Building Infrastructure for Preservation and Publication of Earthquake Engineering Research Data

Stanislav Pejša
Network for Earthquake Engineering
Simulation

Shirley J. Dyke
Purdue University, Network for
Earthquake Engineering Simulation

Thomas J. Hacker
Purdue University, Network for
Earthquake Engineering Simulation

Abstract

The objective of this paper is to showcase the progress of the earthquake engineering community during a decade-long effort supported by the National Science Foundation in the George E. Brown Jr., Network for Earthquake Engineering Simulation (NEES). During the four years that NEES network operations have been headquartered at Purdue University, the NEEScomm management team has facilitated an unprecedented cultural change in the ways research is performed in earthquake engineering. NEES has not only played a major role in advancing the cyberinfrastructure required for transformative engineering research, but NEES research outcomes are making an impact by contributing to safer structures throughout the USA and abroad. This paper reflects on some of the developments and initiatives that helped instil change in the ways that the earthquake engineering and tsunami community share and reuse data and collaborate in general.

The NEEShub Platform for Collaboration

With the inherent uncertainty surrounding earthquake hazards, and the rate of population growth in urban areas, a critical global challenge is to achieve a level of seismic resilience needed to ensure that communities are safe, sustainable, secure, and economically strong. Since 2004, research conducted at 14 state-of-the-art laboratories distributed throughout the USA has generated a wealth of valuable experimental data resulting in new design techniques, improved construction methods, strategies to improve the resilience of existing infrastructure against earthquakes and tsunamis, and a new paradigm for research collaboration in earthquake engineering (NEES Consortium, 2007; Ramirez, 2012). Among the various stakeholders in NEES, there are four related to NEES Operations: (i) the sponsor, National Science Foundation (NSF); (ii) the NEEScomm Center as the administrative headquarters of the network; (iii) the individual research laboratories and their staff, whose expertise and skills are essential for the successful execution of many innovative and interdisciplinary experimental methods; and (iv) the research community.

The venue where these four stakeholders converge is the NEEShub[1] (Hacker et al., 2011) – a collaborative platform based on HUBzero[2] technology (McLennan and Kennell, 2010). A long-standing core goal of the entire NEES effort is that the community of earthquake engineers and practitioners will coalesce around the NEES data repository, the virtual collaborative research environment, state-of-the-art testing capabilities at the laboratories, and the tools to enable visualization, advanced analysis and access to high-performance computing infrastructure. While the NEEShub primarily serves researchers whose projects are funded through NSF's NEESR program, the repository is open to other relevant engineering research projects that require data management capabilities, whether they are funded through federal or private funding agencies. Currently the strength of the repository is in earthquake and tsunami projects, but data from related engineering fields, such as wind, blast, hurricane, tornado, etc can be accommodated.[3]

In just a few years, the NEEShub has established itself as a virtual research environment (Voss and Procter, 2009) that fosters collaboration and provides access to a variety of resources that are of interest to the earthquake engineering community at large. These include tools, theses, articles, educational materials and simulation models. The NEEShub serves as the central US portal for earthquake engineering research data, at this time consisting mainly of the data stored in the NEES data repository, dubbed the Project Warehouse.[4]

The NEES Project Warehouse

The data in the NEES repository are expected to be of high quality, preservable, and accessible for the long term. To meet these expectations, NEEScomm developed a set of data curation guidelines, workflows and services that facilitate controlled and fast upload of data together with the metadata and documentation necessary for their correct

---

1 NEEShub: http://www.nees.org/
2 What is HUBzero? http://hubzero.org/about
3 The US National Institute of Standards and Technology (NIST) was experimenting with HUBzero platform to house data from the 2010 Chile earthquakes and other disasters. (Litvin and Pujol, 2013)
4 Project Warehouse: http://nees.org/warehouse/welcome

interpretation, reuse and preservation. The research data in the NEES data repository must conform to best practices for conducting earthquake engineering research, as described in the *NEEScomm Guidelines for Data Upload* (2012a), and the *NEEScomm Requirements for Curation and Archiving of Research Data* (2012b).

Three types of data are being collected:

- Sensor measurements collected from the instrumentation and data acquisition system (DAQ);

- Data captured as still images or moving images by installed camera systems;

- Required documentation, such as sensor metadata, technical drawings and reports.

