

# Development of a Data Model for EDgrid (E-Defense Grid)

Nobuyoshi Yabuki<sup>1</sup>, Yoshihiro Yoshida<sup>2</sup>

**Abstract:** E-Defense is developing a system which enables all earthquake engineering researchers to freely access to valuable experimental data obtained during shake table tests executed there after some period. EDgrid is a cyberinfrastructure being developed for this purpose by using the cutting-edge information technology with NEESit in the United States of America. In this research, a data model has been developed for storing all the data related to experiments performed at E-Defense in the EDgrid central repository correctly, effectively and efficiently. The data model was developed on the basis of the event-based relational data model theory to correspond to possible future facility expansion and collaboration with other databases in various countries.

**Keywords** : EDgrid, E-Defense, data model, database, NEES, NEESit

## 1. Introduction

E-Defense Grid (EDgrid) is a cyberinfrastructure developed for E-Defense<sup>1)</sup> by making full use of the latest information technology, inspired by NEESgrid of NEES<sup>2)</sup> in the US. Cyberinfrastructure, which is constructed for the purpose of development in the field of civil engineering, is a virtual information world supported by the Internet and all kinds of information techniques<sup>3)</sup>. E-Defense, of which E stands for Earth, is an earthquake engineering experimental facility where the world largest shaking table is situated and which was complete in 2005. Various earthquake engineering experiments are executed at E-Defense. All the data, acquired at E-Defense, will be open to the earthquake engineering research community in Japan and the US after certain period. In the EDgrid development research of National Research Institute for Earth Science and Disaster Prevention (NIED), the authors have developed a data model for storing all the experiment data in the EDgrid central repository by the collaboration with the NEESit, i.e., NEES information technology infrastructure. The EDgrid project has been being proceeded by members shown in **Fig. 1**. In E-Defense, 960 sensors, 27 video cameras, 2 High-Definition (HD) video cameras, etc. exist and large amount of data are generated per shake experiment. All the data are acquired by the EDgrid Data Acquisition (DAQ) and stored in the central repository of EDgrid permanently, and will be open to the earthquake engineering research community after certain period<sup>4)</sup>. In this research, the authors developed a data model for storing all the data related to experiments performed at E-Defense in the

EDgrid central repository correctly, flexibly, clearly, accessibly.

## 2. NEES data models

### (1) NEES

NEES is the George E. Brown, Jr. Network for Earthquake Engineering Simulation (NEES) project in the US. NEES is a shared national network of 15 experimental facilities, collaborative tools, a centralized data repository, and earthquake simulation software, all linked by the ultra-high-speed Internet2 connections of NEESgrid. The goal of NEES is that earthquake engineering researchers can access experimental facilities and data to control experimental equipments and obtain experimental data freely.

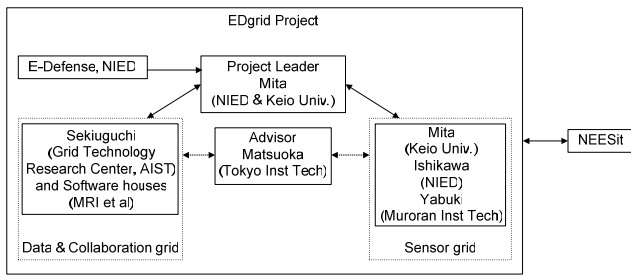
NEESit is in the San Diego Supercomputer Center (SDSC). SDSC is located on the campus of the University of California, San Diego (UCSD). NEESit is an organization to develop a virtual “collaboratory”, i.e., cyberinfrastructure for earthquake engineering experiments and simulations.

### (2) NEESgrid data model

The authors reviewed the data models developed for NEESgrid by Stanford University as our first step to develop the EDgrid data model. In the NEESgrid data model, as shown in **Fig. 2**, the class “NEES”, which is the highest in the hierarchy, is linked with each “Site” class, which represents a distributed experimental facility. Then, each “Site” class is linked to “Experiments”, “Trials”, “Equipment”, and “People” classes. “Experiment” classes

1) M. JSCE, Ph.D., Assoc. Prof., Department of Civil Engineering and Architecture, Muroran Institute of Technology (27-1 Mizumoto-cho, Muroran-shi, Hokkaido, 050-8585, Japan, E-mail: yabuki@news3.ce.muroran-it.ac.jp)

