A Data Model for Storing a Large Amount of Sensor Data

_____

NOBUYOSHI YABUKI & YOSHIHIRO YOSHIDA

ABSTRACT:

As the importance of structural health monitoring is increasingly recognized and cheaper sensor networks are becoming available, very large amount of sensor data will be stored in many computers in a distributed manner in the near future. For each database storing sensor data of a single structure or a set of structures administered by a designated office, various changes such as changes or addition of sensors, relocation of sensors, modification of structural members, refurbishment, administrators changes, etc., will give an impact on the data and querying methods to the database systems. From a broader point of view, data stored in many local database systems should be somehow integrated and utilized for generating or discovering knowledge based on the long term large amount of sensor data.

To control and correspond to various changes of the sensor environment and to utilize the large amount of sensor data effectively, appropriate data models would be necessary. In this research, a new data model was proposed for solving the problems stated above, on the basis of the event-based relational data model. A data model was developed for storing and reusing experimental data along with their related data generated at E-Defense, the world largest 3D shaking table experiment facility in Japan, as a part of the E-Defense Grid (EDgrid) research project. In this paper, the data model is described and the feasibility of applying the developed data model for sensor data of real structures is discussed.

INTRODUCTION

Since sensor prices are decreasing and cost for installing wireless sensors is much cheaper than wired ones, a large number of sensors, especially wireless sensors will be placed on various structures and facilities and a huge amount of sensor data will be stored in a number of computers in the near future for structural health monitoring. The data should be stored in databases based on appropriate data models in a systematic manner so that researchers and engineers can query and obtain necessary data immediately. Such data models should be able to correspond to future changes possibly occurred in the sensor

_____

Dept. of Civil Eng. & Arch., Muroran Inst. of Tech., 27-1 Mizumoto-cho, Muroran-shi, Hokkaido, 050-8585, Japan, yabuki@news3.ce.muroran-it.ac.jp

environment.

Data models for storing various and large amount of experimental and simulation data have been proposed in the George E. Brown, Jr. Network for Earthquake Engineering Simulation (NEES) project in the US [1] [2] [3]. 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 [4]. Such data models are expected to be utilized as a base for developing data models for storing data of various sensors placed on real structures and field.

E-Defense Grid (EDgrid) is a cyberinfrastructure which is developed by National Research Institute for Earth Science and Disaster Prevention (NIED), inspired by NEESgrid. E-Defense [5], of which E stands for Earth, is an earthquake engineering experimental facility where the world largest shaking table is situated. In E-Defense, 960 sensors, 27 video cameras, 2 High-Definition (HD) video camera, 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 in Japan and the US after certain period [6]. In the EDgrid development research of 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.

NEES DATA MODELS

Two data models were developed by Peng and Law [1] [2] [3] for NEES. 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 2000 [7], 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. Both data models was developed by interviewing many earthquake engineering researchers and by using some actual experimental tests, especially for data generated at shaking table facilities. Although these two models were based on different modeling methodologies, both models have a similar data structure and each model contains about 20 classes or tables.

In this research, these models were reviewed based on the following 4 points. 1) serviceability, 2) easiness of developing a system using a database modeling language, 3) flexibility and expandability of the data model, and 4) influence to other computer systems by modifying the data model in the future. Based on our review, although the data models developed for NEESgrid show reasonably good performance, it was deemed that simplification, flexibility and expandability of the data models would be necessary in order to be used for the EDgrid due to its rapid (two year) development plan.

In the discussion in 2005 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. Thus, EDgrid decided to adopt the relational

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

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 NEESit, the event-based data model [8] 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.

However, the event-based data model proposed by NEESit was not the exact star-shape structure but "Project" class had a subclass "Experiment" of which subclass was "Trial" and its subclass was "Repetition," as shown in Figure 1. For simplification and consistency, the authors decided to relocate "Experiment" and "Trial" classes so that they are directly linked to "Event Fact" class and removed "Repetition" class.
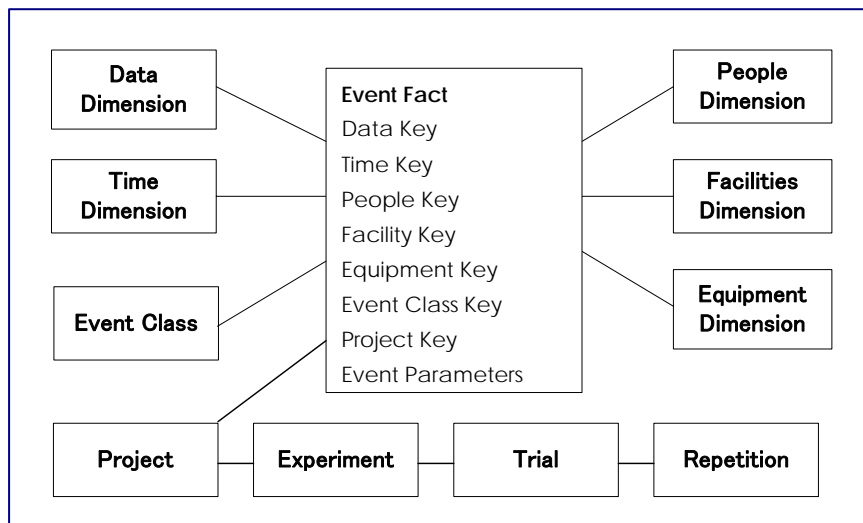