The NEES data model and metadata schema were developed in the early years of NEES with input from the earthquake engineering research community in years 2004-2007 (Peng and Law, 2004; Van Den Einde et al., 2008), but they are constantly updated and expanded based on the requirements from the research community. Compliance with the hierarchy of the earthquake engineering research workflow is key for a correct understanding of the research, but it also helps to navigate to the correct location of the data or documentation. The hierarchy (Figure 1) consists of the collection container *Project* that corresponds to a NSF award; the level of *Experiment* contains the essential metadata and documentation for a given test; *Trial* allows for differentiations of different loads or research approaches while working on the same specimen; and finally the data from instruments are stored on the *Repetition* level. On the Repetition level the data can be further subdivided depending on the type of data processing, allowing for possible verification and more granular reproducibility of the research data.
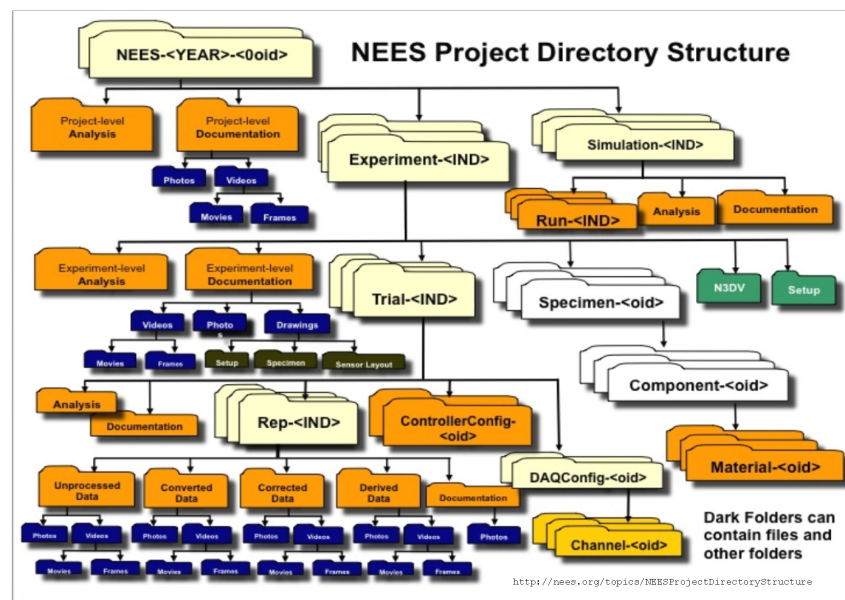

Figure 1. The hierarchy of the research work in the NEES data repository.[5]

Functional data management infrastructure, and transparent and predictable workflows are essential for achieving the NEES data goals (NEES, 2011). For efficient data management, the curation requirements must be communicated to the researchers

---

5   This diagram is available from https://nees.org/topics/NEESProjectDirectoryStructure

early on. While research teams are expected to be familiar with the data archiving requirements, the cyberinfrastructure must facilitate quick, reliable and effective transfer of knowledge. An intuitive interface of the Project Warehouse (Figure 2) assists researchers in the identification of key evidentiary components of the necessary documentation, but also helps to ensure that the pieces of documentation are stored in the proper location. The location also serves as a proxy for metadata of individual pieces of documentation that a researcher would otherwise be required to provide.

Figure 2. The tab-driven interface of the web-based editor of the Project Warehouse.

Curation Services

The Curation Dashboard (Figure 3) is another important feature of the Project Warehouse that helps to communicate the curation requirements and provides researchers with feedback regarding completeness of their research documentation. It dynamically updates researchers on their conformance with the NEES curation requirements.

After researchers complete archiving their data, they submit each dataset for curation review. During that review, the curation team assesses whether the individual pieces of evidence meet the stated requirements. The curation team then provides the research teams with feedback by noting whether a requirement is completed, indicated by a change in the color of the icon on the dashboard. If task is completed the icon remains green; if additional changes are required the icon is changed to yellow; the red icon indicates a missing piece of documentation or metadata. More detailed comments are provided as needed through email. The main goal of curation is to ensure that data are archived with all necessary documentation and metadata, so that an experiment or simulation can be correctly understood and interpreted (Vermaaten, Lavoie and Caplan, 2012).

Given the importance of data collected in the NEES data repository, which contains data that will impact future design codes and serve as a basis for decisions regarding mitigation of natural hazards, curation services are at the centre of the NEES infrastructure (Marchionini, Lee, Bowden and Lesk, 2012). The NEES curation team works to establish communication with the research teams in the early phase of their research, so that the transfer of knowledge and experience, as well as the contextual