2) Student Member of JSCE, Division of Civil Engineering and Architecture, Muroran Institute of Technology



**Fig. 1** EDgrid project organization

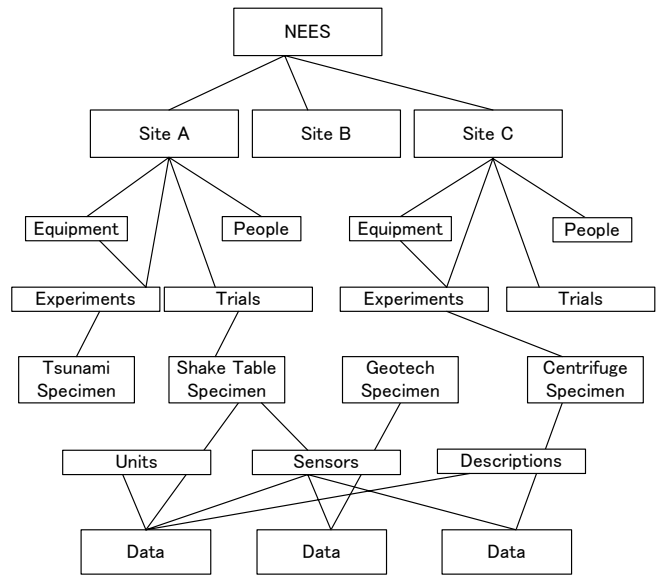
are linked to various “Specimen” classes. Finally, “Specimen” classes are linked to “Sensors”, “Units”, “Descriptions”, and “Data” classes. Both pre- and post-experimental data as well as experimental data are considered to be significant in NEESgrid. Two kinds of data models were developed by Peng and Law for NEES<sup>(5) (6) (7)</sup>. One is an object data model and the other is a relational data model.

The object data model was defined on the basis of the ontology, and Protégé 2008), which is a modeling tool corresponding to Resource Description Framework (RDF) and Web Ontology Language (OWL), was used for implementation.

The relational data model was defined by using tables and relationships among tables. These data models were developed by interviewing many earthquake engineering researchers and by using some actual experimental tests, especially for data generated at shake table facilities. Although these two models were based on different modeling methodologies, they have a similar data structure and each model contains about 20 classes or tables. **Fig. 3** shows NEESgrid relational data model. Conceptually, “Project” is situated at the top and “Task”, “EventGroup”, and “Event” are situated below “Project”, as shown in **Fig. 4**. Thus, in the NEESgrid relational data model, “Project”, “Task”, “EventGroup”, and “Event” are core tables and are related to other tables in either of “Project” level, “Task” level, “EventGroup” level, or “Event” level.

**(3) Evaluating and applying NEESgrid data model**

Both EDgrid and NEESit researchers agreed that the data model to be employed for implementation should be the relational model because the object data model was more difficult for earthquake engineering researchers to understand, and both implementation and maintenance of the object data model seemed to be more time-consuming than the relational model. As EDgrid and NEESgrid will be interrelated, the data models should be similar, if not the same. Thus, EDgrid decided to adopt the relational data



**Fig. 2** Overview of data modeling of NEESgrid<sup>(5)</sup>

model developed by Peng and Law (NEESgrid Data Model) as a basis for the EDgrid data model.

In this research, the authors reviewed NEESgrid data model to examine flexibility and expandability of the data model.

**a) Evaluation of the data model**

The authors investigated the easiness of future changes such as addition of table or definition of new relations in the NEESgrid relational data model. The data model was complicated and could have difficulty in changing the data model in the future. Due to the complexity of the NEESgrid data model, both EDgrid and NEESit researchers had a concern on flexibility and expandability.

**b) Issues in applying NEESgrid data model to EDgrid system**

The EDgrid system consists of connected several systems and its base is database and data model. Lack of flexibility and expandability of the data model could have influence on other computer systems by significant modification of the data model in the future.

Further, in the NEESgrid data model, since there is no table to manage time, contexts between pre- and post-changed data can hardly be understood. Moreover, when data are updated, all previous data are deleted from the database. Thus, NEESgrid relational data model needed some modification and the authors thought that new approach would be necessary in order to build EDgrid data model.

