# The NEES Data Model in Support of Earthquake Engineering Research

**L. Van Den Einde[1], K. Fowler[1], J. Rowley[1], S. Krishnan[1], C. Baru[1], A. Elgamal[2]**

[1] *NEES Cyberinfrastructure Center, San Diego Supercomputer Center, University of California, San Diego, La Jolla, CA,USA*

[2] *Professor, Dept of Structural Engineering University of California, San Diego, La Jolla, CA,USA*

*Email: lellivde@sdsc.edu, kfowler@sdsc.edu, jrowley@sdsc.edu, sriram@sdsc.edu, baru@sdsc.edu, elgamal@ucsd.edu*

**ABSTRACT:**

The George E Brown Jr. Network for Earthquake Engineering Simulation (NEES) is a shared national network of 15 experimental facilities, collaborative tools, a centralized data repository, and earthquake simulation software, all linked together to enable engineers to develop better and more cost-effective ways of mitigating earthquake damage. To support data organization, archiving and dissemination resulting from NEES related research, the NEES Cyberinfrastructure Center (NEESit) has developed an abstract data model for earthquake engineering projects and an implementation of that data model through the NEEScentral Data Repository application. The NEES Data Model represents classes of entities and relationships to provide essential data for reproducing experimental and computational simulation results. It is based on a hierarchical structure, consisting of Projects, Experiments, Trials and Repetitions, that collects highly detailed information from each experiment or computational simulation in order to be able to reproduce the research. The NEEScentral Data Repository provides data management infrastructure to shorten the gap between research and practice through long-term preservation and sharing of data. The ultimate goal is to provide access to different types of experimental and computational data to promote and facilitate collaborative and interactive processes required to address the complex nature of seismic events and their physical and societal impacts, ultimately reducing vulnerabilities from earthquakes. This paper describes the details of the NEES data model and presents an overview of the NEEScentral data repository application. Future directions for the NEES Data Model are provided.

**KEYWORDS:** Data, Data Model, Repository, Information Technology, NEES

## 1. INTRODUCTION

The George E. Brown Jr. Network for Earthquake Engineering Simulation (NEES) is a project funded by the National Science Foundation (NSF) to provide a widely accessible, state-of-the-art science and engineering infrastructure through experimental testing facilities and the application of advanced computers, networking, and software. NEES integrates 15 facilities that provide equipment for NEES research and shared use projects through a unique infrastructure managed by the NEES Cyberinfrastructure Center (NEESit), a multi-institutional organization supporting the IT needs of the earthquake engineering community, including researchers, practitioners, educators, students, and information technology specialists [1].

A principal motivation driving the cyberinfrastructure component is the need for a curated data repository for storing and sharing earthquake engineering data. Historically, this wealth of data becomes inaccessible after experiment/simulation completion because 1) it may be stored in a proprietary format or without adequate descriptive information for reuse, 2) it may be placed on physical media (such as zip disks) that soon become archaic and therefore unreadable, or 3) it is not transferred to an accessible location after the experiment/simulation is completed. As a result, researchers wishing to study results from previous research projects must often track down the data, find the hardware that is required to read it, and decipher the storage format before the data are accessible. Even data that have been preserved by a single facility may be lacking any standard organization or the metadata required to reproduce the experiment or numerical simulation [2]. Long-term preservation of properly curated data in an accessible format is necessary to ensure that future research can benefit from current earthquake engineering experiments, shortening the gap between research and practice. These NEES experimentation and cyberinfrastructure resources provide the means for collaboration and discovery and facilitate more advanced research that investigates the performance of buildings, bridges, utility systems, coastal regions, and geomaterials during seismic events.

A data model is an abstract model that describes how data is represented and used [3]. Data modeling is the process of structuring and organizing data. These data structures are then typically implemented in a database management system. In addition to defining and organizing the data, data modeling imposes (implicitly or explicitly) constraints or limitations on the data placed within the structure. This essentially tries to standardize the data so that easier queries can be accomplished. Data sets are typically associated with metadata, which is structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource. The term metadata is used differently in different communities. An important reason for creating descriptive metadata is to facilitate discovery of relevant information. In addition to resource discovery, metadata can help organize electronic resources, facilitate interoperability and legacy resource integration, provide digital identification, and support archiving and preservation [4].