information, is conveyed by the research teams directly to the repository. An early contact intervention with the research team also helps to address some of the data management issues (Carlson, Johnston, Westra and Nichols, 2013) that teams often face immediately after they complete their tests. Frequent contact between the curators and research teams is complimented by a preservation infrastructure based on the micro-services concept that automatically extracts some of the significant preservation metadata, as well as some key administrative and technical metadata.
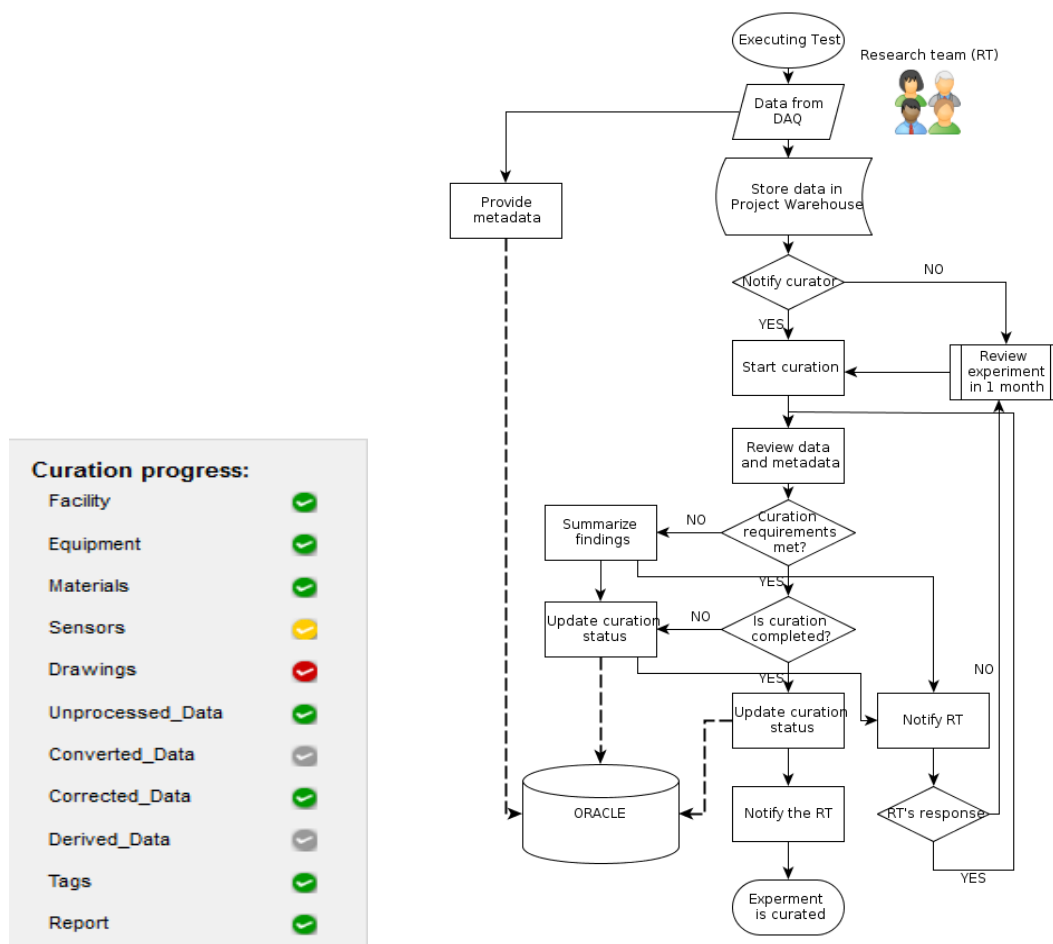
Figure 3. Curation Dashboard.

Figure 4. Curation workflow in the NEES data repository.

Curation in NEES is an interactive and iterative process (Figure 4) that not only assesses the fitness of given data for archiving in the NEES data repository, but also ensures that the data are archived and published on schedule. The data funded through NSF's NEESR program are expected to be published within 12 months after a test has been executed (NEES, 2011).

All research data archived in the NEES data repository are subject to a quality assurance review. Only after the data meet the minimum curation requirements are they accepted into the repository and the NEES formally takes physical ownership of those data files.

The process from planning and generating the data to publication requires meticulous documentation (Faniel and Jacobsen, 2010) over a period of months to years. For this reason, researchers are encouraged to upload documentation and

processed data gradually as they progress with their analysis. To help ensure that the research teams won't be overwhelmed with requirements upon the expected publication date of their datasets, additional deadlines for archiving of data and metadata were devised to more evenly spread the archiving over a 12-month period while the research team has an exclusive control over their data. The intermediate deadlines (Table 1) also provide an opportunity for the periodic review of the archiving progress of the dataset and for intervention and corrective action, if necessary. If a research team does not meet the data archiving schedule, their project is monitored monthly and the team is reminded of the deadlines until they comply with the curation requirements.

Table 1. Intermediate deadlines for delivering documentation required for curation and data archiving.

Required items for curation	Delivery
Unprocessed sensor data	1 month
Metadata and documentation	6 months
Experiment report and analysis	12 months

Besides verifying completeness of research documentation and timeliness of data archiving, ensuring that files are uploaded in interoperable formats and that the data will be accessible, understandable and reusable in the future is the third major concern for the curation team. These efforts are discussed next.

