

George E. Brown, Jr.

Network for Earthquake Engineering Simulation

Data Curation and Quality Assurance at NEES

Stanislav Pejša

HUBbub 2012
Indianapolis, IN
2012-09-24



NEES

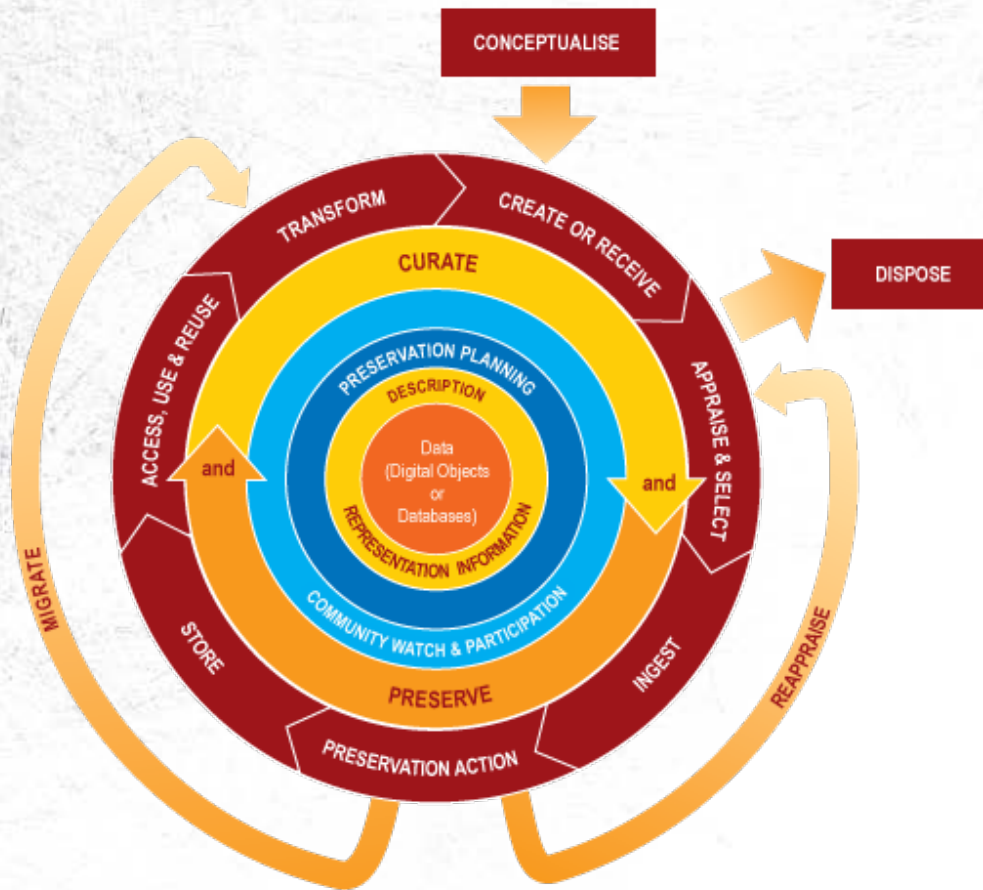


PURDUE UNIVERSITY
Discovery Park

This work is licensed under a [Creative Commons Attribution 3.0 Unported License](https://creativecommons.org/licenses/by/3.0/).



DCC Curation Life-Cycle Model



Data

- Digital objects

Full Lifecycle Actions

- Description and Representation information
- Preservation planning
- Community watch and participation
- Curate and Preserve

Sequential actions

- Conceptualise
- Create or receive
- Appraise and select
- Ingest
- Preservation action
- Store
- Access, use, re-use
- Transform

Occasional actions

- Dispose
- Reappraise
- Migrate

<http://www.dcc.ac.uk/resources/curation-lifecycle-model>

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

Access and re-use

NEES Project Warehouse

Mitigation of Collapse Risk in Vulnerable Concrete Buildings

Project Experiments Team Members File Browser

Experiment-5: Test Series B - Specimen 4

Experiment Type: Field

Description: RC Column tested under cyclic lateral displacement reversals
Uniaxial displacement protocol, 3 cycles per drift level
1.8 aspect ratio
2.0% longitudinal reinforcement ratio
0.07% transverse reinforcement ratio
0.43 axial load ratio
80 ksi longitudinal steel
60 ksi transverse steel
3600 psi concrete