Figure 1: An Event-based Data Model Proposed by NEESit [8]

EDGRID DATA MODEL

The authors developed a data model for EDgrid based on the event-based approach and by trial and error through applying real experimental data to the prototype data models, discussing with EDgrid and NEESit researchers. In the fiscal year of 2005, the EDgrid research group developed a prototype database system based on the data model and demonstrated it to earthquake engineering researchers. They gave us comments and suggestions to the EDgrid system and the authors have modified the data model to meet their requirements.

617

The latest version of the data model for EDgrid consists of the following tables: Event, Project, Experiment, Trial, Setting, Equipment, Time, Facility, Person, ExperimetGroup, Version, EventClass, and Data. Each class contains necessary attributes, as shown in Figure 2.

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 (Uniform Resource Identifiers: 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.
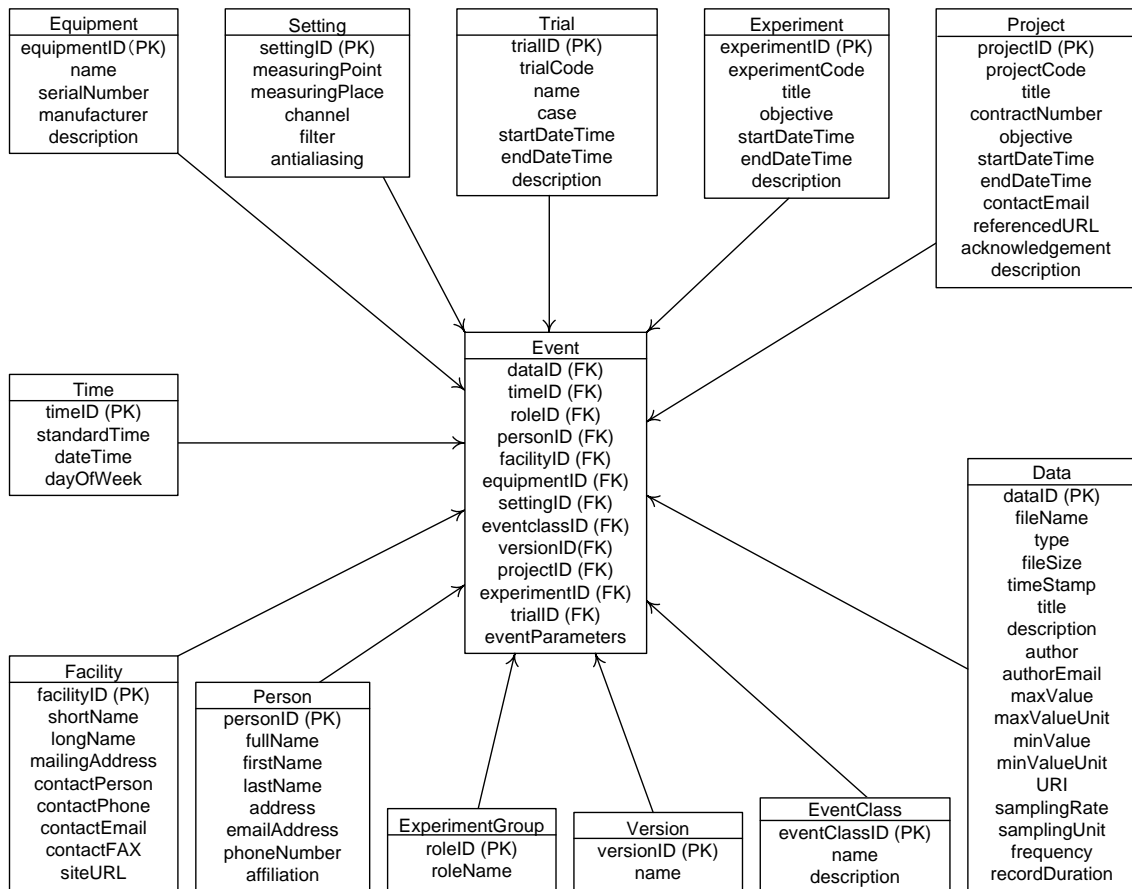