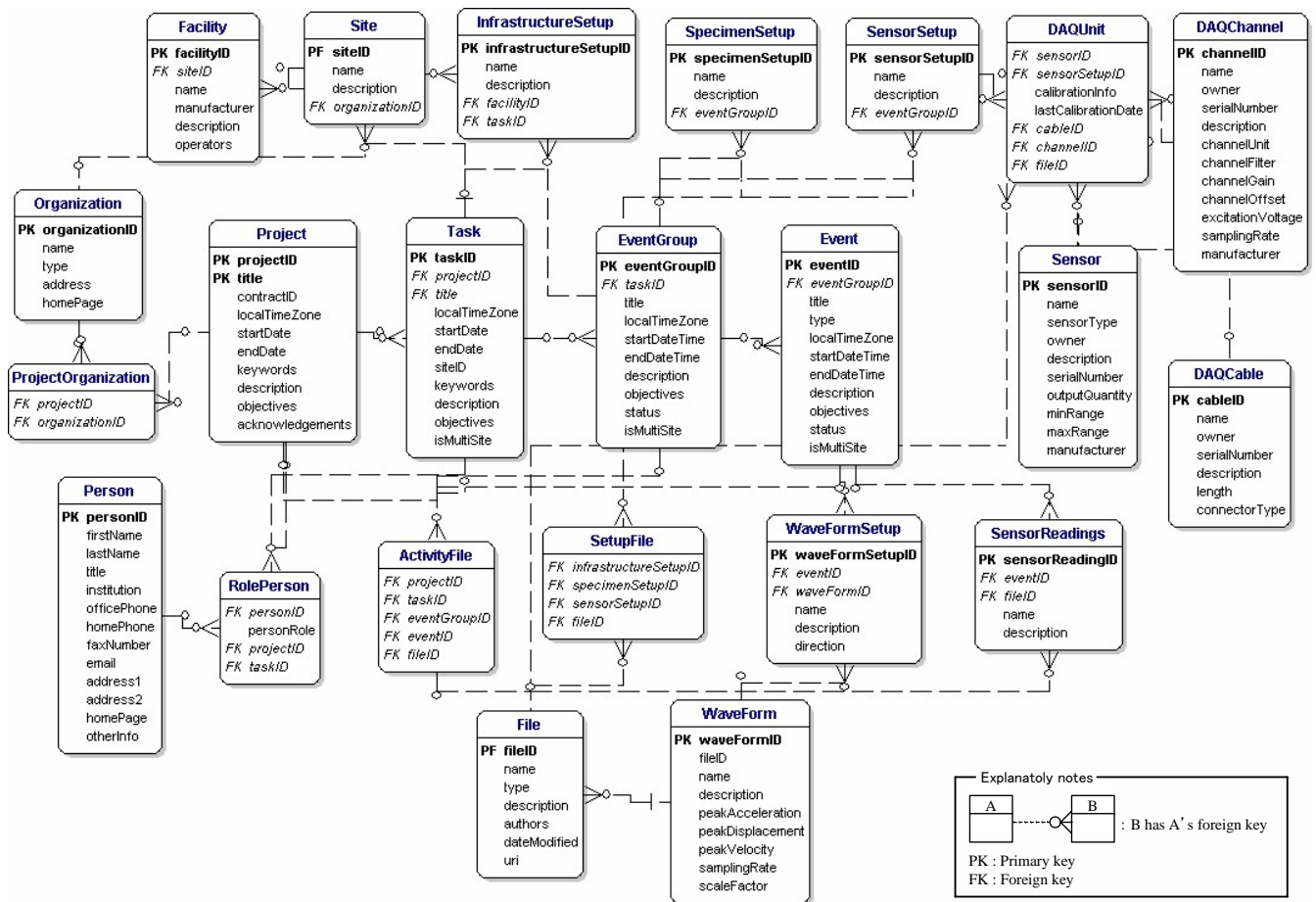


Fig. 3 The referenced data model developed for NEEsgrid

### 3. Event-based data model

In order to simplify the relatively complicated data model and in order for the data model to be flexible and expandable, the authors looked for new concepts. In the collaborative research with NEEsgrid, the event-based data model<sup>9)</sup> was proposed. In the event-based approach, all tables except the table named 'Event' in the relational data model are connected to Event table as its satellites in a two story high hierarchical manner.

The objective of the event-based data model is that it tracks all modifications to the individual classes, providing audit trails for all data and metadata. This approach was deemed suitable and appropriate for EDgrid in this research. For example, by querying "Person" and "Facility" tables we can obtain the facility data where the researcher worked in the past and where he is working now. Thus, the event-based data model allows changing attributes and relationships without losing historic information.