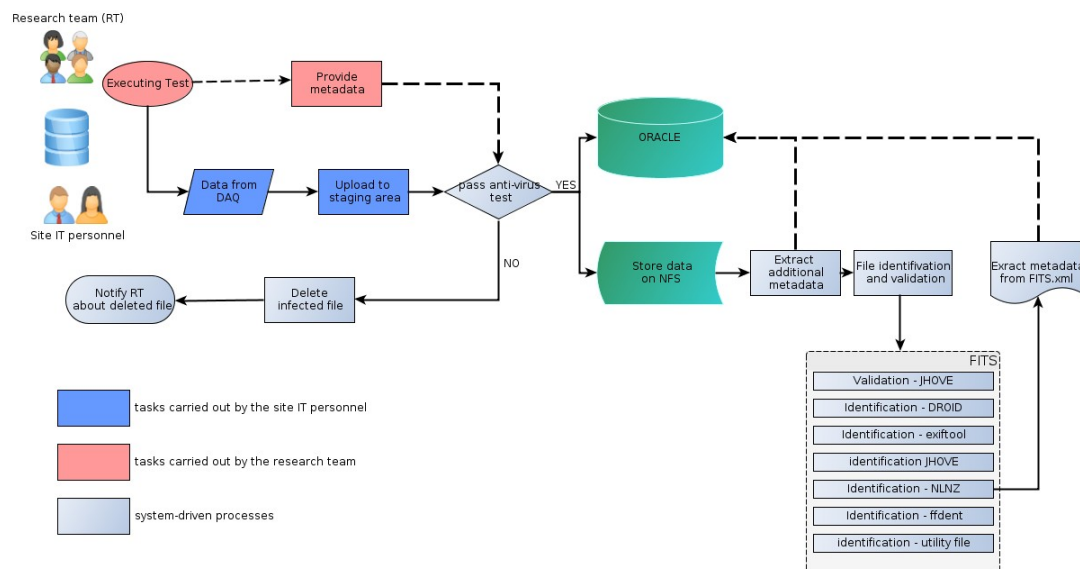**Preservation Efforts in the NEEShub**

As the NEES data repository has not yet reached its first decade (the first files were uploaded to the repository in 2006, although some of the tests date back to the 1970s), the danger of file format obsolescence is relatively small. However, the cutting-edge and interdisciplinary nature of earthquake engineering research requires new software packages, codecs, or new types of sensors and hardware. That often introduces new experimental formats or unknown file extensions. Reliable identification of file formats is therefore a high priority, and the preservation activities in the NEEShub at this point primarily focus on format identification and validation.

Identification and validation (micro-)services (Abrams, Kunze and Loy, 2010) are deployed as part of the NEES preservation pipeline (Figure 5). At the center of these micro-services is FITS[6] - a stack of format identification and validation applications that are complemented by human inspection. The identification services are run as nightly automated jobs within 24 hours after files are uploaded. The identification data are then examined during the curation review and compared with uploaded files on the file system. The extracted mime-types also have to be inspected because files with the same extensions often are identified incorrectly as different formats or conversely different formats have the same extensions. This often happens because the information in the file headers is not reliable or because acquisition systems and other software and hardware used during tests use the same extensions for their output; even if the files are not identical or even interoperable. It is often necessary to further analyze the datasets, taking into consideration which research team uploaded the data and which laboratory was involved. If a new format of data is identified then the curation team collects the

---

6 File Information Tool Set (FITS): https://code.google.com/p/fits/

necessary metadata about the format from the research team. The format-related metadata are based on the PRONOM vocabulary specifications (The National Archives, 2011). Requiring researchers to provide all unprocessed data in ASCII format, either as tab-delimited text files or comma separated values files, is a strategy that mitigates the adverse effect of the proprietary or novel formats. The ASCII format is required and those files are going to be fully preserved; the proprietary formats are optional and will be preserved only on a bit-level.

Figure 5. The preservation workflow in the NEEShub.

Collection of Metadata

The most accurate metadata are those collected from the research teams as they conduct the research. However, the curation team at NEES also recognizes that the research teams have relatively small incentive to spend prolonged periods of time on documenting metadata (Qin, Ball and Greenberg, 2012) that they may not consider useful. Therefore, it is important that researchers archive their data as they analyze them and publish their results. The amount of metadata researchers provide for most documentation files is typically limited to titles and brief descriptions.

A majority of the metadata for individual files is of a technical and administrative nature provided by the infrastructure (Figure 5). They play a key role in maintaining file integrity, in verifying authenticity of uploaded files, and in securing access to the files. The system logs when and who uploaded the files, and modified their metadata. The platform also verifies that only those researchers who are authorized to access a given experiment are allowed to do so. The security is set on the experiment level, so the principal investigators (PI) or another authorized member of the research team can manage access permissions of individual members of their research team on that level.

Each file is scanned for viruses with the ClamAV software. Only after verifying that a file does not contain any virus, it is stored in the repository and its metadata are inserted into the database. Nightly scripts run a series of micro-services that collect additional technical and preservation metadata, such as format identification and validation, mime-type, etc. The file is also check-summed at this time.

The experiment report is one of the final requirements for curation of an archived dataset. As these reports often take the form of theses or pre/post-prints, it makes sense that these reports are shared with the larger community as stand-alone resources, therefore additional metadata modeled on the Dublin Core are required.