Dates: February 17, 2009 to Present

Facility: University of Minnesota-Twin Cities, MN, United States

Specimen Type: Full-Scale Reinforced Concrete Columns

Material: Component 1 (view)

Sensors:

Drawings: Elevation of specimens 3 & 4 showing locations of horizontal LVDTs
Specimens 1-5 strain gage elevations
Elevation of transverse reinforcement layout for specimens 1 & 3 & 4
more...

Data: Specimen 4 Data file (inDEED)
File

Location: /NEES-2008-0637/Experiment-5/Trial-1/Rep-1

Name	Size	Time Stamp	Application
Converted_Data	12 MB	2009-02-23 12:27:52	
Connected_Data (0 files)		2009-02-23 12:27:52	
Derived_Data (0 files)		2009-02-23 12:27:52	
Unprocessed_Data	18 GB	2009-02-23 12:27:52	

Curation progress

- Facility: ✔
- Equipment: ✔
- Materials: ✔
- Sensors: ✘
- Drawings: ⚠
- Unprocessed Data: ✔
- Converted Data: ✔
- Connected Data: ✘
- Derived Data: ✘
- Tags: ✘
- Report: ✘

The experiment contains sensor measurements, but some metadata and documentation are missing. The experiment is missing experimental set-up report.

Incomplete - 03/12/2012
Last curation request: 03/09/2012

What's this?
Once the curator starts working with your submission, monitor the

Testing

project 637

Collected sensor measurements

Visualisation

West (-Y) East (+Y)

HW_CHANs:	SC1Mod1/ai1	SC1Mod1/ai1	SC1Mod1/ai2
CHAN_OFFSETS:	2.4990000000000000e-001	-6.0590000000000000e-001	7.6250000000000000e-001
CHAN_UNITS:	in	kips	in
CHAN_NAMES:	time	X_Displ	X_Force
			Y_Displ
2009-03-17	12:13:14.056	1.366342103823637e-003	-2.660552609949137e+00
2009-03-17	12:13:15.056	-5.872865660678239e-004	-2.794839395738763e+00
2009-03-17	12:13:16.056	8.779349363507438e-004	-2.714267324264988e+00
2009-03-17	12:13:17.056	-5.872865660678239e-004	-2.848554110054613e+00
2009-03-17	12:13:18.056	3.895277688779064e-004	-2.767982038580838e+00
2009-03-17	12:13:19.056	-9.887939859493100e-005	-2.848554110054613e+00
2009-03-17	12:13:20.056	-1.075693733540661e-003	-2.767982038580838e+00
2009-03-17	12:13:21.056	1.854749271296530e-003	-2.714267324264988e+00
2009-03-17	12:13:24.056	3.895277688779064e-004	-2.741124681422313e+00
2009-03-17	12:13:25.056	-5.872865660678239e-004	-2.821696752896686e+00

NEEScomm Data Goals

Aligned with NSF Data Management Plan (DMP) requirements*

- **All research data and documentation will be archived**
- **Archived data will be of high quality**
- **Archived data will be accessible and shareable**
- **Archived data will be re-usable**
- **Archived data will be preserved**

