large-Displacement Facility



Image source:
Cornell

Geotechnical Centrifuge Facilities

- ▶ Need to understand the interaction of geological materials (dirt and rock) with structures
- ▶ Geotechnical centrifuges
 - Used to study the effects of gravity on small-scale models of structures
- ▶ University of California, Davis
 - 9m 80G centrifuge
- ▶ Rensselaer Polytechnic Institute
 - 3m centrifuge

UC Davis Centrifuge



Image source: UC Davis

Mobile and Field Facilities

- ▶ Researchers also need to conduct experiments in-situ
- ▶ UCLA
 - Mass shakers for measuring structure response to vibration
- ▶ University of California at Santa Barbara
 - Monitoring facilities instrumented to measure data from real earthquakes when they occur



Mobile and Field Facilities

- ▶ University of Texas at Austin
 - Operates several Mobile Shakers
- ▶ T-Rex (Tri-Axial)
 - Buggy-mounted vibrator
 - 64,000 lbs (29,030 kg)
- ▶ Liquidator
 - Low-frequency vibrator to study soil liquefaction
 - 70,000 lbs



Image source: University of Texas, Austin

NEES Cyberinfrastructure

- ▶ Provide a reliable, robust, and responsive IT infrastructure for the NEES network
 - Curated research data repository
 - Telepresence and teleoperation tools and services
 - Data visualization tools
 - Tools for hybrid and multi-site hybrid simulation
 - Simulation tools for earthquake engineering research and education
 - Access to national supercomputing resources
 - Middleware and cybersecurity

NEES Project

- ▶ Tremendous amounts and variety of data
 - Type of equipment at experimental facility site
 - E.g. shake table, wave basin
 - Unique to individual experiment
 - NEES sites utilize local data standards
- ▶ NEEShub
 - Cyberinfrastructure to support the upload and dissemination of data
 - A first step towards a comprehensive data management system

NEES Data

- ▶ The most enduring asset of the NEES network
- ▶ Many forms of data
 - Numerical data from sensors, simulation, and analysis output
 - Visual data from photographs, images, and figures
 - Video data from high definition cameras
 - Reports, papers, documentation, and calibration information
- ▶ Data contributed from the community
 - Lots of useful data available for researchers and practitioners
 - Encourage practitioners, researchers, and international contributions
- ▶ Example of the types of data from a project
 - Collaborative Research: Using NEES as a Testbed for Studying Soil–Foundation–Structure Interaction, PI Saiidi
 - Stored in NEEShub

Raw sensor data

4b-1 - Microsoft Excel

Home Insert Page Layout Formulas Data Review View Add-Ins Acrobat

Clipboard Font Alignment Number Styles Cells Editing

Normal Good Bad Neutral

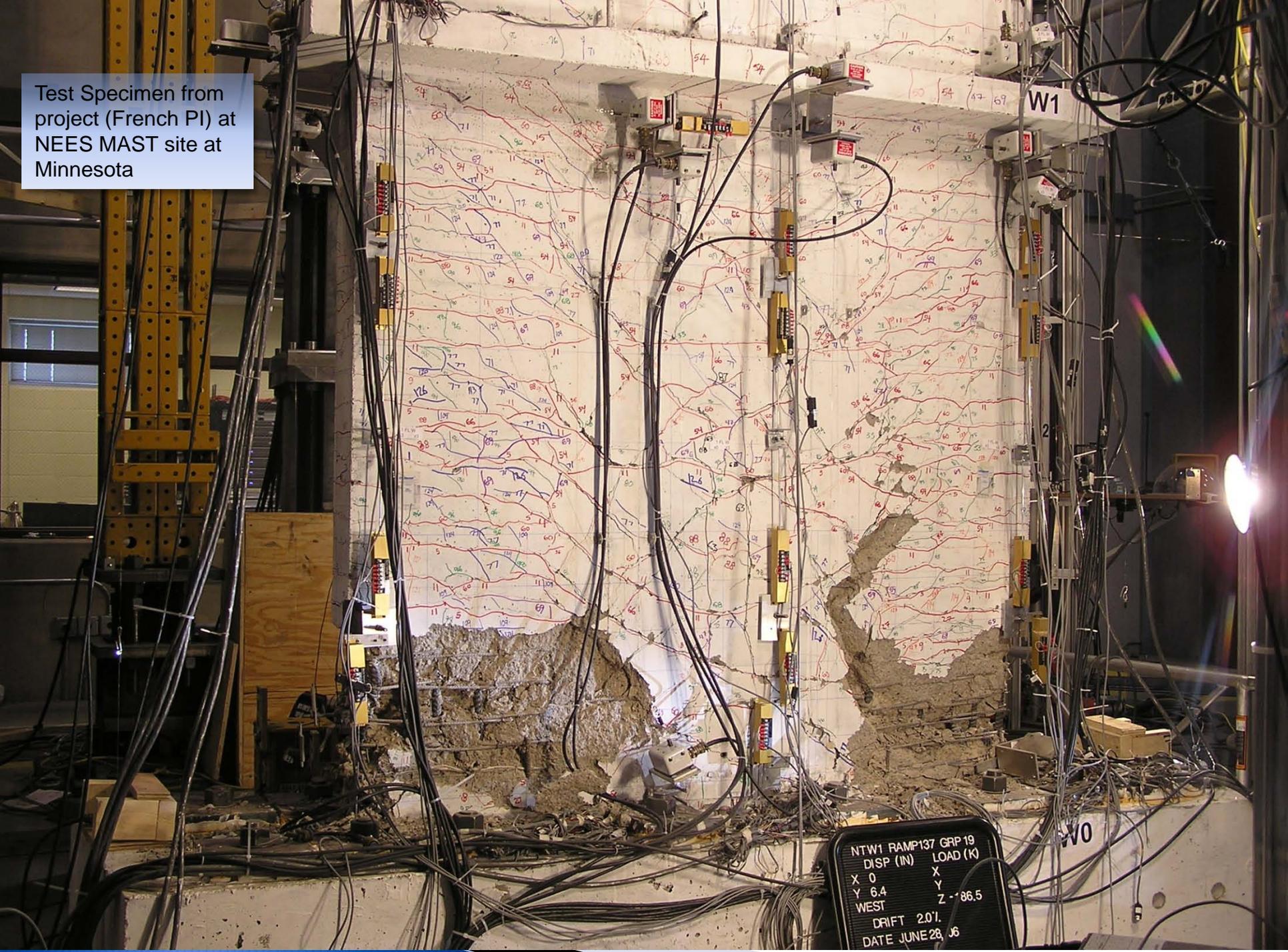
AutoSum Fill Clear Sort & Filter Find & Select