The NEES Cyberinfrastructure center has developed the NEES Data Model for Earthquake Engineering in close collaboration with researchers from the domain community. The goal is providing enough metadata about an experiment or simulation so it can be effectively reproduced and to foster the open exchange of data, information, and knowledge among earthquake engineering researchers and practitioners. The data model provides an easy mechanism for researchers to structure and organize their data using predefined hierarchies and classes of information. Additionally, the structure of the NEES Data Model enables researchers to archive their data in digital formats, which ultimately provides opportunities for the development of automated tools for visualization and data knowledge extraction, with the vision of automated pipelines for data processing into higher-order data products. The following paper presents current efforts in the development of data models for earthquake engineering, and describes the detailed structure and hierarchy of the NEES Data Model.

### *1.1. Related Data Models for Earthquake Engineering*

Several efforts have been undertaken to develop data models for earthquake engineering research. Experimental databases have been developed over the years to archive results from structural testing. The SAC Steel Project Database [5] provides results from an extensive test program that investigated the damage to welded steel moment frame buildings in the 1994 Northridge earthquake and developed repair techniques and new design approaches to minimize damage to steel moment frame buildings in future earthquakes. Test summaries and results of steel connection tests, as well as latest developments in design and construction methods for steel moment frame buildings are provided. The PEER Structural Performance Database [6] provide researchers with the data needed to evaluate and develop seismic performance models for reinforced concrete columns with results from over 400 cyclic, lateral-load column tests.

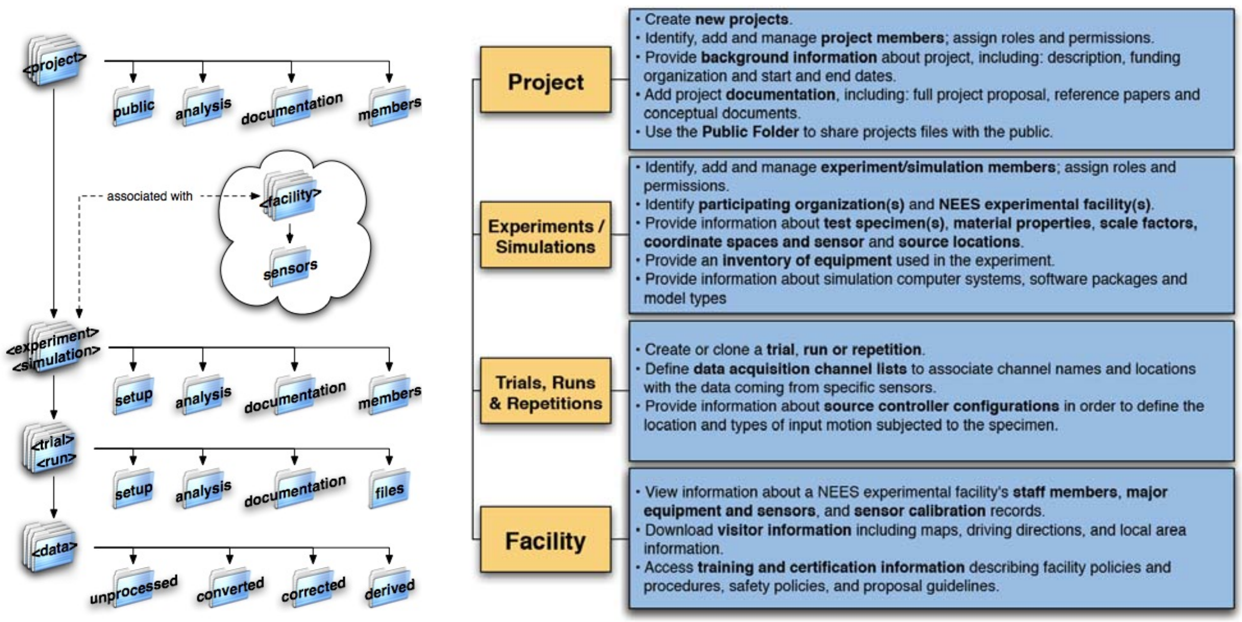
Before NEES became operational in 2004, a task force comprising of earthquake engineering researchers evaluated existing data standards and models such as SensorML, ISO, and STEP. The NEES data model development effort was then formally led by Dr. Kincho Law with the culmination of the NEESgrid Reference Data Model [2]. The reference model tried to capture commonalities among various disciplines of research with the ultimate goal to support the major activities involved in earthquake engineering simulations. Although the reference data model focused on shake table experiments, many of the features are applicable or can be extended to centrifuge, field tests, tsunami, pseudo-dynamic and other types of experiments. Six base classes and their subclasses were created, and the relationships among these classes were defined [2].