NEEScomm Data Sharing and Archiving Policies, <https://nees.org/resources/2811>

* http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/gpg_2.jsp#dmp



Data Archiving at NEES

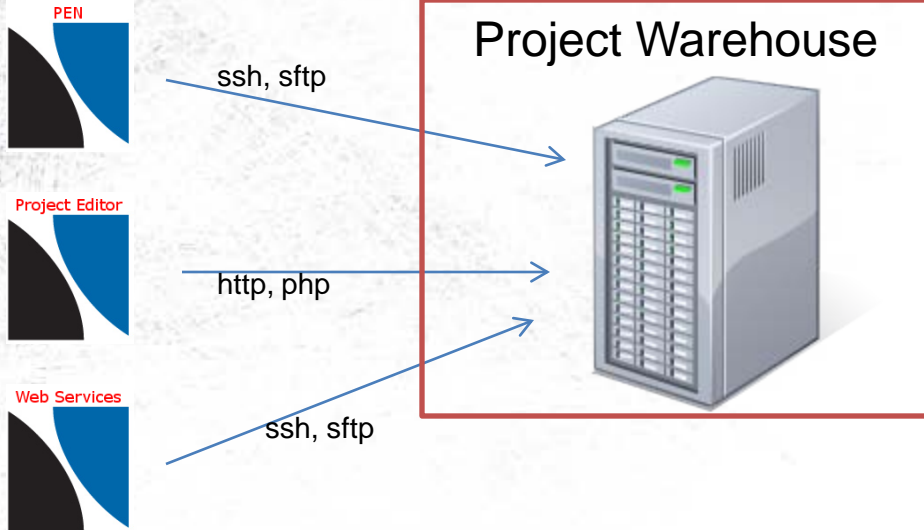
- **Who**
 - research team, site personnel, curator, NEEScomm
- **What**
 - sensor measurements
 - sensor calibrations
 - observations
 - analyses
 - numerical simulations
 - reports (including publications and presentations)
- **When**
 - Dates are stated in the Data Sharing and Archiving Policies (1 month, 6 months, 12 months)
 - For as long as the data are useful ~ indefinitely ~ for 20 years
- **Where**
 - Project Warehouse <http://nees.org/warehouse/welcome>
- **Why**
 - increases researcher's impact
 - saves work, time, money
 - facilitates knowledge transfer
 - maintained authenticity and integrity of data
 - good practice
 - advances research

What kind of data?

- diverse
 - shared facilities, not always practices
 - research domain
 - structural engineering
 - geotechnical engineering
 - geophysical research
 - material engineering
 - tsunami research
 - type of data
 - experimental
 - observational
 - computational
- increasingly complex
 - number of sensors
 - interdisciplinarity
 - experimenting with computational modeling

Where are the data coming from?

- DAQ systems on the site
- DAQ systems of research teams
- Researcher's computers
- Thumb-drives

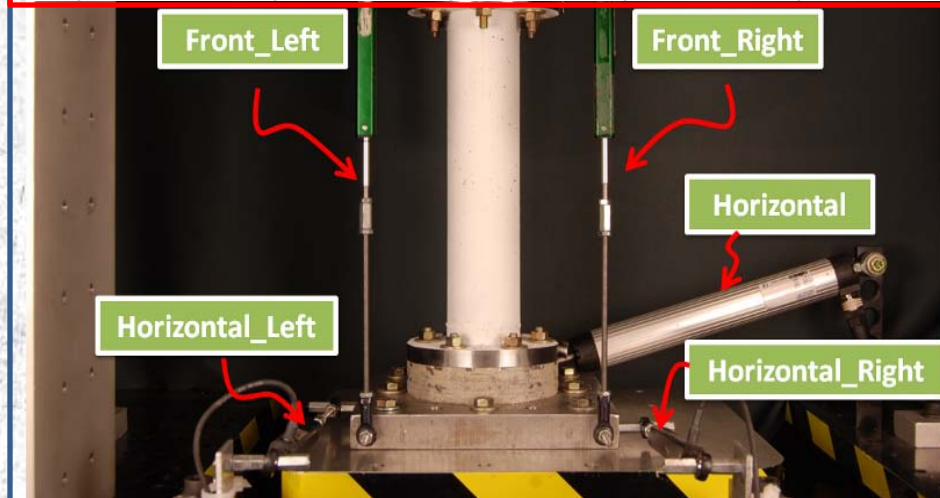


Quality assurance

- Content
 - dependent on human input and monitoring
 - Metadata
 - Completeness
 - Based on standards and requirements
 - » On the level of research hierarchy
 - » Timeline
- Technical
 - machine-actionable
 - Formats
 - Interoperability
 - Accessibility
 - Preservability
 - File integrity