A1 4b, Transverse Direction

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	4b, Transv	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
2	Channel N	c0	c1	c2	c3	c4	c5	c6	c7	c8	c9	c10	c11	c12	c13	c14	c15	c16	c17	c18	c19
3	Units	V	in	in	in	in	in	in	in/sec	in/sec	in/sec	in/sec	in/sec	in/sec	g	g	g	g	g	g	V
4	Description	DAQ_Sync	Table1_dis	Table1_dis	Table2_dis	Table2_dis	Table3_dis	Table3_dis	Table1_ve	Table1_ve	Table2_ve	Table2_ve	Table3_ve	Table3_ve	Table1_ac	Table1_ac	Table2_ac	Table2_ac	Table3_ac	Table3_ac	FunGe
5	Location	asdrtrPCI6	493AO	493AO	493AO	493AO	493AO	493AO	493AO	493AO	493AO	493AO	493AO	493AO	493AO	493AO	493AO	493AO	493AO	493AO	asdrtr
6	8.57	0.00061	0.48193	-0.53363	0.05554	-0.01709	0.92071	1.4458	0	-0.00183	0	0.00366	-0.00366	-0.00183	0.00092	-0.0003	0.00061	0.00061	0.00061	0.00061	-4.5
7	8.58	0.00092	0.48151	-0.53448	0.05682	-0.01709	0.92071	1.44623	0	0.00366	0	-0.00549	0.00366	0.00366	0.0003	-0.0003	-0.0003	0.00092	0	0.00092	-4.5
8	8.59	0.0003	0.48236	-0.53406	0.05597	-0.01794	0.922	1.44623	0.00183	0.00183	-0.00183	0.00366	0.00549	0.00366	0.00061	-0.0003	0.00061	0.00092	-0.00061	0.00092	-4.5
9	8.6	0.00122	0.48108	-0.53448	0.05511	-0.01709	0.92071	1.44623	0.00183	0.00183	0.00366	-0.00183	0.00183	0.00366	0.00061	-0.0003	-0.0003	-0.00061	0	-0.0003	-4.5
10	8.61	0.00061	0.48151	-0.53448	0.0564	-0.01709	0.92029	1.44537	0	0.00183	0	0.00366	0.00366	0.00366	0.0003	0.00061	-0.0003	0.00305	-0.00092	0.00061	-4.5
11	8.62	0.00092	0.48193	-0.53406	0.05597	-0.01709	0.92071	1.44623	-0.00183	0.00549	0	0	0.00549	-0.00183	0.0003	-0.00061	0.00061	0.00061	0.00061	0.00061	-4.5
12	8.63	0.00061	0.48108	-0.53448	0.05597	-0.01794	0.92114	1.44537	0.00366	0.00366	0	-0.00183	0.00183	0.00732	0.0003	0.00061	-0.00061	0.00092	0.0003	0.00092	-4.5
13	8.64	0	0.48108	-0.53406	0.05597	-0.01794	0.92071	1.44537	0.00366	0.00183	0	0	-0.00183	-0.00183	0.00061	-0.00061	0.0003	0.00244	-0.00122	0.0003	-4.5
14	8.65	0.0003	0.48108	-0.53448	0.05597	-0.01837	0.92071	1.44623	0	0.00366	0	-0.00183	0.00366	0	0.00061	-0.0003	-0.0003	-0.00061	0	-0.0003	-4.5
15	8.66	0.00061	0.48065	-0.53406	0.05597	-0.01709	0.92114	1.44623	0.00183	0.00183	0	0.00366	0.00366	-0.00183	0.0003	0.00061	0.00061	0.00305	-0.00122	0.00061	-4.5
16	8.67	0.00061	0.48193	-0.53406	0.05597	-0.01709	0.92071	1.44495	0	0.00366	0	-0.00183	0.00183	0.00366	0.00061	-0.0003	-0.0003	-0.0003	0.00061	-0.0003	-4.5
17	8.68	0.0003	0.48151	-0.53448	0.05597	-0.01837	0.92114	1.4458	0	0.00366	0	0.00366	0.00183	0.00183	0.00061	0.00061	0	0	0	0.0003	-4.5
18	8.69	0.00122	0.48151	-0.53406	0.05597	-0.01837	0.92071	1.4458	-0.00183	0.00366	0	-0.00366	0.00366	-0.00183	0.00061	-0.0003	-0.00122	0.00305	-0.00061	0.00244	-4.5
19	8.7	0.00061	0.48193	-0.53406	0.05597	-0.01709	0.92029	1.44623	0.00366	0.00366	0	0.00366	0.00366	0.00366	0.00092	-0.0003	0.00061	-0.00092	0	-0.0003	-4.5
20	8.71	0.00061	0.48236	-0.53406	0.05597	-0.0188	0.92071	1.4458	0	0.00366	0	0.00183	-0.00183	0.00366	0.00061	-0.0003	-0.00061	0.00305	-0.00122	0.00061	-4.5
21	8.72	0.00061	0.48193	-0.53363	0.05597	-0.01837	0.92114	1.4458	-0.00183	0.00366	0	-0.00549	0.00549	0	0.00061	-0.0003	0.00061	-0.0003	-0.0003	-0.0003	-4.5
22	8.73	0.00092	0.48193	-0.53406	0.05682	-0.01794	0.92114	1.44666	0.00366	0.00366	-0.00183	0.00732	0	0.00183	0.00092	-0.0003	-0.0003	0.00092	-0.0003	0	-4.5
23	8.74	0.0003	0.48065	-0.53448	0.0564	-0.01794	0.92071	1.44623	0.00183	-0.00183	0.00366	-0.00183	0.00366	-0.00183	0.0003	-0.0003	0	0.00061	-0.00061	0.00061	-4.5
24	8.75	0.00061	0.48193	-0.53406	0.05554	-0.01837	0.92114	1.44623	-0.00549	0.00366	0	0.00549	0.00183	0.00366	0.00061	-0.0003	-0.00061	-0.00092	0.00061	-0.0003	-4.5
25	8.76	0.00061	0.48279	-0.5332	0.05597	-0.01837	0.92157	1.4458	0.00366	0.00183	0	0.00183	0.00183	0.00183	0.00061	-0.0003	-0.00061	0.00305	-0.00122	0.00061	-4.5
26	8.77	0.00061	0.48108	-0.53363	0.05597	-0.01837	0.92029	1.4458	0	0.00183	0.00183	0.00183	-0.00183	0.00183	0.00061	-0.0003	0.00061	-0.0003	-0.0003	-0.0003	-4.5
27	8.78	0.00061	0.48065	-0.53363	0.05597	-0.01837	0.92114	1.4458	-0.00549	0.00183	0	0.00183	0.00366	0.00366	0.00061	-0.0003	0.0003	0.00061	-0.0003	0	-4.5
28	8.79	0.00061	0.48193	-0.5332	0.0564	-0.01709	0.92071	1.4458	0	0.00183	0	-0.00183	0.00366	-0.00183	0.0003	0.00061	-0.00061	0.00061	-0.00061	0.00244	-4.5
29	8.8	0.0003	0.48108	-0.53406	0.05554	-0.01837	0.92114	1.4458	-0.00549	-0.00183	0	0.00732	-0.00183	0.00732	0	-0.0003	0.00122	-0.00061	0.00061	-0.0003	-4.5
30	8.81	0.00092	0.48108	-0.53406	0.05597	-0.01794	0.92114	1.4458	-0.00183	0.00366	0.00183	0.00366	0.00366	-0.00183	0.00122	0.00061	-0.00092	0.00305	-0.00092	0.00061	-4.5
31	8.82	0.00061	0.48065	-0.53448	0.05554	-0.01794	0.92029	1.44623	-0.00183	0.00183	0.00183	0.00183	0.00183	0.00183	0.00366	0.0003	-0.00061	0.00061	-0.0003	0.00061	-4.5
32	8.83	0.00061	0.48108	-0.53406	0.05597	-0.01794	0.92071	1.44623	-0.00183	-0.00183	0	0.00366	0.00183	0.00183	0.0003	0.00061	-0.00092	0	0	0	-4.5
33	8.84	0.0003	0.48193	-0.53406	0.05597	-0.01837	0.92157	1.4458	-0.00366	0.00732	0.00183	-0.00366	0	0.00061	-0.0003	-0.0003	0.00275	-0.00092	0.00061	-4.5	