Oregon State University and the Northwest Alliance for Computational Science and Engineering (NACSE) have developed a data model for describing laboratory tsunami experiments [7]. The model contains relationships among projects, experiments, researchers, equipment, experimental results, etc. The model is designed for storing the experimental data in a relational database. The general hierarchy structure of the NEES Data Model closely represents the NACSE work.

Researchers at the NEES Real-Time Multi-Directional (RTMD) testing facility at the ATLSS Center at Lehigh University developed a data model for large-scale structural laboratory experiments [8]. The model supports both typical structural experiments as well as hybrid pseudo-dynamic tests, which require specific information describing physical and numerical components of the hybrid simulation, as well as classes to represent the communications among the simulation coordinator at the site.

## 2. ELEMENTS OF THE NEES DATA MODEL

A data model represents classes of entities that are of interest to store information, the attributes of that information, and relationships among those entities and attributes. The NEES Data Model is organized in a conceptual hierarchy as shown in Figure 1a. Figure 1b provides a general overview of the main data organizational functions performed at each level of the hierarchy. Included in Figure 1b is a description of the Facilities class which is described in more detail in Section 3.7.




a) Hierarchy Structure    b) Functions at Each Level

Figure 1: NEES Data Model

The hierarchy provides a single, high-level, standardized model for storing data that is universal to all disciplines in NEES. It is based on four directory levels: project, experiment or simulation, trial or run, and data. At the coarsest level of the hierarchy is the project directory, which provides a general container for secure data sharing within a collaborative team. The second level provides extensive mechanisms for inputting metadata about the setup of an experiment or simulation. Experiments and simulations contain one or more trials or runs that define further changes to configuration parameters. At the lowest level of the hierarchy are directories for archiving the data associated with each experiment. Although the focus of the data model is to collect structured data and metadata to improve querying capabilities, the NEES Data Model still supports the archival of files.

## 3. CLASS AND ATTRIBUTES OF THE NEES DATA MODEL

The four main levels of the NEES Data Model hierarchy are represented by numerous interrelationships to accurately characterize an earthquake engineering experiment or computational simulation. The current NEES Data Model consists of hundreds of entities implemented by over 130 tables in a relational database, making a traditional entity-relationship diagram difficult to graphically represent. Figure 2 describes primary classes in the NEES Data Model with some of the key entities (or metadata fields) for each class, which are described in more detail in the following sections [12].