The first data model developed by collaborating with NEEsgrid is shown in Fig. 5. In Fig. 5, "Event" table is situated at the center of the model and has foreign keys of

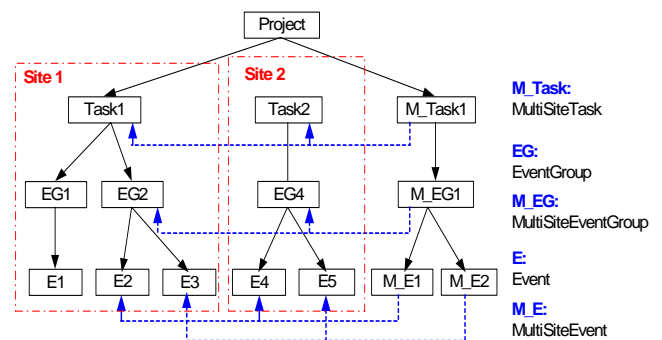


Fig. 4 The concept of NEEsgrid data model

all other tables in order to link all tables except event table as its satellites in a two story high hierarchical manner. However, because "Project" table had a sub table "Experiment" of which sub table was "Trial", the first data model was not the simple exact star-shape structure and since Fig. 5 consisted of only fundamental tables on the basis of event concept, it needed to add tables for using as shake experimental data model.

## 4. EDgrid data model

### (1) E-Defense

E-Defense, constructed on the tenth anniversary of Hanshin-Awaji earthquake in January 2005, is world's largest three-dimensional full-scale shake table and its keywords are “full-scale”, “destruction”, and “destruction”. At this facility, full-scale architectural and civil engineering structures are constructed and destroyed by strong motion and their destruction process is investigated. Through those experiments, earthquake engineering researchers can inspect earthquake-resistant design standards, analytical methodologies, analytical software, and input data and contribution to structural damage reductions is expected. **Table 1** shows basic specifications of its shaking system<sup>1)</sup>.

### (2) EDgrid

Objectives of EDgrid research project are storing all the data acquired by using E-Defense efficiently and safely and realization of remote access through its Internet. EDgrid project consists of two subsystems, data and cooperation grid and sensor grid. **Fig. 6** shows a system architecture of EDgrid Central, which is the EDgrid data repository system. In this research, the authors developed a data model for Relational Database Management System (RDBMS) which is backend of systems shown in **Fig. 6** and implemented a data model by using the open source MySQL<sup>10)</sup> database system.

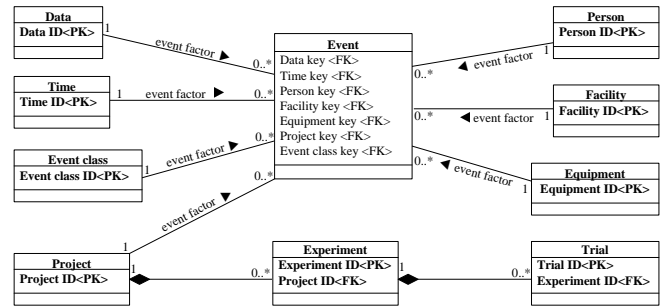
### (3) Investigation of event-based data model

As described before, the authors thought that the event-based data model has more flexibility and expandability than the NEESgrid reference data model and decided to develop the EDgrid data model by applying the abstract concept of “event”.

Since the event-based data model has a star-shape structure, it was assumed that it can change tables and relationships relatively easily compared to NEESgrid reference data model. Further, since event-based data models allow data model developers to define basic and necessary tables relatively easily and without missing in advance, there would be no significant change in the future. If modification of a data model is necessary, this model can be changed easily and the modification can be reserved as an event, the data model change would not give a severe impact to other systems.

### (4) Optimization of the data model

Based on our investigation, the event-based data model has flexibility and expandability adequately and has been adopted as a new approach to develop the EDgrid data model. Since all the data stored in the central repositories of



**Fig. 5** An event based data model

**Table 1** Specification of shake table

Payload	12MN (1,200tonf)
Size	20m×15m
Maximum Acceleration	XY-Horizontal: >900cm/s <sup>2</sup> , Z-Vertical: >1500cm/s <sup>2</sup>
Maximum Velocity	XY-Horizontal: 200cm/s, Z-Vertical: 70cm/s
Maximum Displacement	XY-Horizontal: ±100cm, Z-Vertical: ±50cm
Maximum Allowable Moment	Overturning Moment: 150MN·m Yawing Moment: 40MN·m
Driving Type	Accumulator Charge/Electro-Hydraulic Servo Control

EDgrid and NEESit will be open and shared, harmonization and similarity between the EDgrid and NEESit data models was necessary. The authors developed a data model for EDgrid based on the event-based approach and by trial and error through applying actual experimental data to the initial prototype data model shown in **Fig. 5**, discussing with EDgrid, Stanford University, and NEESit researchers.