Ready

Test Specimen from project (French PI) at NEEs MAST site at Minnesota





Network for Earthquake Engineering Simulation (NEES) is a shared national network of 14 experimental facilities, collaborative tools, a centralized data repository, and earthquake simulation software. Together, these resources provide the means for collaboration and discovery in the form of advanced research based on experimentation and computational simulations of the ways buildings, bridges, utility systems, coastal regions, and geomaterials perform during seismic events. [Learn more](#)



In the Spotlight

- Specimen Authoring Tool for inDEED:** Define the specimen views for your experiment with the types, measurement ... - in Tools
- FlexTpsSource:** FlexTpsSource acquires data from a FlexTPS video stream and sends it to the ... - in Tools
- EMERSE:** Enhanced Messaging for the Emergency Response Sector - in Multimedia

Use NEEShub to...

- Access NEESR Proposal resources
- Access NEES projects- Project Warehouse
- Run simulators and other Tools
- Learn with earthquake data and simulators - NEES Academy

How-To Videos



NEES Videos on YouTube



Events and Activities

View All | Submit an event

JANUARY

S	M	T	W	T	F	S
						1
2	3	4	5	6	7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29

News and Announcements

View All | Subscribe to Newsletters

NEES (NEES) Platform for Cyber Collaboration
21 Dec 2010
"(NEES) Platform for Cyber Collaboration," a workshop on NEEShub usage, will be presented by ...

UCLA NEES@UCLA Team Conducts Trio of Global Field Studies
21 Dec 2010
NEES@UCLA Team Conducts Trio of Global Field Studies UCLA shake table test at E-Defense lab in ...

Latest Earthquake Reports

View All

JAN 02 M 5.7, southern Sumatra, Indonesia
Sunday, January 2, 2011 15:19:31 UTC
Sunday, January 2, 2011 10:19:31 PM at epicenter
Depth: 21.50 km (13.36 mi)

JAN 02 M 5.0, Loyalty Islands, New Caledonia
Sunday, January 2, 2011 09:32:55 UTC
Sunday, January 2, 2011 08:32:55 PM at epicenter
Depth: 24.20 km (14.95 mi)

NEES Project Warehouse

- ▶ Tremendous amount of data, in many different forms
 - Difficult for a researcher to navigate and distill meaning from data
 - Project Warehouse seeks to simplify data navigation
- ▶ Summarizes the key aspects of an experiment
 - Allows end user to decide whether the experiment is of value to them.
 - Supports the visualization and inspection of data *directly in the repository*
 - No need to download large data amounts the user's computer
- ▶ Provides a simple search and review interface
- ▶ Each project in the Warehouse has an overview page
 - Abstract of project
 - Tabs to quickly navigate to experiments, team members, and data
 - Data (PDF, excel, text) is stored in files

NEES Project Warehouse

NEES - Network for Earthquake Engineering Simulation - Windows Internet Explorer

https://nees.org/warehouse/featured

NEES - Network for Earthquake Engin...