Publication in NEEShub

The HUBzero framework allows for the publication of several different types of resources out of the box. The NEEShub leverages this publication mechanism through the use of several of these resources and further expands types of publishable materials. There are two types of publications in NEEShub: (i) resources that are voluntarily contributed by the members of the earthquake engineering community which are released to the public via HUBzero channels; and (ii) products that are considered "published" by the NEES and are assigned a DOI. This paper focuses on the latter group. While all resources are monitored before being released to the public, mainly for completeness and intellectual rights, resources that are assigned a DOI receive closer scrutiny. At this time, three types of resources are published with DOI in the NEEShub:

- Curated and public experimental data,

- Models for computational simulation,

- Compiled databases.

Data Publication

From its inception, the NEES data repository was intended to contain shareable and publicly accessible research output. However, the real value of archiving the important experiments lies in the ability of other researchers to reuse these data and expand the impact of those experiments. Over the course of almost a decade of the existence of the NEES repository, it became obvious that it is not sufficient only to deposit data for public access. Users, not familiar with the existence, purpose, and structure of the NEEShub would have difficulty identifying the NEEShub as a source of research data. Thus, efforts have been made to further expose these data externally to the NEEShub.

In December 2012, NEES started to issue a DOI for each curated research dataset. This step, amounting to publication of the data in the mind of the researcher, motivates proper documentation as well as encourages reuse of and improves access to research data (Piwowar and Vision, 2013). The primary pre-requisite for a DOI is an upgrade of descriptive metadata, thus metadata on the experiment level need to be enhanced so that the metadata can support the expected activities of the users/researchers, primarily discoverability, identification and retrievability (Socha, 2013).

First and foremost, these improved metadata enable a more effective discovery of NEES data in third party services, such as Google Scholar and DataCite, to potentially increase impact of the data outside of the NEES data repository. NEEShub users also benefit from these adjustments, as they can more effectively search for data within the NEEShub. Apart from the requested improvements – including more informative and meaningful titles and descriptions of the datasets – researchers are also asked to provide the names of additional members of the research team that can be considered authors of the dataset, as it is recognized that the authorship of the dataset may differ from

authorship of related print publications. Often laboratory personnel contribute documentation, instrumentation, or improve testing methodology – all steps that contribute to successful testing – and recognition for that contribution should be reflected in the data citation. By December 2013, over 500 DOIs have been issued for individual datasets out of 1353 that are public and curated in the Project Warehouse.

The research data sets are made public after a one year embargo period under the Open Data license with attribution ODC-BY 1.0[7] that allows researchers to copy, distribute and use the dataset, to produce works from the data stored in the NEES data repository, and build upon the data stored there, as long as they attribute any public use of the repository and the authors of the data set. The 'Open Data' icon is visible on each publicly available dataset (Figure 6).

Figure 6. Display of a curated and public dataset in the Project Warehouse.

The recommended citation for the dataset is prominently displayed on the page of the experiment within the NEEShub to make identification of the data easier and to promote the data outside of the environment of the NEES data repository. A file with the text of the attribution formula is also appended to any data downloaded from the repository.

The tracking of usage statistics for data use is becoming an increasingly popular indicator of the usefulness and impact of a given dataset. Each curated experiment in the NEEShub also contains a page that lists key statistics about the number of views and downloads, as well as timeline for each of these categories. Another available metric lists formats present in the dataset. This should also enable data reusers to assess the tools they may need to view and analyse the data before they start digging deeper into the dataset.

A citable published dataset and the accompanying metrics are seen as instruments to incentivize researchers to archive their data in the NEES repository, and to get more actively involved in the NEEShub community.

---

7   ODC Attribution Summary: http://opendatacommons.org/licenses/by/summary/

Models[8]

Computational simulations are becoming more comprehensive, complex and integrated within earthquake engineering, a capability that is of considerable interest to the NSF and other funding agencies. The NEES repository contains a variety of computational models for specific structures and systems, but has just recently started to formalize the review process to provide the community with high quality models that will further advance research into computational simulation in the earthquake engineering field and contribute to other related fields.

In order to publish a computational model with a DOI in the NEEShub, it not only needs to work, but it also needs to be properly documented. Such a model requires:

- A stated purpose;

- A description of how the model is intended to be used;

- A list of required software, version and environment;

- Descriptions of individual files that constitute the model;

- Units (i.e. system of units used for force, displacement, Young's modulus);

- Descriptions of variables and parts of the file defining the model, if applicable;

- Descriptions of additional files (i.e. input files), if applicable;

- Descriptions of outputs and output files.

Database Compilations

Another new data product that was introduced in the NEEShub is an aggregation of testing results from a variety of experiments that focus on a particular problem (e.g. a type of component or class of test). These resources represent useful instruments for quickly transferring earthquake engineering research to engineering practitioners. The databases reuse previous data results and analysis, and can be further enhanced through adding a visualization component (Browning, Pujol, Eigenmann and Ramirez, 2013). To provide a generalized solution that enables researchers to quickly deploy these databases, NEES supported the development of DataStore[9], an application that quickly turns formatted spreadsheets into searchable databases and makes them quickly accessible to the community.