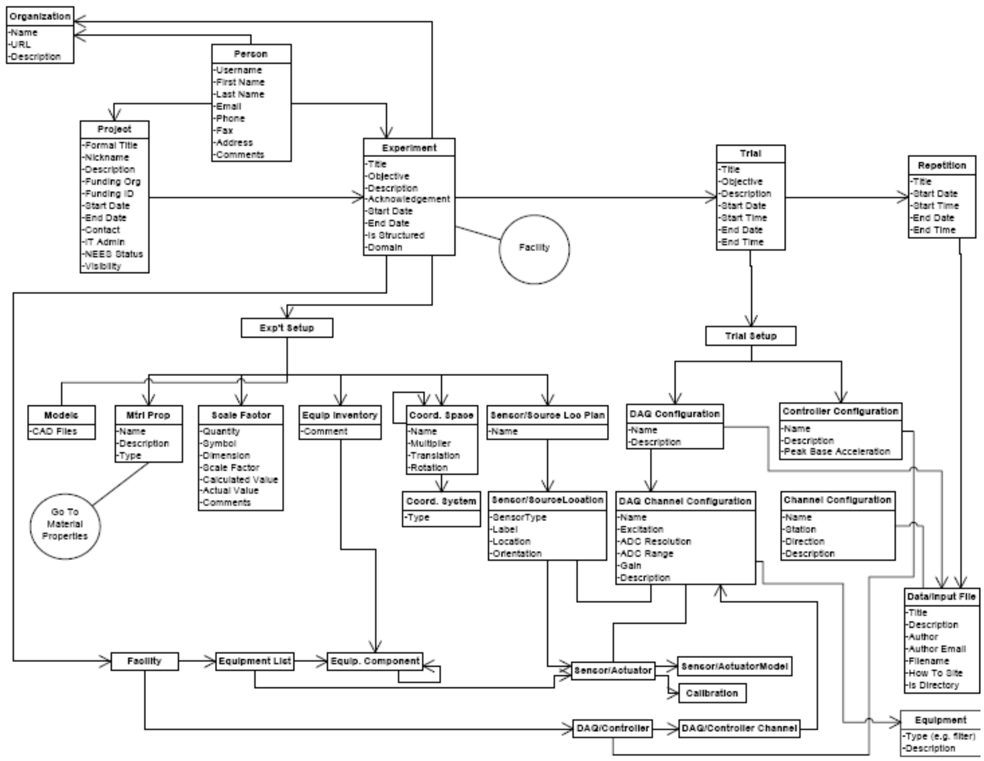



Figure 2: High Level Classes in the NEES Data Model

### *3.1. Core Data Model*

3.1.1 Project

The project entity in the NEES Data Model represents a large earthquake engineering endeavor. Projects can be associated with experimental, computational simulation, and reconnaissance tasks. High level information collected for each project within NEES consist of formal project title and nickname, description, funding agency, project visibility (whether the project is viewable in the NEEScentral database web application), publishing status (publicly available or private), whether the research is NEES related, project type (either structured following the NEES Data Model or an unstructured file system), and contact information.

3.1.2 Experiment

An experiment (physical or computation simulation) represents a subset of a project. High level information for each experiment/simulation include title, experiment structure (either structured following the NEES Data Model or an unstructured file system), experiment domain (shake table, centrifuge, tsunami, large-scale, field, other), start/end dates, objective, description, acknowledgement, participating organizations, NEES Facility where the research was conducted, experiment visibility (whether the experiment is viewable in the NEEScentral database web application), and the ability to upload an experiment thumbnail that represents the research with one graphical image.

*3.1.2.1 Experiment Setup*

Within the Experiment Setup section, researchers are able to describe details about their experiment such as material properties, coordinate spaces, sensor location plans, source location plans, equipment inventory, scale factors, and models. Information about each of these classes is described below:

- *Measurement Units:* This section provides a system for defining default measurement units used for all numeric fields. Numeric fields are found in many of the Experiment Setup classes such as material properties, sensor location plans, source location plans, and coordinate spaces. This feature simplifies metadata ingestion for a researcher. However, a researcher is still able to customize units for each field.

Categories of measurements include acceleration, angle, area, density, distance, mass, pressure, temperature, time, and velocity. SI and English systems are supported.

- *Material Properties:* This section provides the capability to describe and store information about the characteristics of the materials used in an experiment or simulation. It consists of material name, description of the material, and material type. Predefined material type choices include Composite Steel, Concrete, Masonry, Rebar, Rock, Soil–Clay, Soil–Sand, Steel, Wood, and Other. Many of the material types also have the ability to provide more detailed information such as modulus of elasticity, yield stress, ultimate stress, compression strength, tensile strength, poison's ratio, specific gravity, permeability, and relative density, to name a few. Additional information about the materials such as stress-strain data from material testing can be included as files.
- *Coordinate Spaces:* Each experiment can define Coordinate Spaces to describe important geospatial information about a test setup. For example, a cantilever bridge column test may have sensors located around the column that are measured from a particular reference point such as the top of the footing. Other instrumentation and equipment may be measured from a more global reference point such as the corner of the laboratory floor. Researchers are able to define the global coordinate space, and one or more "child" coordinate spaces can be defined relative to the global coordinate space.
- *Sensor Location Plans:* Every experiment is composed of numerous sensors/instruments to help characterize the performance of the specimen. It is imperative to know where the sensors are located in three dimensional space in order to understand the data coming from the data acquisition systems. The NEES Data Model defines Sensor Location Plans that allow researchers to describe what type of sensor they are using (i.e., accelerometer, temperature, pressure transducer, etc.), and where it is located in space relative to a predefined Coordinate Space.
- *Source Location Plans:* Every experiment is composed of one or many input sources that impose a load, displacement or vibration on the specimen. The NEES Data Model defines Source Location Plans that allow researchers to describe what type of actuator or source they are using (i.e., actuator, hammer, other), and where they are located in space relative to a predefined Coordinate Space.
- *Equipment Inventory:* Equipment Inventory allows you to identify the pieces of equipment used in your experiment. This feature is only available for experiments performed at NEES equipment sites. To enable the dynamic equipment inventory, an experiment must be associated with the NEES facility (or facilities) where the experiment was performed.
- *Scale Factors:* In support of shake table and centrifuge tests where scale modeling of any large-scale nonlinear problem for which gravity is a primary driving force is critical, the NEES Data Model provides the ability to define geometric, material, and loading scale factors based on standard similitude laws. Researchers are able to provide values for several independent scale factors and automatically calculate values for dependent scale factors. A researcher is able to provide actual scale factors if they differ from derived scale factors, and provide documentation.
- *Specimen/Models:* Currently, specimen geometry and test setup configuration is not part of the NEESit relational data model but are included as files (specimen drawings, photos). The current use of drawings can sufficiently represent the specimen and its setup, but does not properly characterize a specimen's geometry in a relational database that can be queried upon. Several groups such as STEP [9] are working on defining complex models to support computer aided drawing applications.