NEEShub
George E. Brown, Jr. Network for Earthquake Engineering Simulation

1379 New Messages
Thomas Hacker (thacker)

Tools & Resources | Learning & Outreach | Project Warehouse | Sites | Collaborate | Explore | Support

You are here: Home » Project Warehouse

NEES Project Warehouse

Search Enhanced Projects

Quick Links to Enhanced Projects

- [21 - Semiactive Control of Nonlinear Structures](#)
- [22 - Non-rectangular Walls](#)
- [27 - Sllickensided Surfaces](#)
- [32 - SFSI Testbed: Shaking Table Tests](#)
- [33 - Bridge Systems with Conventional and Innovative Materials](#)
- [42 - Passive Pressure on Pile Caps](#)
- [78 - High-Strength-Concrete Structural Walls](#)
- [84 - Sidesway Collapse of Deteriorating Structural Systems](#)
- [116 - Damage-Tolerant Slab-Column Frame Systems](#)
- [138 - Full-Scale Two-Story Wood Building](#)
- [180 - SFSI Testbed: Centrifuge Tests](#)
- [411 - Elastomeric Structural Damper](#)
- [412 - Landslides in Cohesive Slopes](#)
- [414 - Using the NEES Field Shakers to Induce Liquefaction at Previous Liquefaction Sites](#)
- [625 - Full-Scale RC Flat-Plate Structure](#)
- [637 - Vulnerable Concrete Buildings](#)
- [665 - Wave Loading on Residential Structures](#)
- [672 - Deep Shear Wave Velocities in the Las Vegas Basin](#)
- [711 - Earthquake Performance Evaluation Tool \(EPET\)](#)
- [904 - Concrete Columns Reinforced with High-Strength Steel](#)
- [905 - Shear Strength Decay in Reinforced Concrete Columns](#)
- [915 - 10 Story Reinforced Concrete Walls](#)

https://nees.org/home Internet | Protected Mode: On 100%

You are here: [Home](#) » [Resources](#) » [Tools](#) » [inDEED](#) » [Session: 8072 "inDEED"](#)

inDEED

File Edit View Plot
Help

	Name	Type
<input type="checkbox"/>	Table1_vel_y	veloci
<input type="checkbox"/>	Table2_acc_x	accel
<input type="checkbox"/>	Table2_acc_y	accel
<input type="checkbox"/>	Table2_disp_x	displa
<input type="checkbox"/>	Table2_disp_y	displa
<input type="checkbox"/>	Table2_vel_x	veloci
<input type="checkbox"/>	Table2_vel_y	veloci
<input type="checkbox"/>	Table3_acc_x	accel
<input type="checkbox"/>	Table3_acc_y	accel
<input checked="" type="checkbox"/>	Table3_disp_x	displa
<input type="checkbox"/>	Table3_disp_y	displa
<input type="checkbox"/>	Table3_vel_x	veloci
<input type="checkbox"/>	Table3_vel_y	veloci
<input type="checkbox"/>	T2YFrameAccel	accel
Transverse SGs_Bent1		

Find sensors

name

type a→ δ→ ←ε→ ↑ V→ ↻

Plot selected sensors

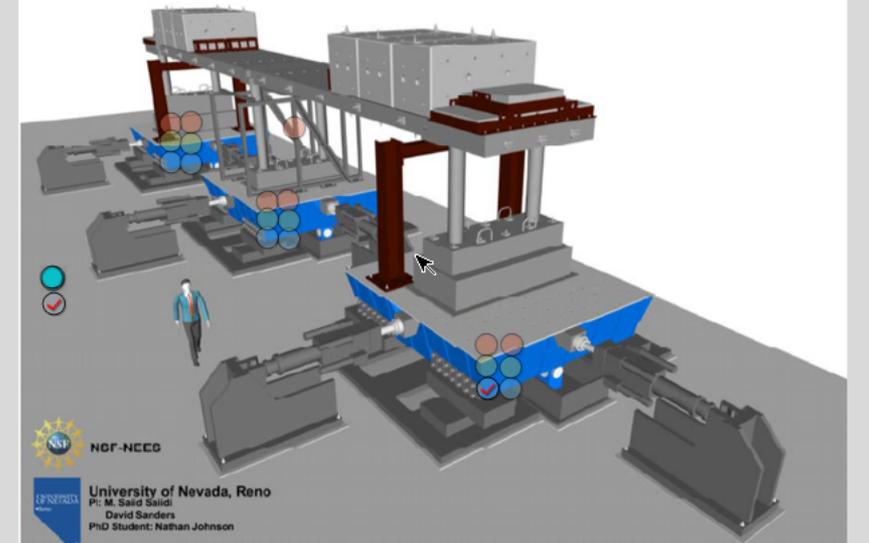
Automatic Plot

Accelerometers

Shake table

Column curvature g...

Plot



NSF-NECCO
 University of Nevada, Reno
 PI: M. Saïd Saïdi
 David Sanders
 PhD Student: Nathan Johnson



Images courtesy of Ann Catlin, Purdue

NEES - Network for Earthquake Engineering Simulation - Windows Internet Explorer

https://nees.org/warehouse/project/22

Convert Select

Favorites Suggested Sites Web Slice Gallery

NEES - Network for Earthquake Engin...

Logout My Account My NEEShub About NEES Sitemap Feedback

1379 New Messages
Thomas Hacker (tjhacker)

Search

Tools & Resources Learning & Outreach Project Warehouse Sites Collaborate Explore Support

You are here: Home » Project Warehouse » NEES-2005-0022

NEES Project Warehouse

Experimental and Analytical Investigation of Non-rectangular Walls under Multidirectional Loads OPEN DATA

Project Experiments Team Members File Browser

warehouse search

PI(s): [Catherine French](#)

Dates: May 16, 2005 - October 31, 2008

Facility: [University of Minnesota](#)

Organization(s): Iowa State University, University of Minnesota

Description: Specimens to be tested include 1 half-scale, four story T-shaped wall; 1 three-quarter scale, two story T-shaped wall; and 3 half-scale, four story rectangular walls. This testing will be conducted in the MAST laboratory at the University of Minnesota.

Sponsor: NSF - 0324504

Website(s): NEES@Minnesota ([view](#))
[NEES University of Minnesota T-wall Experiment \(YouTube\)](#) ([view](#))

Equipment: [View Details](#)

Tools: inDEED

Publications: [Benton Johnson](#), "Anchorage Detailing Effects on Lateral Deformation Components of R/C Shear Walls" ([view](#))
[Beth Brueggen](#), "Performance of T-shaped Reinforced Concrete Structural Walls under Multi-Directional Loading" ([view](#))

Documentation: [view](#)



Testing of a half-scale T-shaped structural wall.

484 Views

0 Downloads

Curation progress:
Discovered existence of final report so completed curation and republished the PRPP.