Understandable data

Step	Horizontal Right inch	Horizontal Left inch	Back_Vertical inch	Horizontal inch	Front_Left inch	Front_Right inch
0	+1.2463610e+000	+1.1243380e+000	+7.3573610e-001	+2.4332460e+000	+1.0029440e+000	+8.6135440e-001
1	+1.2462570e+000	+1.1242340e+000	+7.3583960e-001	+2.4332460e+000	+1.0029100e+000	+8.6128560e-001
2	+1.2461880e+000	+1.1242000e+000	+7.3546010e-001	+2.4333470e+000	+1.0029790e+000	+8.6132000e-001
3	+1.2461880e+000	+1.1243030e+000	+7.3528760e-001	+2.4333470e+000	+1.0030130e+000	+8.6121670e-001
4	+1.2461180e+000	+1.1242340e+000	+7.3521860e-001	+2.4333470e+000	+1.0031160e+000	+8.6128560e-001



Hybrid Simulation LVDT Set-up

Channel Name	Label	Sensor Type	Comment	Orientation	XYZ Coordinates
Horizontal_Right	Horizontal_Right	LVDT		0, 1, 0	-7.21 in, 12.112 in, 1.65 in
Horizontal_Left	Horizontal_Left	LVDT		0, 1, 0	6.709 in, 12.014 in, 1.717 in
Back_Vertical	Back_Vertical	LVDT		0, 0, 1	1 in, -4.65 in, 19.875 in
Horizontal	Horizontal	LVDT		1, 0, 0	-14.913 in, -8.579 in, 5.084 in
Front_Left	Front_Left	LVDT		0, 0, 1	4 in, 4.15 in, 19.875 in
Front_Right	Front_Right	LVDT		0, 0, 1	-2 in, 4.15 in, 19.875 in

Metadata need to be:

- meaningful
- purposeful
- consistent
- accurate
- predictable
- "standardized"

Relationship among:

- Instrumentation plan
- Sensor metadata
- Data

CONTENT - Metadata

- Auto complete
 - Names
 - Organization
 - Facility
- Check boxes
 - Equipment
- System-generated folders/Tabs
 - Genre/Format
 - Proxy for metadata
 - Location on file system
 - File format constrains
- Templates
 - Sensors
- Forms
 - Material properties

test 3

About
Specimen
Sensors
Drawings
Data
Videos
Photos
Documentation

Title:* ?

Type of Test:* ?

Description: ?

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Phasellus ornare mauris ut dui egestas sed tristique ante adipiscing. Donec sit amet lorem elit. Aenean vitae turpis in nunc dignissim tristique et ac augue. Morbi ultricies bibendum augue, at laoreet nibh sagittis id. Morbi massa elit, laoreet nec congue a, lobortis ut justo. Integer sit amet nisi nec elit cursus consectetur. Fusce accumsan ipsum eu elit facilisis congue. Pellentesque imperdiet turpis at tellus vestibulum ut porta turpis adipiscing. Vestibulum id adipiscing est.

Start Date:* ?

End Date: ?

Facility: ?

Equipment: ?

UMN - 6 DOF Test System :: North Strongwall, 7' x 35' x 35'
 UMN - 6 DOF Test System :: West Strong Wall, 7' x 35' x 35'
 UMN - 6 DOF Test System :: Strong Floor, 5.5
 UMN - 6 DOF Test System :: 47 ton Steel Cruciform Crosshead
 UMN - 6 DOF Test System :: Hydraulic Pump, 190,000



Metadata - Autocomplete

- Personal Names

Tags (keywords): [?](#)

Document Type:

Contribute a NEEShub Resource? No Yes
 If yes, this file will become **PUBLIC**.