*3.1.2.2 Simulation Setup*

- *Computer Systems:* If "Simulation" is selected as the experiment domain, the data model changes to support the metadata required to reproduce computational simulation research. Under Computer Systems, information such as the computer system name, the organization it belongs to, the hardware and operating system used, and the software and version number are collected.
- *Model Types:* Under Model Types, a researcher can provide a model name, description and then select which type of model best characterizes the analysis such as beam, bridge, building, column, fluid, foundation, pipeline, or soil. Files can be uploaded to better describe the model geometry.

3.1.3 Trials/Runs

Experiments can contain multiple trials where only minor changes to the configuration parameters defined at the experiment level occur. Minor changes may include modified loading protocol, removal or relocation of a small subset of instruments, or slightly different material properties. For some experiments, such as a single column tested pseudo-dynamically, only one trial should be created to store the collected data. For larger experiments like multiple shake-table tests using the same structure/specimen, separate trials should be created. For Simulations, trials are often referred to “Runs” and special metadata is collected to better describe each computational run such as the simulation type (cyclic-pushover, dynamic, pushover, other), load type (elemental, multi-support, nodal, uniform excitation, other), element type (beam-column, shell, solid, other), and also includes the objective, description, and start and end times fields like “Trials”.

*3.1.3.1 Trial Setup:*

- *Source Controller Configurations:* Researchers can associate a channel number or name with each of the locations defined in the source location plan(s), and indicate other details and descriptions about each source. Information collected for source controller configurations include channel name (or number associated with the source), location plan (the source location plan in which the researcher’s source's location is defined), location (the name that references the source's location), and sensor (assign a particular source to the chosen location). Additionally, station location of the recording equipment and direction of motion can be inputted if using a real earthquake record.
- *DAQ Configurations:* Researchers can associate DAQ channels with locations defined in the researcher’s sensor location plan(s), and indicate details about the sensor placed at that location. Key information collected include channel number (column of data in the output file that is associated with this sensor), location plan (sensor location plan associated with this particular sensor), and label (name that references the sensor's location). Other fields that help describe individual sensors include ADC Range, ADC resolution, gain, excitation, and a description of the channel's input, output, and purpose.

3.1.4 Repetitions

In some cases, a trial may be run multiple times without any changes to the setup or trial parameters and the data averaged in an attempt to improve data quality. In these cases, a user should create a trial repetition. A trial repetition only allows a researcher to store multiple data sets for different runs without changing any setup or configuration information. If the configuration changes in any way, a user must create a new trial. The only difference between the terms "trial" and "run" is that trials are associated with experiments and runs are associated with simulations.