In the earthquake engineering context, the user contributed databases proved particularly suitable for cases such as display of survey data after natural catastrophes, compilations of data used by code committees that review changes for revisions of design codes, etc.

Visualization and Analytical Tools

The research data, models, and databases are complemented at NEEShub by tools that facilitate analysis of data, telepresence, simulation, and visualization of data. Among the most popular tools at NEEShub is OpenSEES Laboratory[10],which provides access to the

---

8 Computational models: https://nees.org/resources/computationalmodels
9 DataStore: https://nees.org/resources/databases
10 OpenSEES Laboratory: http://nees.org/resources/openseeslab

Grid infrastructure where researchers run computationally demanding operations. Another example is the educational version of the SAP2000[11], a tool that can be launched directly in the NEEShub. Unlike SAP2000, the majority of the tools are developed by the researchers themselves, e.g. SLAMMER[12] that enables sliding block displacement analyses for seismic slope stability or PocketStatics, a structural analysis program.

inDEED[13] is a visualization tool that plays a particularly significant role on the NEEShub. inDEED allows researchers to rapidly visualize large quantities of experimental data without the need to download them. Thus, the tool is integrated with the NEEShub, co-locating the tool with the data. Additionally, comparisons of experimental and numerical simulation data are possible, and some analysis capabilities are provided. 3DDV[14] is a visualization tool that has been recently integrated into the NEEShub. It is designed for visualization and comparison of 3D models and 2D plots. 3DDV was developed by the staff at two NEES sites – RPI and Oregon State University.

A New Paradigm for Research Collaboration

Over the last decade, the capabilities, policies and activities discussed in this article have fostered the development of a community of researchers working within a new paradigm for research collaboration. During the planning phase of NEES it was not clear that the community would embrace the idea of sharing of research data. Admittedly, the barriers were mostly cultural within the earthquake engineering community. However, through active promotion of the cyberinfrastructure and persistent updates and introduction of new capabilities, the earthquake engineering community has evolved and progressed toward openly sharing not only data, but tools and computational models as well. The NEES community is accelerating the pace of discovery by influencing and expanding the way in which the earthquake engineering community generates and consumes data.

NEEScomm promotes and incentivizes reuse of the research data, and these efforts have further encouraged researchers to publish their data. Metrics documenting the number of views and downloads of individual datasets are available for each dataset. Persistent identifiers, specifically DOIs, are being requested by research teams for individual published datasets so that they can be easily discovered, identified and cited. The NEEShub provides each dataset with recommended citation format. New knowledge is being generated from the data sets leading to safer and more resilient infrastructure systems.

In addition to the efforts of NEEScomm to actively promote data reuse, funding agencies, specifically the NSF, have funded projects that reuse data sets. Since 2011, NSF has been soliciting proposals for awards that require significant use of one or more of the NEES laboratories "...and/or require **significant reuse of data** that is curated and archived" in the NEES repository (stress added by authors) (National Science Foundation, 2011). Such a wording of solicitation indicates the significance the NSF attributes to data reuse. Through this program, several research teams are reusing

---

11 SAP2000 Educational Version: http://nees.org/resources/sap2000edu
12 Seismic LAndslide Movement Modelled using Earthquake Records (SLAMMER):
     http://nees.org/resources/slammer
13 Interactive Display for Earthquake Engineering Data (inDEED): http://nees.org/resources/indeed
14 3D Data Viewer: http://nees.org/resources/3DDV

multiple datasets from the NEES data repository[15] Input from the community is always welcome to simplify the use of the repository even further.

This new paradigm extends beyond the existing population of researchers. Graduate and undergraduate students who have worked within this new research paradigm have graduated and become practicing engineers or faculty members. Their active participation in NEES committees and memberships to NEEShub is evidence that this new generation of researchers and faculty are continuing in this mode of publishing and reusing published data sets. These observations are a clear sign that they see great value in the establishment of NEES data repository.