**Fig. 7** shows the data model developed for EDgrid and characteristics of this data model are described in the following.

- The “Setting” table was added in the data model in order to manage equipment setting data. It could link same settingID to each equipmentID. Because it was added as single table. It was impossible for the initial prototype data model shown in **Fig. 5** to link those IDs.
- The “Experiment Group” table was added in the data model in order to manage roles in each trial. It could separate individual roles per trial. Because it was added as single table. Since the initial prototype data model

shown in Fig. 5 had one attribute of individual roles, it was impossible to manage roles separately.

- Relationships among “Project”, “Experiment”, and “Trial” tables were removed in order to change the data model into the star-shape two-story hierarchical structure and to manage separately in “Project”, “Experiment”, and “Trial” levels.
- “maxValue” and “minValue” attributes were added in order to express an outline of data. “maxUnit” and “minUnit” were also added in order to reserve numerical and unit data separately.

### 5. Application case of the EDgrid data model

The authors applied actual experimental data to the initial prototype data model developed for EDgrid for verification.

#### (1) Sensor data

Table 2 shows an example of linking each ID in event table by using following tables: equipment, setting trial, and event. “Equipment” table contains the data of sensor types, serial numbers, etc. “Setting” table contains information on how a particular sensor is set up. This table is necessary because the same equipment may be set up differently for various trials or experiments in a project. “Trial” table has information on the trial name, trial date, etc. and these three tables are linked each ID at the lowest table in Table 2.

Equipment and setting information used each experiment can express by using “Event” table. In “Event” table, settingID of 13 used at trial and trialID is 11 and 13. Since “Setting” table is defined as different table, previous setting can reuse in the case of repetition trial and there is no need to input same setting data. Hence, as same setting data, data administrators can manage data by using only ID and volume of data stored in the central repository of EDgrid can reduce. In the NEESgrid data model, a part of data model related to sensor and setting consists of plural tables and relationships, but, in the EDgrid data model, it consists of “Sensor” table and “Setting” table. Thus, EDgrid data model is easy to understand for data administrators.