Authors: [?](#) +

Collaborators: [?](#)

Language:

Publisher: [?](#)

Published Year:

Created Year:

Coverage: [?](#)

Source: [?](#)

Mitsuyuki, Steven (smitsuyu) -

Smit, Peter (peter) +

Smith, Bruce (brucesmith7998) +

Smith, Chris (csmith)

Smith, Christopher (smithcm)

Smith, Curtis (curfissmith)

Smith, Jhon (jpsmith)

Smith, Joseph (jsmith)

Smith, Joseph (jdsmith57)

Smith, Katherine (ksmith)


Smith, Matthew (mds002)


Smith, Pat (study01)


Smith, Pat (study02)


Metadata - Autocomplete



- Personal Names
- Organization / Facility

Organization: 

Description: 

Sponsor: 

Website(s): 

Universidad Tecnica Particular de Loja  

Universidad de Costa Rica

Universidad del Valle, Cali, Colombia

Universiti Sains Malaysia

University of Akron, OH, United States


University of Alabama, AL, United States


University of Alabama at Birmingham, AL, United States

University of Alabama in Huntsville, AL, United States


University of Alaska Anchorage, AK, United States

University of Alaska Fairbanks, AK, United States


University of Arizona, AZ, United States 

University of Auckland, New Zealand 

University of Bologna, Italy

University of British Columbia, Canada 

University of Calgary, Canada







Metadata - Checkboxes

- Personal Names
- Organization/Facility
- Facility/Equipment

Facility: ?

Equipment: ?

- UMN - DAQ :: Sensor Connect Boxes
- UMN - DAQ :: NI Data Acquisition Board
- UMN - DAQ :: Data Acquisition Computer
- UMN - Krypton Measurement System :: Space Probe







System-generated folders/Tabs

- control of proper position
- provides metadata
- file format constrains

[About](#)
[Specimen](#)
[Sensors](#)
[Drawings](#)
[Data](#)
[Videos](#)
[Photos](#)
[Documentation](#)
[Analysis](#)
[Security](#)

Drawings: [?](#)

[Upload Drawing](#)

<input type="checkbox"/>		Title	Description	Type	Manage
<input type="checkbox"/>		test file		Sensor Layout	[Edit] [Delete]
<input type="checkbox"/>		test02 file		Drawing	[Edit] [Delete]
<input type="checkbox"/>		test03		Specimen	[Edit] [Delete]
<input type="checkbox"/>		replace test 04		Drawing	[Edit] [Delete]
<input type="checkbox"/>		test-of-teplace 02		Sensor Layout	[Edit] [Delete]



Templates

- sensor metadata difficult to collect
- variety of measuring instruments
- different research needs

Sensor Label	Sensor Type	Comment	Position			Position Units	Orientation		
			X	Y	Z		I	J	K
A0F2NX	Accelerometer		144	131	0	in	1	0	0
A0F2NY	Accelerometer		144	131	0	in	0	1	0
A0F2NZ	Accelerometer		144	131	0	in	0	0	1
A0F3WX	Accelerometer		277	131	0	in	1	0	0
A0F3WY	Accelerometer		277	131	0	in	0	1	0
A0F3WZ	Accelerometer		277	131	0	in	0	0	1
A0F4NX	Accelerometer		10	245	0	in	1	0	0
A0F4NY	Accelerometer		10	245	0	in	0	1	0
A0F4NZ	Accelerometer		10	245	0	in	0	0	1
A0F7NX	Accelerometer		277	245	0	in	1	0	0
A0F7NY	Accelerometer		277	245	0	in	0	1	0
A0F7NZ	Accelerometer		277	245	0	in	0	0	1
A1S1EX	Accelerometer		10	131	105	in	1	0	0

Completeness

- **Research teams**
 - Professional standards
 - Local Site guidelines
 - Team guidelines for data management
 - NEEScomm documentation and metadata requirements
 - NEEScomm data archiving milestones
 - T0 - completion of an experiment
 - T + 1month: Unprocessed data uploaded
 - T + 6 months: Metadata and Documentation are uploaded
 - T + 12 months: Rest of the data and experimental report are uploaded [OPEN DATA](#)

- **NEES Sites**
 - Certifications
 - Professional standards
 - Local guidelines (naming conventions, etc.)

- **NEES repository**
 - OAIS
 - PREMIS
 - Dublin Core
 - NEEScomm documentation and archiving policies

Ready for Curation?		Curation progress:	
Facility	✓	Facility	✓
Equipment	✓	Equipment	✓
Materials	✓	Materials	✓
Sensors	✓	Sensors	✓
Drawings	✓	Drawings	✓
Unprocessed_Data	✓	Unprocessed_Data	✓
Converted_Data	✓	Converted_Data	✓
Corrected_Data	✓	Corrected_Data	✓
Derived_Data	✓	Derived_Data	✓
Tags	✓	Tags	✓
Report	✓	Report	✓

Test Curation

Experiment 2 contains unprocessed sensor measurements in Volts. Some issues are still pending.