Over the past ten years there has been great progress toward the development of a community of researchers contributing open data and advancing discovery using past data. In the past few years we have seen this trend accelerate. NEEShub, now in its fifth year of existence, has grown considerably in all respects since its launch in July 2010 (Hacker, Eigenmann and Rathje, 2013). The number of registered users almost quadrupled. The NEES data repository had some 600 users in 2006. It expanded to 2,500 users by 2010, and at the end of the year 2013 it was over 7,900. NEEShub is used by 95,000 users annually, out of them over 80,000 downloaded at least one resource.[16] All other significant metrics measuring success of the network, such as the number of uploaded files (Figure 7), number of curated projects and experiments (Figure 8), have also significantly increased.
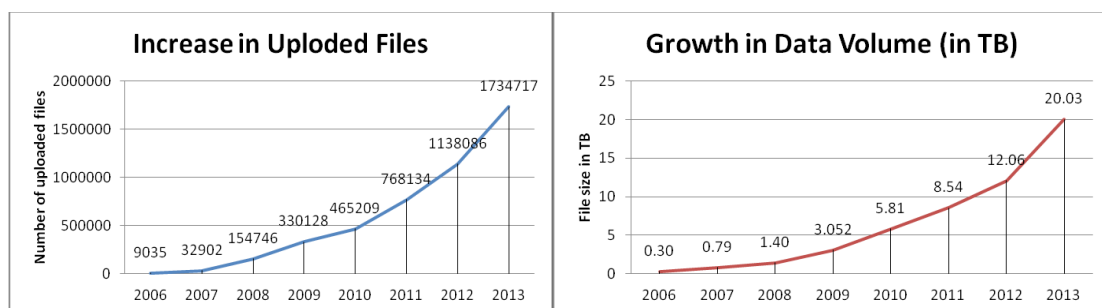
Figure 7. The growth of the NEES data repository.
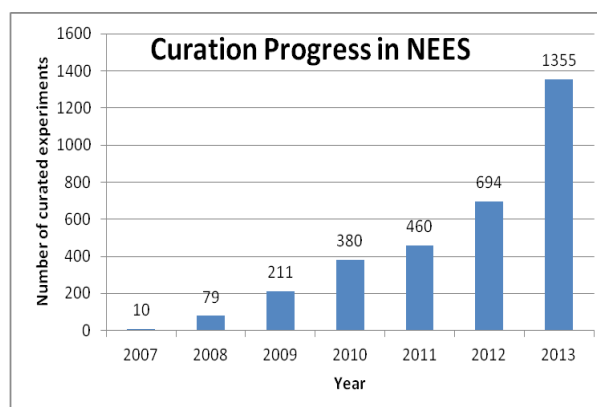
Figure 8. Growth of curated and publicly available datasets in the NEES data repository.

---

15 Several studies resulting from this effort to encourage reuse were submitted for publication, e.g. Balafas and Kiremidjian (in press), Cheung and Kiremidjian (in press), and Bernal (n.d.). Several papers on data reuse in NEEShub were also presented during the SERIES Concluding Workshop – Joint with US-NEES – "Earthquake Engineering Research Infrastructures". See http://www.series.upatras.gr/workshop_Ispra

16 NEEShub Usage Overview: http://nees.org/usage/overview

In the future we anticipate this cultural shift to expand. Recent earthquakes in Chile, Haiti, Italy, and Japan are reminders that there is a great deal of work yet to be performed to increase the resilience of our communities (National Research Council, 2011a, 2011b). NEES aspires to lead the earthquake engineering community toward embracing the potential of cyberinfrastructure to drive advances in fundamental research and in faster transfer of research innovation into practice. And while the NEEShub is currently populated mostly by earthquake engineering data funded through NSF's NEESR program, it is certainly available to and useful for fields beyond earthquake engineering (Dyke et al., 2010), as the NEES data repository is useful for researchers and practitioners in the broader community, including international researchers, social scientists, government agencies, emergency responders, and seismologists.

Acknowledgements

References

Abrams, S., Kunze, J., & Loy, D. (2010). An emergent micro-services approach to digital curation infrastructure. *International Journal of Digital Curation, 5*(1), 172-186. doi:10.2218/ijdc.v5i1.151

Balafas, K., & Kiremidjian, A.S. (in press). Reliability assessment of the rotation algorithm for earthquake damage estimation. *Structure and Infrastructure Engineering*.

Bernal, D. (n.d.). *Damping estimation from seismic records.* Manuscript submitted for publication.

Browning, J., Pujol, S., Eigenmann, R., & Ramirez, J.A. (2013). NEEShub databases: Quick access to concrete data. *Concrete International, 35*(4), 55-58. Retrieved from http://nees.org/resources/5768

Carlson, J., Johnston, L., Westra, B., & Nichols, M. (2013). Developing an approach for data management education: A report from the data information literacy project. *International Journal of Digital Curation, 8*(1), 204-217. doi:10.2218/ijdc.v8i1.254

Cheung, A., & Kiremidjian, A.S. (in press). Development of a rotation algorithm for earthquake damage diagnosis. earthquake spectra. *Earthquake Spectra.* doi:10.1193/012212EQS016M

Dyke, S.J., Stojadinovic, B., Arduino, P., Garlock, M., Luco, N., Ramirez, J.A., & Yim, S. (2010). *2020 vision for earthquake engineering research: Report on an OpenSpace Technology Workshop on the future of earthquake engineering.* Retrieved from https://nees.org/resources/1636

Faniel, I.M., & Jacobsen, T.E. (2010). Reusing scientific data: How earthquake engineering researchers assess the reusability of colleagues' data. *Computer Supported Cooperative Work, 19*(3-4), 355-375. doi:10.1007/s10606-010-9117-8