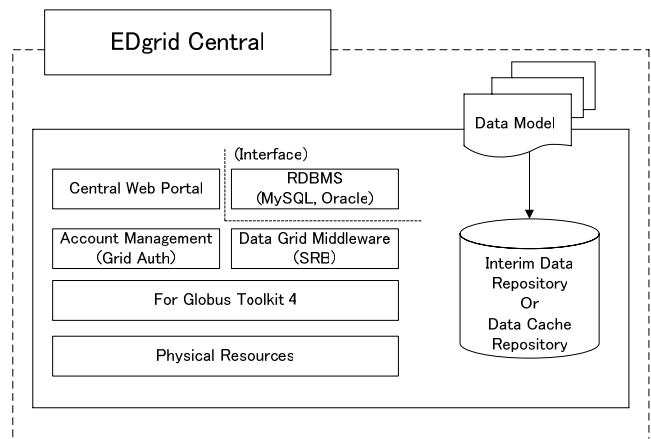


Fig. 6 System architecture of EDgrid

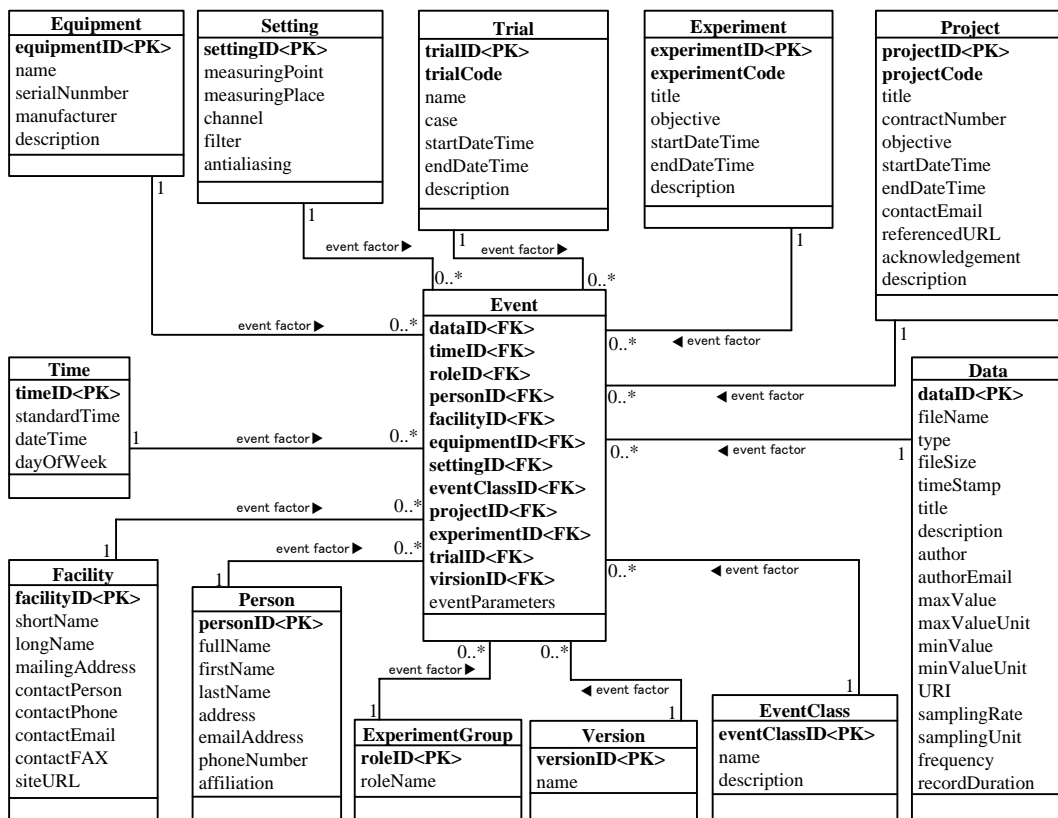


Fig. 7 EDgrid data model

**Table 2** An example of ID relationships at Event table in the EDgrid data model

equipmentID(PK)	name	serialNumber
392	accelerometer	AC1
502	displacement sensor	DPI

settingID (PK)	measuringPoint	measuringPlace	channel	filter	antialiasing
13	91	X direction	32	on	off
393	12	Y direction	12	off	off
401	456	Z direction	43	off	on
598	65	X direction	56	off	off

trialID	trialCode	name	case	startDateTime	endDateTime
11	TRL20050706_001	frequency characteristic trial	random	2005/7/6 9:00:00	2005/7/6 10:30:00
12	TRL20050706_002	frequency characteristic trial	random	2005/7/6 13:00:00	2005/7/6 14:00:00
13	TRL20050708_001	seismic wave	koubeNS	2005/7/8 10:00:00	2005/7/8 11:45:30
14	TRL20050726_001	specimen characteristic trial	random	2005/7/26 10:30:00	2005/7/26 12:10:00
15	TRL20050727_001	seismic wave (applied control)	kushiroNS	2005/7/27 9:30:00	2005/7/27 11:30:00

dataID	experimetGroupID	timeID	personID	facilityID	equipmentID	settingID	eventClassID	projectID	experimentID	trialID	eventParameters
					392	13				11	
					502	393				11	
					392	13				13	
					502	598				13	

The database itself does not contain all sensor data but only their metadata. **Fig. 8** shows a part of the accelerometer data obtained in the demonstration experiment. The sensor data is stored in a hard disk of the central repository and has a Uniform Resource Identifier (URI). “Data” table contains only the metadata including the URI of the data file. Other data such as drawings of experiment models, sensor locations and directions, and various reports, notes, and minutes are represented as electronic data files and their metadata including their file locations URI are stored in the database. Data files are stored in a central repository and can be retrieved by clicking the URI in the database.

**(2) Management of changed data**

Example of modification of data in database is explained by using following tables: ”Person”, ”Time”, “Event”, and “EventClass” tables. “Person” table has information on the personal name, Email address, etc., “Time” table has information on date, day of week, and “Event” table has information on each table ID.

Explanation of “EventClass” table is the following. It is one of the tables situated around “Event” table and specifies combination of keys linked in event table.