Error on page. Internet | Protected Mode: On 100%

You are here: Home » Project Warehouse » NEES-2005-0022 » File Browser

NEES Project Warehouse

Experimental and Analytical Investigation of Non-rectangular Walls under Multidirectional Loads

Project Experiments Team Members File Browser

Tremendous number and variety of files

Location: / [NEES-2005-0022](#)

<input type="checkbox"/>	Name	Size	Timestamp	Application
<input type="checkbox"/>	Analysis (0 files)			
<input type="checkbox"/>	Documentation	31 MB		
<input type="checkbox"/>	Experiment-2 : RWN - continuous reinforcement	40 GB		
<input type="checkbox"/>	Experiment-3 : RWC - couplers	40.6 GB		
<input type="checkbox"/>	Experiment-4 : RWS - splices	31.5 GB		
<input type="checkbox"/>	Experiment-5 : NTW2	98.4 GB	2008-01-16 09:15:37	
<input type="checkbox"/>	Experiment-6 : NTW1	114.8 GB	2008-03-11 14:01:45	
<input type="checkbox"/>	Public (0 files)			

Download Approximate Download File: 0 b (max is 100 MB)

Find by
Title
Search:
GO

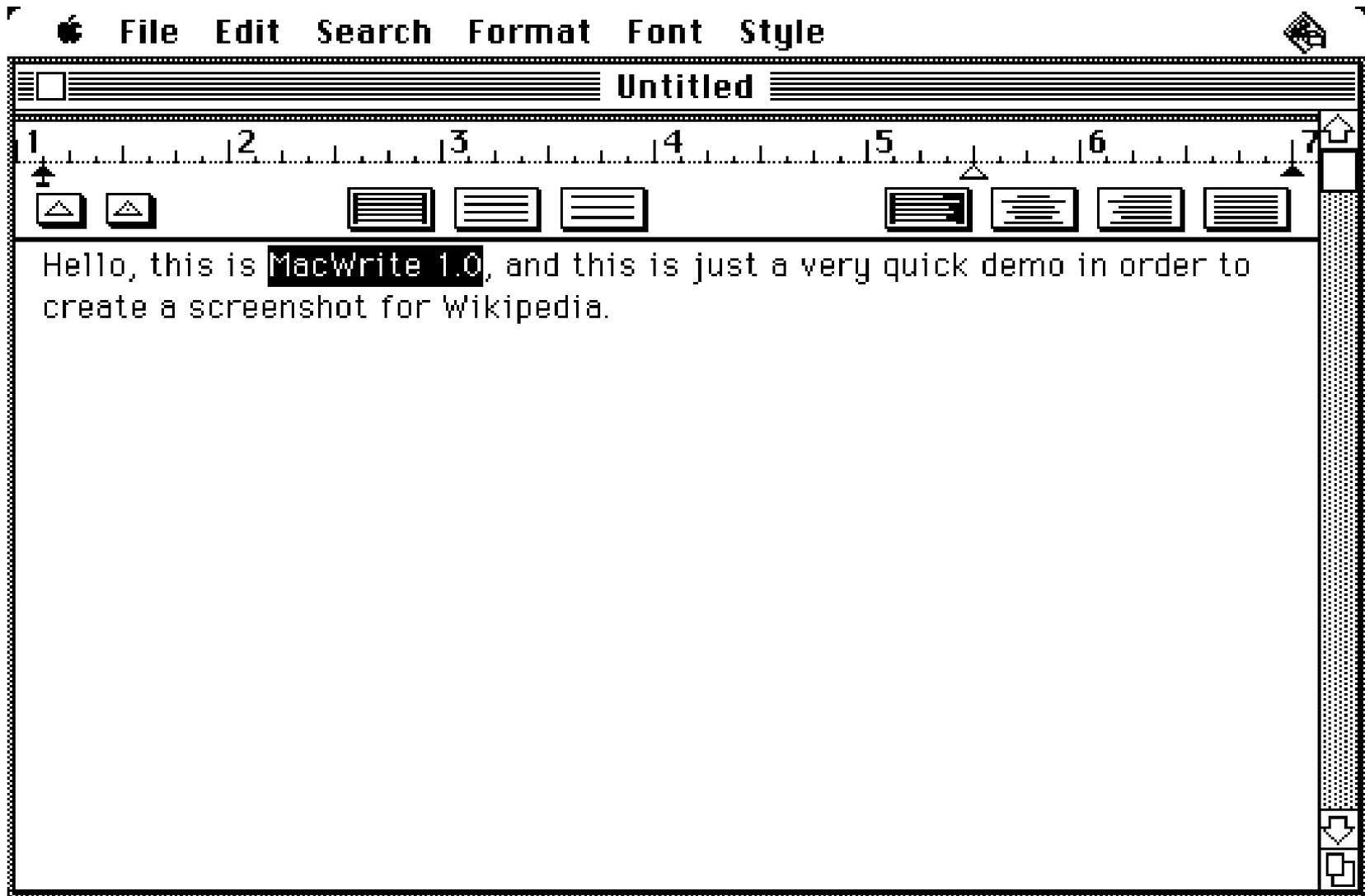
- File By Type
- [Charts \(0\)](#)
 - [Data Files \(6135\)](#)
 - [Drawings \(75\)](#)
 - [Photos \(5995\)](#)
 - [Presentations \(0\)](#)
 - [Publications \(0\)](#)
 - [Reports \(5\)](#)
 - [Videos \(0\)](#)

Many data types and formats

- ▶ Depending on the source of data
 - Sensor data
 - Photos
 - Calibration data
 - Papers
 - Engineering drawings
 - Field notes and comments
 - RDV
- ▶ What about binary data?
 - Many data formats (e.g. JPEG) store data in binary format.
- ▶ Will data encoding still be readable and valid 20 years from now?