*3.1.4.1 Data Folders*

For each trial, the data model supports a file folder structure to collect the output data measured by the data acquisition systems from the sources and sensors. The four folders are Unprocessed Data, Converted Data, Corrected Data, and Derived Data. The Unprocessed Data folder is for the lowest, most basic form of data coming from the DAQ sensor. Usually this is in volts or some other non-meaningful unit, but some equipment sites have sensors that automatically convert data into engineering units. Data in the three other data folders should be traceable back to data in the Unprocessed Data folder. Converted Data is for data that has been converted to more useful units. These are simple conversions, such as changing voltages to strain, curvature, displacement, etc. Corrected Data is for cases when problems occur during testing and data must be cleaned or revised to compensate for calibration problems, to eliminate noise, or to apply an overall correction factor. Finally, Derived Data is for data generated by using information from the Converted Data folder to plot, compare, or analyze results. Examples of derived data include plots of strain or displacement profiles along a specimen axis; shear profiles derived from accelerometer or other instrumental data; and derived FFTs. Documents that fully describe all modifications, versioning information, and assumptions made when converting the unprocessed data to converted, corrected or derived data should be included.

***3.5. Security and Authentication***

Each project can associate team members who usually belong to different organizations but are assigned to activities for the same project, thereby allowing them to collaborate as one team. Each project team member is required to register for a NEEScentral account. Information such as first name, last name, password, primary organization they belong to, role within the organization, email, address, phone and fax are collected upon registration. Each project can manage their membership and edit roles and permissions for each member. The various roles available on a project consist of Principal Investigator, Co-PI, Grad Student, Undergrad, Collaborator, Post Doc, Research Scientist, Industry Partner, Visiting Scholar, IT Administrator, IT Programmer, Curator, Site Operations Manager, Technicians, and Other. Each pre-defined role (e.g., principal investigator) is associated with a default set of permissions that control how the member is allowed to interact on the project. These default permissions are easily customized by project administrators to allow for flexible access. Permissions include View, Create, Edit, Delete, and Grant.

### *3.7. Facilities*

The NEES Data Model includes a significant portion related to fully characterizing the NEES equipment site facilities. This includes basic information about the facility, contact information, staff, training and certification, and education and outreach as well as very specific details about the equipment and sensors that are available at the facility. Goals of "Facilities" are to provide enough information to potential researchers about the current state of a facility so that they can write reasonable proposals to utilize the equipment site(s) and provide calibration information about sensors and instruments and keep an historical archive of these records.

#### 3.7.1 Equipment and Sensors

The Equipment class describes major equipment and sensors available at the facility. Each piece of equipment can be described by specific attributes and are associated with downloadable documents obtaining further information. If the piece of major equipment is comprised of multiple subcomponents, details about the subcomponents can also be stored. Each facility's equipment list includes a Sensor List. This inventory of sensors with calibration information is important for fulfilling the vision of the NEES Data Model in enabling researchers to associate specific facilities with their experiments so they may quickly and accurately indicate which sensors (specific serial numbers) were used in an experiment.

## 4. IMPLEMENTATION OF NEES DATA MODEL: NEEScentral

NEEScentral (http://central.nees.org) is the data repository for the NEES initiative. It provides a centralized location for researchers to securely organize, store, and share data and metadata in a nonproprietary format. NEEScentral consists of an Oracle based implementation of the NEES data model, with a PHP based front-end to provide secure Web based access to the back-end database implementation. It is the mechanism by which researchers can upload, view, and download data and metadata. In NEEScentral, several methods to upload data and metadata are available. The most straightforward method is through the user interface, where different pages of forms are provided to allow the user to add information about their project, experiment, specimen, simulation, etc. ASCII data files as well as other file formats representing specimen configurations, instrumentation plans, research proposals, video and images, to name a few can be uploaded to many locations in the file hierarchy of the data model using the NEEScentral upload applet. This is a bulk upload capability that allows the user to upload one or many files to a directory. Once the files have been uploaded, the researcher may add metadata (details about the file) to each individual file to enhance search capabilities. NEEScentral provides secure methods to store, modify, search, and retrieve data and metadata to be used in data manipulation and visualization tools. NEEScentral offers the NEES Web Services which provide secure interfaces for external access to the data repository. This enables application developers to create applications for researchers to control the data in the repository without needing to understand the implementation details.