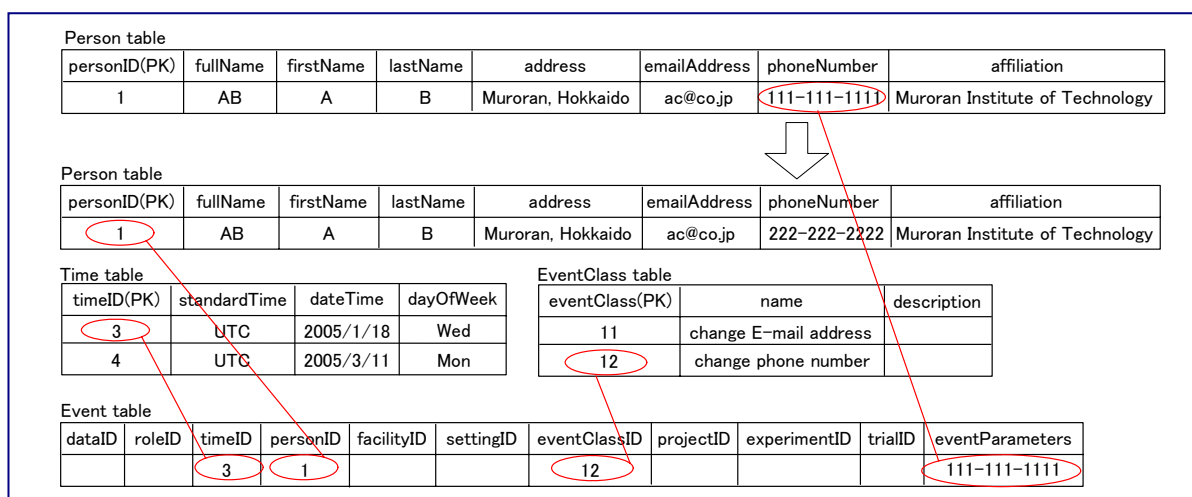
Explanation of eventParameters attribute is the following. **Fig. 9** shows an example of researcher’s telephone number change. In the “Person” table, the personID “1” changed the telephone number from “111-111-1111” to “222-222-2222”. This change was done on Wednesday, January 18, 2005 as shown in “Time” table. This change can be represented in

Date	Time	Acceleration
2005-11-15	09:15:00.00000	1.03000000000000
2005-11-15	09:15:00.01000	0.52806244950202
2005-11-15	09:15:00.02000	0.14515958735064
2005-11-15	09:15:00.03000	0.04865171102717
2005-11-15	09:15:00.04000	0.25213525678815
2005-11-15	09:15:00.05000	-0.30563957891648
2005-11-15	09:15:00.06000	-0.79380155583873
2005-11-15	09:15:00.07000	-0.34439991847798
2005-11-15	09:15:00.08000	0.36189959108080
2005-11-15	09:15:00.09000	1.07372017013149
2005-11-15	09:15:00.10000	1.02225275795905
2005-11-15	09:15:00.11000	-0.08528954703312
2005-11-15	09:15:00.12000	-0.68121178405656
2005-11-15	09:15:00.13000	0.11009846413430
2005-11-15	09:15:00.14000	-0.33580332724757
2005-11-15	09:15:00.15000	-2.37471264646424

**Fig. 8** A part of sample sensor data

“Event” table by filling the attributes of timeID, personID, eventclassID, and eventParameters. The eventParameters attribute can contain any textual data such as the previous telephone number in this example. The “EventClass” table can contain meaning of each event. In **Fig. 9**, the “name” of “EventClass” table linked to “Event” table is “change phone number” and this “name” means the event by linking keys.

The event-based data model can manage historical changing data information in database and “EventClass” table can specify its event. NEESgrid relational data model



**Fig. 9** An example of data change management by using eventParameter

shown in **Fig. 3** didn't have the table to contain changing information. Thus, the event-based data model can deal with modification flexibly.