Example of the problem

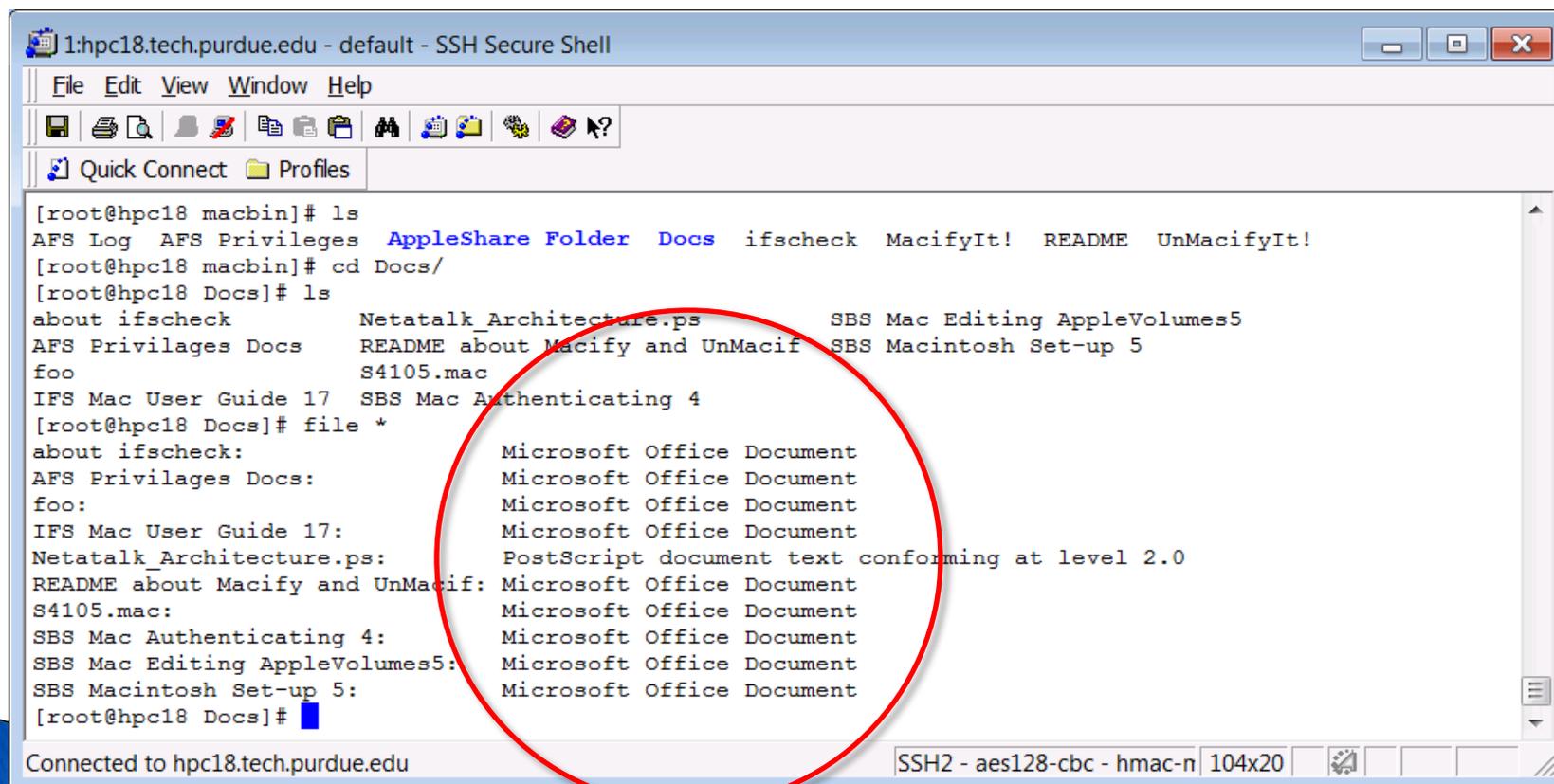
- ▶ I have access to files I wrote in 1992 for the University of Michigan Institutional File System (IFS) project
 - Files are still stored in IFS at Michigan
- ▶ Kernel code written for IBM AIX to implement EtherTalk/AppleTalk for UNIX – provided AppleShare from a UNIX file system
- ▶ Being a diligent developer, I wrote documentation for the software using state-of-the-art word processing software



Source: Wikipedia

What does Linux think the files are?

- ▶ % file *



```
1:hpc18.tech.purdue.edu - default - SSH Secure Shell
File Edit View Window Help
Quick Connect Profiles

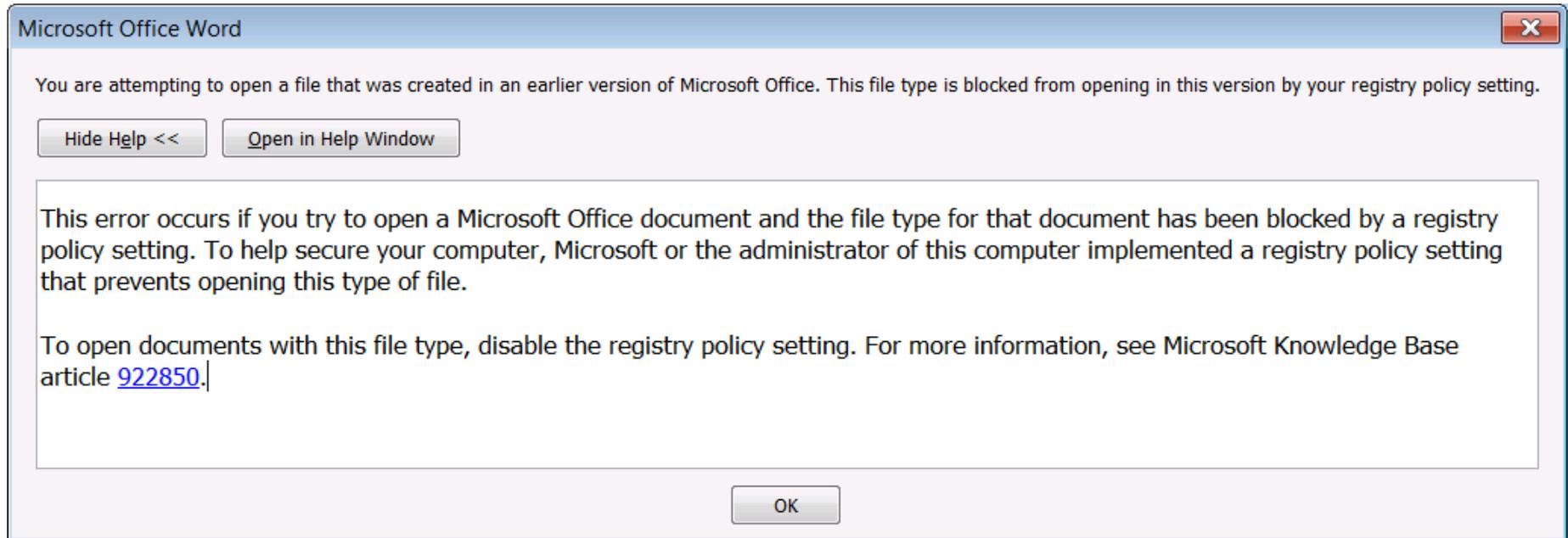
[root@hpc18 macbin]# ls
AFS Log  AFS Privileges  AppleShare Folder  Docs  ifscheck  MacifyIt!  README  UnMacifyIt!
[root@hpc18 macbin]# cd Docs/
[root@hpc18 Docs]# ls
about ifscheck          Netatalk_Architecture.ps          SBS Mac Editing AppleVolumes5
AFS Privilages Docs    README about Macify and UnMacif  SBS Macintosh Set-up 5
foo                    S4105.mac
IFS Mac User Guide 17  SBS Mac Authenticating 4
[root@hpc18 Docs]# file *
about ifscheck:          Microsoft Office Document
AFS Privilages Docs:    Microsoft Office Document
foo:                    Microsoft Office Document
IFS Mac User Guide 17:  Microsoft Office Document
Netatalk_Architecture.ps:  PostScript document text conforming at level 2.0
README about Macify and UnMacif: Microsoft Office Document
S4105.mac:              Microsoft Office Document
SBS Mac Authenticating 4: Microsoft Office Document
SBS Mac Editing AppleVolumes5: Microsoft Office Document
SBS Macintosh Set-up 5: Microsoft Office Document
[root@hpc18 Docs]#
```