Hacker, T.J., Eigenmann, R., Bagchi, S., Irfanoglu, A., Pujol, S., Catlin, A., & Rathje, E. (2011). The NEEShub cyberinfrastructure for earthquake engineering. *Computing in Science and Engineering, 13*(4) 67-78. doi:10.1109/MCSE.2011.70

Hacker, T., Eigenmann, R., & Rathje, E. (2013). Advancing earthquake engineering research through cyberinfrastructure. *Journal of Structural Engineering*, *139*(7), 1099-1111. doi:10.1061/(ASCE)ST.1943-541X.0000712

Marchionini, G., Lee, C. A., Bowden, H. & Lesk, M. (Eds.). (2012). *Curating for Quality: Ensuring Data Quality to Enable New Science*. Chapel Hill, NC: University of North Carolina. Retrieved from http://datacuration.web.unc.edu/files/2012/10/NSF_Data_Curation_Workshop_Report.pdf

McLennan, M., & Kennell, R. (2010). HUBzero: A platform for dissemination and collaboration in computational science and engineering. *Computing in Science and Engineering, 12*(2), 48-53. doi:10.1109/MCSE.2010.41

The National Archives. (2011). *PRONOM vocabulary specification: DRAFT.* Retrieved from http://test.linkeddatapronom.nationalarchives.gov.uk/vocabulary/ pronom-vocabulary.htm

National Research Council. (2011a). *Grand challenges in earthquake engineering research: A Community Workshop report*. Retrieved from National Academies Press website: http://www.nap.edu/catalog.php?record_id=13167

National Research Council. (2011b). *National earthquake resilience: Research, implementation, and outreach.* Retrieved from National Academies Press website: http://www.nap.edu/catalog.php?record_id=13092

National Science Foundation. (2011). *George E. Brown, Jr. Network for Earthquake Engineering Simulation Research (NEESR) program solicitation* (NSF 11-512). Retrieved from http://www.nsf.gov/pubs/2011/nsf11512/nsf11512.htm

NEES Consortium, Task Group on Information Technology Vision. (2007). *Information technology within the George E. Brown, Jr. Network for Earthquake Engineering Simulation: A vision for an integrated community* (NEES Report No. 07-01). Davis, CA: NEES. Retrieved from http://www1.coe.neu.edu/~jfhajjar/home/ NEES%2007-01%20NEES%20IT%20Vision%20Report.pdf

NEES. (2011). *NEEScomm data sharing and archiving policy.* Retrieved from https://nees.org/resources/2811

NEES. (2012a). *NEEScomm guidelines for data upload.* Retrieved from https://nees.org/resources/4757

NEES. (2012b). *NEEScomm requirements for curation and archiving of research data.* Retrieved from https://nees.org/resources/4759

Peng, J., & Law, K.H. (2004). *Reference NEESgrid data model* (Technical Report NEESgrid-2004-40). Retrieved from https://nees.org/resources/798

Piwowar, H., & Vision, T.J. (2013). Data reuse and the open data citation advantage. *PeerJ PrePrints, 1*, e1v1. doi:10.7287/peerj.preprints.1v1

Letvin, E., & Pujol, S. (2013, September). *Using the HUB to Study Disaster Events.* Presentation given at HUBbub 2013, Indianapolis, IN. Retrieved from https://hubzero.org/resources/1042

Qin, J., Ball, A., & Greenberg, J. (2012). *Functional and architectural requirements for metadata : supporting discovery and management of scientific data.* Paper presented at the 12th International Conference on Dublin Core and Metadata Applications, Kuching, Singapore. Retrieved from http://dcevents.dublincore.org/IntConf/dc-2012/paper/view/107

Ramirez, J.A. (2012). The George E. Brown, Jr., Network for Earthquake Engineering Simulation (NEES): Reducing the impact of EQs and tsunamis. In *Proceedings of the 15th World Conference on Earthquake Engineering, Lisbon, Portugal.* Retrieved from https://nees.org/resources/13178

Socha Y.M. (Ed.). (2013). Out of cite, out of mind: The current state of practice, policy, and technology. *Data Science Journal, 12*, CIDCR1-75. doi:10.2481/dsj.OSOM13-043

Van Den Einde, L., Fowler, K., Rowley, J., Krishnan, S., Baru1, C., & Elgamal, A. (2008). The NEES data model in support of earthquake engineering research. In *Proceedings of the 14th World Conference on Earthquake Engineering, Beijing, China.* Retrieved from https://nees.org/resources/654

Vermaaten, S., Lavoie, B., Caplan, P. (2012). Identifying threats to successful digital preservation: The SPOT model for risk assessment. *D-Lib Magazine, 18*(9/10). doi:10.1045/september2012-vermaaten

Voss, A., & Procter, R. (2009). Virtual research environments in scholarly work and communications. *Library Hi Tech Journal, 27*(2), 174-190. doi:10.1108/07378830910968146