Figure 2: EDgrid Data Model

APPLICATION OF THE EDGRID DATA MODEL

Sensor Data

The EDgrid data model was implemented by using MySQL and the experiment data obtained in the demonstration experiment of E-Defense was applied for verification of the data model and the prototype system.

Sensor data are managed as the following. 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. Each ID of these three tables is linked each other in Event table so that equipment, equipment setting and the trial used can be linked and represented correctly. In this way, the same setting data for repetitive experiments does not have to be input repeatedly, which reduces the amount of work and data.

The database itself does not contain all sensor data but only their metadata. Figure 3 shows a part of the accelerometer data obtained in the demonstration experiment. The sensor data like this 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.

Figure 3: A Part of an Accelerometer Data File (Date, Time, Acceleration)

Tracking Data Change

Data change quite often. For example, researchers move from organization to organization, and equipment tends to be replaced every several years. If the previous data is deleted from the database, the user cannot retrieve such data forever. Further, the user needs to retrieve the data that was valid at the time of the experiment done in the past. Thus, the data model must be able to keep tracking record of changes and show the proper data if requested.

Figure 4 shows an example of a researcher's telephone number change. In Person table, the Person ID of 1 changed the telephone number from 111-111-1111 to 222-222-2222.

This change was done on Wednesday, January 18, 2006 as shown in Time table. This change is in EventClass table and the EventClassID is 12. This change can be represented in 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.
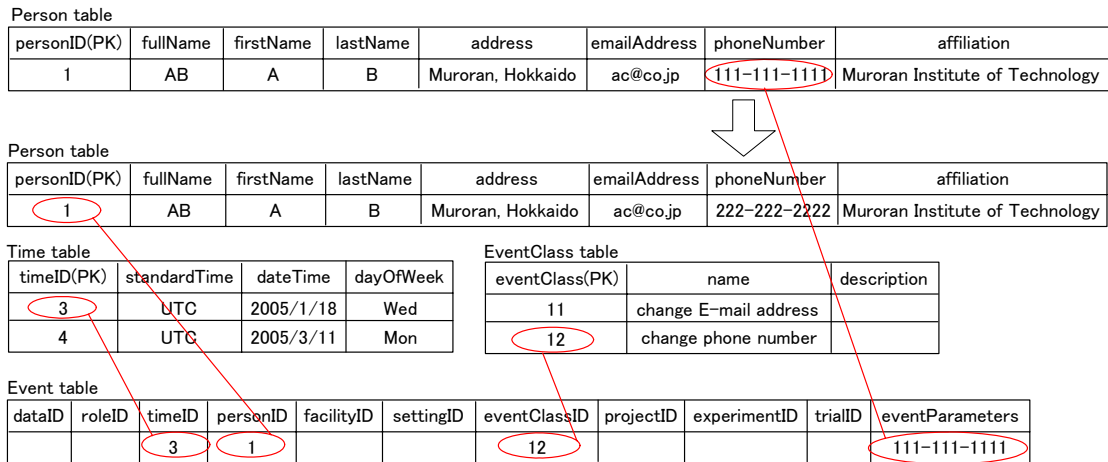

Person table

personID(PK)	fullName	firstName	lastName	address	emailAddress	phoneNumber	affiliation
1	AB	A	B	Muroran, Hokkaido	ac@co.jp	111-111-1111	Muroran Institute of Technology

Person table

personID(PK)	fullName	firstName	lastName	address	emailAddress	phoneNumber	affiliation
1	AB	A	B	Muroran, Hokkaido	ac@co.jp	222-222-2222	Muroran Institute of Technology

Time table

timeID(PK)	standardTime	dateTime	dayOfWeek
3	UTC	2005/1/18	Wed
4	UTC	2005/3/11	Mon

EventClass table

eventClass(PK)	name	description
11	change E-mail address	
12	change phone number	

Event table

dataID	roleID	timeID	personID	facilityID	settingID	eventClassID	projectID	experimentID	trialID	eventParameters
		3	1			12				111-111-1111

Figure 4: An Example of Data Change Management in the EDgrid Data Model

Extending the Data Model

In the future, new tables or new attributes may have to be added to the data model in order to extend the capability of the database. A new table can be added to Event table directly as shown in Figure 5. The primary key of the added table, addedTableID, is inserted to Event table as a foreign key. The event of adding a new table is recorded as an event. New attributes of an existing table can be added to the table as an added attribute table as shown in Figure 6. In this example, Trial table has new attributes and the primary key, trialID, is added to the added attribute table as a foreign key. The event of adding new attributes is recorded as an event.