Connected to hpc18.tech.purdue.edu

SSH2 - aes128-cbc - hmac-n 104x20

Is it really a word document?

- ▶ Try opening in Windows, but Microsoft Office is suspicious...



NEES Data Issues

- ▶ Long-term viability of data
 - If we are going to keep all of this data, is the data alone sufficient?
- ▶ There are many issues and problems that must be resolved
 - Metadata
 - Representation
 - Relationships among data
 - Data integrity
 - Provenance and chain of custody
 - Curation processes
 - Security
 - Display of data (User Data Model)
- ▶ Publications and related work that refer to these data need to have a stable repository and identifier (DOI) that will persist and be available for a long period of time.
- ▶ How can we best manage the data life cycle for the NEES project to ensure the long-term usability and integrity of valuable NEES data?

Scientific Data Management

- ▶ Problems extend beyond a small set of scientific projects
- ▶ New NSF data management plan requirements
 - Institutions need to develop an institutional data repository and cyberinfrastructure
 - Collect, archive, and deliver digital data
- ▶ Efforts ongoing at institutions
- ▶ Purdue, Cornell, UNC, Florida, California Digital Library, Australian National Data Service
 - The California and UNC efforts seem to be the most advanced currently.

Solutions

- ▶ There are no integrated software systems today that can solve all of these problems in a generic way for a broad category of sciences
- ▶ There are many systems for a very specific science domain
 - Purdue Ionomics Information Management System (PiiMS)
 - Commercial Life Sciences Laboratory Information Management Systems (LIMS)
 - Compact Muon Solenoid (CMS) project
 - Three tier data distribution and management system

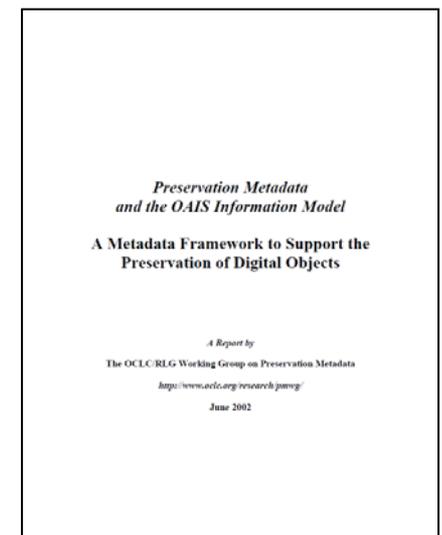
Data Cyberinfrastructure

- ▶ We need a more generic framework and model for a data cyberinfrastructure
- ▶ Fortunately the library community has been thinking about this problem
 - Defining and promoting a generic framework called ISO-OAIS
 - Provides an excellent, cogent, and clear framework and model for the collection, curation, and delivery of information

ISO-OAIS

- ▶ ISO Open Archive Information System (OAIS)
 - Created by Management Council of the Consultative Committee for Space Data Systems (CCSDS)
- ▶ Described in 2002 document by OCLC/RLG working group
- ▶ Some key terms
 - *information object* – a data object combined with “representation information” which is the structural and semantic information needed to provide meaning and interpretation of the data object.
 - *information package* – information objects combined with metadata needed for preservation, packaging, and provenance.

OAIS - “A conceptual framework for an archival system dedicated to preserving and maintaining access to digital information over the long term.”



OAIS components

- ▶ *Submission Information Package (SIP)*
 - Sent from the information producer to the archive
 - NEES upload tools that are part of SIP
 - PEN and SyncroNEES
- ▶ *Archival Information Package (AIP)*
 - Information package actually stored in the archive
 - NEES AIP
 - Curation tools used by NEES curator
- ▶ *Dissemination Information Package (DIP)*
 - Information package transferred to a user in response to an access request
 - NEES DIP
 - NEES Project Warehouse
- ▶ Clear four stage workflow
 - ingest-> archiving -> data management->dissemination