Incomplete - 02/10/2012

Last curation request: 01/20/2012

Technical quality

- format identification
 - FITS tool
 - JHOVE
 - NLNZ Metadata Extractor
 - Exiftool
 - DROID
 - FFident
 - File utility (windows)

Objects with file extension '*.jpg' in PW

FORMAT_NAME	COUNT(*)
Exchangeable Image File Format	344041
JPEG File Interchange Format	93678
Conflict	68304
empty	574
(null)	141
Extensible Markup Language	37
Portable Network Graphics	31
JTIP (JPEG Tiled Image Pyramid)	5
JPEG	1
Plain text	1

PREMIS_ID	DF_ID	CREATED_DATE	MODIFIED_DATE	LOCATION	FORMAT_NAME	FORMAT_NAME_MIMETYPE
3528106	3528106	23-MAY-12	23-MAY-12	Experiment Setup Report_Int.docx	OpenDocument Text	application/vnd.oasis.opendocument.text
3523121	3523121	22-MAY-12	22-MAY-12	Experiment Setup Report_Int.docx	OpenDocument Text	application/vnd.oasis.opendocument.text
3523120	3523120	22-MAY-12	22-MAY-12	Experiment Setup Report_Int.docx	OpenDocument Text	application/vnd.oasis.opendocument.text
3523119	3523119	22-MAY-12	22-MAY-12	Experiment Setup Report_Cir	OpenDocument Text	application/vnd.oasis.opendocument.text
3523118	3523118	22-MAY-12	22-MAY-12	Experiment Setup Report_Cir	OpenDocument Text	application/vnd.oasis.opendocument.text
3523043	3523043	21-MAY-12	21-MAY-12	C1_P Experiment Setup Report	OpenDocument Text	application/vnd.oasis.opendocument.text



Technical quality

- format identification
 - FITS tool
 - JHOVE
 - NLNZ Metadata Extractor
 - Exiftool
 - DROID
 - FFident
 - File utility (windows)

- format validation
 - JHOVE

FORMAT_NAME	COUNT(*)
Exchangeable Image File Format	344041
JPEG File Interchange Format	93678
Conflict	68304
empty	574
(null)	141
Extensible Markup Language	37
Portable Network Graphics	31
JTIP (JPEG Tiled Image Pyramid)	5
JPEG	1
Plain text	1

```

<identity format="JPEG" mimetype="image/jpeg" toolname="FITS" toolversion="0.6.1">
<filestatus>
  <well-formed toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">true</well-formed>
  <valid toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">false</valid>
  <message toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">File does not begin with
SPIFF, Exif or JFIF segment offset=115</message>
</filestatus>
  
```



Technical quality

- format identification
 - FITS tool
 - JHOVE
 - NLNZ Metadata Extractor
 - Exiftool
 - DROID
 - FFident
 - File utility (windows)

- format validation
 - JHOVE

FORMAT_NAME	COUNT(*)
Exchangeable Image File Format	344041
JPEG File Interchange Format	93678
Conflict	68304
empty	574
(null)	141
Extensible Markup Language	37
Portable Network Graphics	31
JTIP (JPEG Tiled Image Pyramid)	5
JPEG	1
Plain text	1

```

<identity format="JPEG File Interchange Format" mimetype="image/jpeg" toolname="FITS" toolversion="0.6.1">
<filestatus>
  <well-formed toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">true</well-formed>
  <valid toolname="Jhove" toolversion="1.5" status="SINGLE_RESULT">true</valid>
</filestatus>

```



Technical quality

- format identification
 - FITS tool
 - JHOVE
 - NLNZ Metadata Extractor
 - Exiftool
 - DROID
 - FFident
 - File utility (windows)