## 6. CONCLUSIONS

The NEEScentral Data Repository and underlying data model provides data management infrastructure to shorten the gap between research and practice through long-term preservation and sharing of data. The ultimate goal is to provide access to different types of experimental and numerical data to promote and facilitate collaborative and interactive processes required to address the complex nature of seismic events and their physical and societal impacts, ultimately reducing vulnerabilities from earthquakes. Through this structuring of data and metadata, searchable archived data sets can be shared with the broader community and used to further advance the understanding of current earthquake engineering problems. The diverse research within the NEES network requires data models that are flexible and extensible, as well as capabilities for data federation that seamlessly tie together multiple data sources into a federated system [10]. Data federation in NEES involves several concepts including tools like DB2 Information integration [10], GEON Data Integration Cart™ [11], and Custom Integration tools. These federation tools can allow disparate data sources residing locally or in remote locations to be integrated and searchable. This will allow for flexibility in data access.

Once published, scientific data should remain available forever so that other engineers can reproduce the results and do new science with the data. Data is incomprehensible and hence useless unless there is a detailed and clear description of how and when it was gathered, and how the derived data was produced. Librarians would describe documenting the metadata as curating the data [13]. Through the structured data model described in this paper, NEES has already begun the process of curating the data in the NEEScentral repository, a process which requires community feedback and participation. Curation should be able to show the evolution of data in the repository. It should also provide automated checklists to determine that the research data conforms to the curation standards (data completeness). Integral to the curation process is a community rating system, which can show how complete and/or effective a dataset is based on the opinions of earthquake engineering colleagues.

## 7. ACKNOWLEDGEMENTS

This research project is sponsored by the National Science Foundation through NEES under award number CMS-0402490. The authors would like to acknowledge the support of NEESinc, as well as Dr. Kincho Law and Dr. Jun Peng, and NACSE at Oregon State University for their significant and early contributions to the development of a NEES-wide data model. Additionally, development was based on extensive collaborations with the earthquake engineering community through the data use case group including The authors acknowledge the contributions of the NEES Data User Committee consisting of Bruce Kutter, Dan Wilson, Sharon Wood, Adolfo Matamoros, Saiid Saiidi, Sherif Elfass, Gokhan Peckan, Christopher Stanton, Claude Trottier, M. Llankatharan, and A. Vosooghi. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the sponsor.

## REFERENCES

[1] Whitmore, S., Van Den Einde, L., Warnock, T., Diehl, D., Hubbard, P. and Deng, W., 2006, NEESit Software Overview: IT Tools that Facilitate Earthquake Engineering Research and Education, 100th Anniversary Earthquake Conference (8th National Conference on Earthquake Engineering, 8NCEE), San Francisco, CA.
[2] Peng, J., and Law, K., 2004, Reference NEESgrid Data Model, Technical Report NEESgrid-2004-40, Network for Earthquake Engineering Simulation, http://it.nees.org/documentation/pdf/TR-2004-40.pdf.
[3] http://en.wikipedia.org/, data modeling, data models
[4] Understanding Metadata, http://www.niso.org/standards/resources/UnderstandingMetadata.pdf
[5] SAC Steel Project, http://www.sacsteel.org/
[6] Pacific Earthquake Engineering Research Center (PEER) Structural Performance Database, http://nisee.berkeley.edu/spd/
[7] Oregon State University and Network Alliance for Computational Science and Engineering. NEES Database and Metadata Structure, Version 1.3, white paper, Network for Earthquake Engineering Simulation, 2003.
[8] Lehigh Data Model Report, http://www.nees.lehigh.edu/index.php?page=lehigh-data-model
[9] STEP, http://www.steptools.com/
[10] IBM, 2005, "Using data federation technology in IBM WebSphere Information Integrator: Data federation design and configuration", http://www.ibm.com/developerworks/db2/library/techarticle/dm-0506lin/.
[11] GEON Data Mediation, http://geongrid.org/
[12] NEEScentral User Guide, http://it.nees.org/library/data/neescentral-users-guide-17.php
[13] Gray, J., Szalay, A.S., Thakar, A.R., Stoughton,C., and vandenBerg, J., Online Scientific Data Curation, Publication, and Archiving, July 2002, http://research.microsoft.com/research/pubs/view.aspx?msr_tr_id=MSR-TR-2002-74.
[14] Van Den Einde, L., et al, "The NEEScentral Data Repository: A Framework for Data Collaboration in Earthquake Engineering", Companion paper for 14WCEE, Beijing, China, October 2008.