OAIS Repository Functions

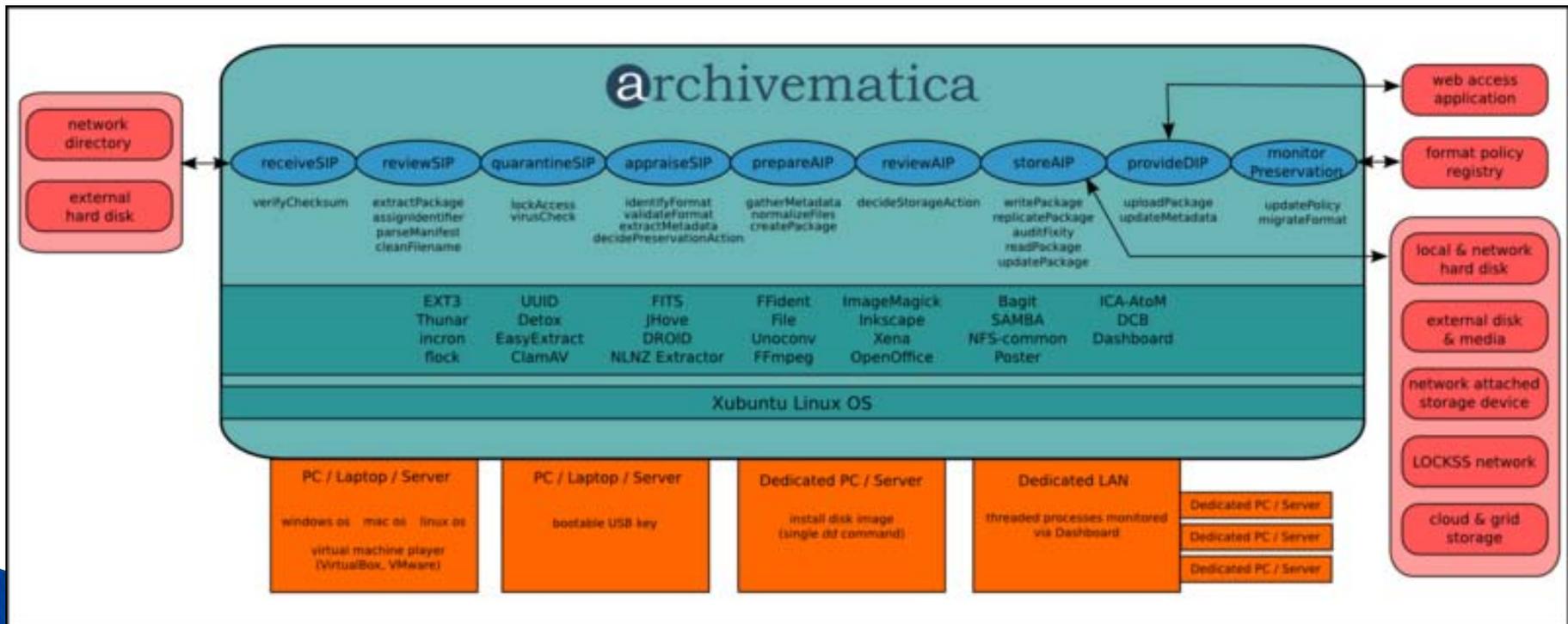
- ▶ Accept appropriate information from information producers.
 - ▶ Obtain sufficient control of the information needed to ensure Long-Term Preservation.
 - ▶ Domain user community should be able to understand the information without needing the assistance of the experts who produced the information.
 - ▶ Follow documented policies and procedures which ensure that the information is preserved against all reasonable contingencies, and which enable the information to be disseminated as authenticated copies of the original, or as traceable to the original.
 - ▶ Make the preserved information available to the domain science community
- 

Microservices

- ▶ To implement the OAIS workflow
 - Need a collection of software tools, repositories, and dissemination mechanisms
- ▶ ***Microservices*** is a new approach
 - Breaks down the OAIS workflow
 - Series of small scale tools that can be combined
 - Follows the UNIX philosophy of combining simple tools.

Archivematica

- Archivematica is an approach we are exploring for curation and to develop a more complete OAIS compliant repository in NEEShub



Source: Archivematica

Archivematica

- ▶ Open source digital preservation system
 - Developed by Artefactual Systems in British Columbia
- ▶ Uses *microservices* to provide a framework for implementing the OAIS model
 - SIP, AIP, DIP
- ▶ Integrates a broad suite of existing tools
 - Format detection
 - Virus detection
 - Metadata generation
- ▶ Creates a curatable data object and format based on the type of information
 - E.g. creates an Open Document Format (ODF) file for PPT, Word
 - Uses open source conversion tools
 - Automatic conversion as part of the microservices workflow
- ▶ Archivematica can be downloaded and installed as a virtual machine

Integration into the NEEShub

- ▶ Investigating the use of archivematica for NEES to address the data challenges
 - Metadata
 - Representation
 - Relationships among data
 - Data integrity
 - Provenance and chain of custody
 - Curation processes
 - Security
 - Display of data (User Data Model)

Integration into the NEEShub

- ▶ NEEShub provides tools as SaaS
- ▶ Archivematica could be added as a tool within the NEEShub
- ▶ SIP linked to ingestion directories for new data
 - NEES curator would be notified to process new data once it is available
- ▶ Automated processing
 - Check and learn formats
 - Create & validate checksums
 - Virus check
 - Create curation version of data
 - Create Dublin Core XML metadata
 - Create any additional metadata and XML data representations
- ▶ Flexible workflow
 - NEES can modify and fit archivematica microservices workflow to closely fit NEES requirements

Integration into the NEEShub

- ▶ Direct file access to NEES data file system, NEES tools, NEEShub web interface
- ▶ Authentication
 - Can use existing NEEShub authentication scheme
- ▶ Archiving tool
 - We can actually create and archive older versions of tools as VMs using OpenVZ that can be archived for long-term use later
 - We are already doing this using VMware for the older versions of NEESit infrastructure we inherited from SDSC
- ▶ These VM-based active and archived tools and data sources can be integrated into a microservices curation workflow

NEES Data Curation

- ▶ NEES Data Curator can use the system as part of a curation workflow
- ▶ Users upload project data using PEN/SynchroNEES
 - Curator sees new data in SIP inbox
 - Curator processes data using archivematica adopted to NEES project
- ▶ A copy of the data is transformed into a version suitable for long-term archiving
- ▶ DIP version of the data is produced and made available in the NEEShub

Expected Outcomes

- ▶ Better integrated curation workflow that can support the data life cycle
 - ▶ Extensible and flexible data framework that can be incrementally improved and evolved over time without drastic changes to the system
 - ▶ Ability to make software tools an archivable and usable object for future use
 - ▶ Movement towards compliance with ISO-OAIS standards and long-lasting data representation (i.e. XML) schemas and standards
- 

Conclusions

- ▶ Working to provide innovative cyberinfrastructure for the network
 - Goal of integrating the geographically distributed NEES resources and the user community.
 - Focusing on the management of research data
- ▶ NEEShub
 - Provides easy access to data
 - Combines analysis and simulation tools with data
 - Provides advanced functionality through a web portal
- ▶ Framework for community driven cyberinfrastructure