### (3) Expansion of the data model

The data model shown in **Fig. 7** is a simple star-shape structure as its satellites in a two story high hierarchical manner and can expand the model simply. If new data which can't deal by current data model appears, new table is added in the data model to deal with new data. The method of table addition is shown in **Fig. 10** and is described as the following. New table is situated in the data model and its primary key is added into event table as a foreign key. The method of attribute is also done as well as the method of table addition. The method of attribute addition is shown in **Fig. 11** and is described as the following. New table is situated right under the table that is needed to add attribute and the foreign key is added into new table. The star shape of the data model is not changed and the administrator can manage the data model modification easily and consistently. In the NEESgrid relational data model shown in **Fig. 3**, since relationships of the whole data model were complex and there are a lot of tables compared to event-based data model, expansion of the data model would be difficult. Event-based data model can be expanded relatively easily, but frequent modification causes complexity of data model. Thus, it is data models should be developed so that frequent modification would not be needed.

## 6. Conclusion

In this research, the authors developed the initial data model for experimental data generated at E-Defense to open to earthquake engineering researchers. Characteristics of

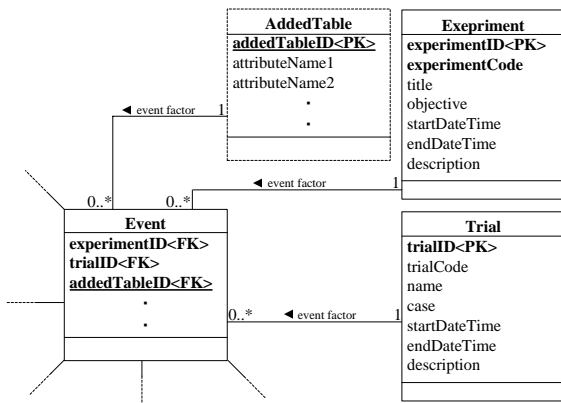
EDgrid data model are as the following.

- The data model was developed by applying the event concept, which was proposed through the collaboration with NEESit.
- Many relationships among tables were cut in order to reduce complexity, and to simplify the whole data model compared to NEESgrid reference data model.
- EDgrid data model can manage time data when data stored in database is changed or updated. However, NEESgrid data model was not able to manage changed data.
- By using eventParameters, previous data is able to be stored in database.

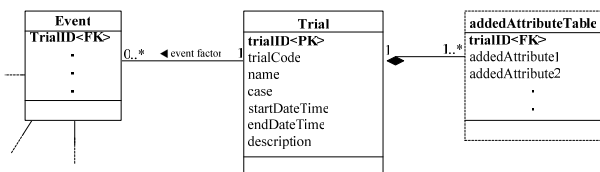
The EDgrid data model developed on the basis of the event-based data model has a simple star-shape structure. The data model can record the data change and can keep both the old and new data in the same table. The databases system can keep the track record and the user can retrieve old but appropriate data. Furthermore, even when the data model itself is modified, such as adding new tables and attributes, the star shape of the data model is not changed and the administrator can manage the data model modification easily and consistently.

The current data model has been implemented in the EDgrid system and has been tested. With some modification, the data model will be complete and the EDgrid system is expected to commence its official operation from April, 2007.

**Acknowledgments:** This research has been partially supported by National Research Institute for Earth Science and Disaster Prevention. The authors would like to thank the EDgrid and NEESit researchers, Prof. Kincho H. Law of Stanford University and Dr. Jun Peng.



**Fig. 10** An example of new table addition



**Fig. 11** An example of new attribute addition

## References

- 1) E-Defense: <http://www.bosai.go.jp/hyogo/index.html>
- 2) NEES: <http://www.nees.org/index.php>
- 3) Yabuki, N: Toward Value Creation by Developing Cyber-Infrastructures, Journal of Construction Management and Engineering, JSCE, No.805/VI-69, pp.1-13, 2005.12.
- 4) Mita, A.: Edgrid as Cyber Infrastructure for E-Defense, Proc. The First Int'l Symp. on Innovation & Sustainability of Structures in Civil Engineering, Nanjing, China, pp.255-260, 2005.
- 5) Peng, J. and Law, K.H.: A Brief Review of Data Models for NEESgrid, NEESgrid TR-2004-01, 2004.
- 6) Law, K.H.: Summary Report on NEESgrid's Data Curation Summit, NEESgrid TR-2004-43, 2004.
- 7) Peng, J. and Law, K.H.: Validity and Usability of the NEESgrid Reference Data Model, NEESgrid TR-2004-44, 2004.
- 8) Protégé2000: <http://protege.stanford.edu/>
- 9) Warnock, T.: NEES Data Architecture, NEESit TR-2005-49, 2005.
- 10) MySQL: <http://www.mysql.com>