In this way, extending the data model can be relatively easily done without changing the star-shape two-story hierarchical structure of the data model owing to the flexibility of the event-based data model.

CONCLUSIONS

In this research, EDgrid data model for storing all the data obtained during the experiments executed at E-Defense was developed. As the EDgrid data and NEESgrid data are shared among Japanese and US earthquake engineering researchers, EDgrid and NEESit researchers have been collaborating to develop data models for storing heterogeneous experiment data. The authors reviewed the NEESgrid data model and identified the necessity of simplification and flexibility in order to apply to EDgrid partially due to the rapid development. In the collaborative research with NEESit, the

event-based data model was proposed and the authors adopted it for developing the EDgrid data model.

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.
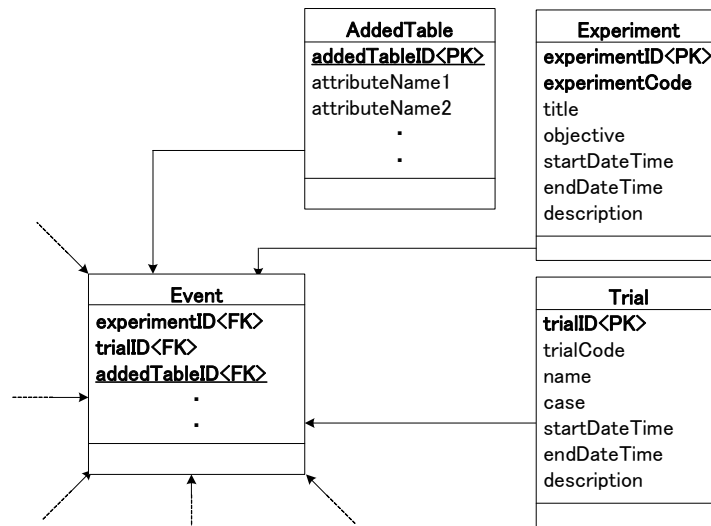

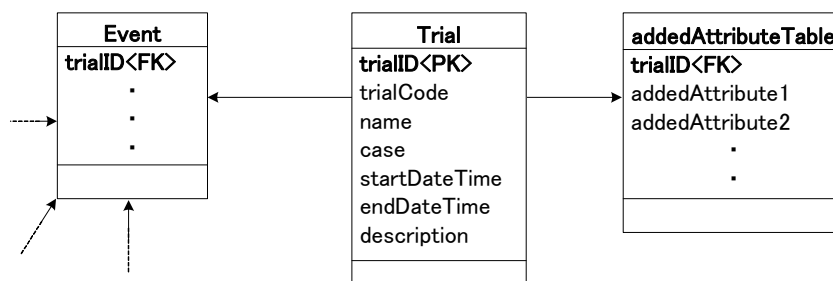Figure 5: An Example of New Table Addition in the EDgrid Data Model

Figure 6: An Example of New Attributes Addition in the EDgrid Data Model

ACKNOWLEDGMENTS

REFERENCES

1. Peng, J. and Law, K. H.: A Brief Review of Data Models for NEESgrid, *NEESgrid TR-2004-01*, 2004.

2. Law, K. H.: Summary Report on NEESgrid's Data Curation Summit, *NEESgrid TR-2004-43*, 2004.

3. Peng, J. And Law, K. H.: Validity and Usability of the NEESgrid Reference Data Model, *NEESgrid TR-2004-44*, 2004.

4. NEES: http://www.nees.org/index.php

5. E-Defense: http://www.bosai.go.jp/hyogo/index.html

6. Mita, A.: Edgrid as Cyber Infrastructure for E-Defense, *Proceedings of the First International Symposium on Innovation and Sustainability of Structures in Civil Engineering*, Nanjing, China, pp.255-260, 2005.

7. Protege2000: http://protege.stanford.edu/

8. Warnock, T.: NEES Data Architecture, *NEESit TR-2005-49*, 2005.