- format validation
 - JHOVE

- file integrity
 - MD5

FORMAT_NAME	COUNT(*)
Exchangeable Image File Format	344041
JPEG File Interchange Format	93678
Conflict	68304
empty	574
(null)	141
Extensible Markup Language	37
Portable Network Graphics	31
JTIP (JPEG Tiled Image Pyramid)	5
JPEG	1
Plain text	1

PREMIS_ID	DF_ID	CREATED_DATE	MODIFIED_DATE	LOCATION	CHECKSUM	PREMIS_CHECKSUM
3528106	3528106	23-MAY-12	23-MAY-12	Experiment Setup Report_Int.docx	775c26a1f97cecd8f6881f27ca722846	775c26a1f97cecd8f6881f27ca722846
3523121	3523121	22-MAY-12	22-MAY-12	Experiment Setup Report_Int.docx	775c26a1f97cecd8f6881f27ca722846	775c26a1f97cecd8f6881f27ca722846
3523120	3523120	22-MAY-12	22-MAY-12	Experiment Setup Report_Int.docx	775c26a1f97cecd8f6881f27ca722846	775c26a1f97cecd8f6881f27ca722846
3523119	3523119	22-MAY-12	22-MAY-12	Experiment Setup Report_Cir	791ea95f4df85c6fab115c1bba2cf270	791ea95f4df85c6fab115c1bba2cf270
3523118	3523118	22-MAY-12	22-MAY-12	Experiment Setup Report_Cir	791ea95f4df85c6fab115c1bba2cf270	791ea95f4df85c6fab115c1bba2cf270
3523043	3523043	21-MAY-12	21-MAY-12	C1_P Experiment Setup Report	791ea95f4df85c6fab115c1bba2cf270	791ea95f4df85c6fab115c1bba2cf270

Curation - Path through SWAMP

- The harder you try to get out, the deeper you sink



Image credit: <http://www.machinery.uk.com/interestingpics/fendt1.jpg>

Curation - Path to SWAMP

- **Straightforward** (relatively)
- **Way to**
- **Authorship**
- **Merit and**
- **Publication**

Curation progress:

Experiments contain data and required metadata and documentation. The experimental setup reports and the final report were uploaded, as well.

Complete - 08/06/2012

Curation progress:

Facility	✓
Equipment	✓
Materials	✓
Sensors	✓
Drawings	✓
Unprocessed_Data	✓
Converted_Data	✗
Corrected_Data	✓
Derived_Data	✗
Tags	✓
Report	✓

Cite this work:

Matthew Smith; Chia-Ming Uang (2012), "Full-Scale Metal Building Frame with Heavy Wall Attachment and Mezzanine", Network for Earthquake Engineering Simulation (database), Dataset, DOI:10.4231/D3N00ZS79

Experiment-4:

[Full-Scale Metal Building Frame with Heavy Wall Attachment and Mezzanine](#)

OPEN DATA

Dates:

June 17, 2010 - June 18, 2010

Description:

This full-scale metal building specimen is the third and last in a series of specimens for shake-table testing. This specimen has a 20' eave ...



Photo credits

Slide 3:

- Specimen 5 after axial failure at 1.0% drift in Y direction.
<http://nees.org/warehouse/experiment/1622/project/637>

Slide 7:

- NEES@University of Buffalo. Stanislav Pejša (2011)
- DAQ setup. Ian Prowell (2009)
https://nees.org/data/download/NEES-2008-0661/Experiment-6/Documentation/IMG_2614.JPG
- Vibpilot.
<http://www.mpihome.com/images/vibpilot2.jpg>
- Thumbdrive.
<http://openclipart.org/detail/167170/usb-thumb-drive-3-by-rygle>

Slide 9:

- LVDT layout 1
http://nees.org/data/download/NEES-2009-0685/Experiment-3/Documentation/Drawings/Sensor%20Layout/SensorLayout_FrontView.png

Slide 21:

- The Fendt. Stuck on peat marshes at Hardley in Norfolk
<http://www.machinery.uk.com/interestingpics/fendt1.jpg>



Thank you.

Questions?

Comments?

Standa Pejša - spejsa@purdue.edu

NEEScomm Data Curator

Network for Earthquake Engineering Simulation

