

Establishing the Worthiness of Information Systems Data through Information Engineering

Eghosa Ugboma
eugboma@fmuniv.edu/eugboma@aol.com
Department of Computer Science and Mathematics
Florida Memorial University
Miami Gardens, Florida 33054, U.S.A.

Abstract

The data table planning sheet is the third tool used at the information-level (conceptual) phase of the database design aspect of an information system. Although rarely used in comparison to the frequencies with which normalization process and entity-relationship model are used to illustrate the information-level phase of a data structure, the planning sheet describes each data field's requirement, including the field's rules for storing reliable data that defines the structure of the data table. The tool uses the declarative data restriction approach to mark out the required rules for each field of a user table. These rules are the underlying factor for establishing the reliability and the dependability of data within the data table of an information system. In addition, the instrument follows the data modeling principles of the second level of the planning phase of information engineering and it is used to model data to meet the reliability requirement stipulated for the data table used by an information system. This paper shows how the planning sheet is used to optimize the data table an information system uses. In other words, the paper demonstrates how the tool helps achieve the quality of data for an information system, as well as, maintaining the worthiness of data for the system. Using this instrument (a) documents the characteristics of a data table, (b) ensures that the restrictions the table's data must obey are enforced, and (c) make certain that the user requirement is satisfactorily met.

Keywords: Rules, Restrictions, Planning Sheet, Information System, Information Engineering, Data table, Database.

1. INTRODUCTION

This paper uses a design tool called the planning sheet to optimize a data table of an information system to maintain the accuracy and the usefulness of the data generated by the information system. Since the major concern of this article is to establish and ensure the worthiness of data in a user table, the relational bind that exists among data tables is not discussed and remains outside the scope of this presentation.

The subject of this article is to assure that the rules governing the reliability of data in the user table of an information system is enforced in order to retain the worthiness and the usefulness of data used by an information system. The planning sheet

defines the data types and restrictions that are applied to each data field in the table an information system uses. The tool satisfies the data modeling requirement as described in the second level of the planning phase of information engineering.

The planning phase of information engineering, among other things, model data that is dependable and the planning sheet contains the necessary details needed for data in a user table to be accurate and complete, thus, keeping the table in a reliable and complete state.

Note: To avoid misunderstanding of terms, data table, user table, and table are synonymous and they are used

interchangeably throughout the paper. The same applies to data table planning sheet, planning sheet, sheet, instrument, and tool. Also the terms paper, presentation, and article are synonymous. In this paper data table is used in the singular form to present individual table structure that supports an information system.

2. INFORMATION ENGINEERING

Hoffer, Prescott, and McFadden (2002) define information engineering as "data-oriented methodology to create and maintain information system." They emphasize that within the context of information engineering, data are modeled to meet the needs of the enterprise, not by data usage. In his article, *From Data Engineering to Information Engineering*, Wiederhold (2004) writes that since information engineering focuses on "information", dealing with the planning of information requires new engineering concepts if the data used to generate information is to retain its worthiness.

The fundamental function of information engineering is finding new ways of creating an efficient information system, including its components such as the database that stores the information system's data, for improved performance. The fundamental role of an information system is to generate reliable and dependable information.

The second level of the planning phase of information engineering that conforms to the planning sheet focuses on planning, data modeling, and establishing data worthiness for a data table during the database design phase of an information system. The planning phase of information engineering governs database rules and the authenticity of a table's data.

3. DATA TABLE PLANNING SHEET

The data table planning sheet is a database design instrument that is used to establish a table's structure and content type, and define the data restriction for that table at the information-level phase of a database design. In other words, the planning sheet is used to define and retain the necessary restrictions for a data table of an information system. The sheet addresses the issues of defining and ensuring the necessary data restrictions for each data field of a user table

in a database that will support an information system, both at the information-level design phase and the physical-level design phase to maintain accuracy, consistency, and the completeness of data.

These restrictions administrate the validity of data in the table and ensure correctness of the database in terms of design, implementation, and manipulation. The tool, although rarely used compared to the high frequencies with which normalization process and entity-relationship model are used to illustrate a table structural configuration, is also used to develop the logical representation of data in the database of an information system and to model data at the information-level of a database design to meet information system's user requirement.

The planning sheet does not only establish the worthiness of data in a user table, it defines as well a table's data restrictions. The sheet uses the declarative data restriction approach to define the restrictions. The approach is a method where rules are the primary means of establishing and enforcing the defined restrictions. Restrictions are affirmed, as required, to each data field of a data table.

Using the planning sheet, an information system designer plans a table structure that includes the types of data restrictions that must be applied and retained by each data field in the table. In other words, the planning sheet establishes the requirement for each field of a user table of an information system. The tool provides an effective method for defining the rules that will guide the physical-level phase of a database design that will support a information system to ensure the authenticity of data in the user table. . The sheet is both a descriptive and a problem-solving tool that become part of an information system documentation.

4. OPERATION METHODS

Most information systems and database textbooks hardly present the establishment of data worthiness at the information-level of an information system design instead they discuss the implementation of the usefulness of data at the physical-level phase. The lack of not seeing the overall picture of the functionality of a data table prior to the

physical-level phase, most of time, leads to a unreliable system design.

The function of a user table that is part of an information system must fully be analyzed if the table is to generate the information needed by the user. This analysis, which must include the definition of restrictions the data must obey to maintain its worthiness, is performed at the information-level design phase, and this is where the planning sheet comes into play.

The information-level phase, which can be regarded as the housekeeping or preparatory phase, is the first level of the two levels of a database design. This level is the design phase on which the implementation of the physical-level phase is based. The physical-level phase of a database design is the level where the actual database and its data tables are created. The planning sheet is used to ensure the authenticity of data within a user table prior to its creation in a database that will support an information system.

Data Table Planning Sheet versus Normalization and Entity-Relationship Model

In a database design, normalization process is used to identify and remove potential problems as well as to set up controlled redundancy. In addition, normalization process shows table names, field names, and primary key fields of the tables. Normalization process does not indicate the fields' data restrictions nor does it show the fields' width sizes.

On the other hand, entity-relationship (E-R) model main functions during a database design process are to indicate tables' relationships and cardinalities. Chen and Crow's Foot E-R models show table names, but optionally indicate tables' field names, primary keys, and foreign keys. The Air Force IDEF1X E-R model shows all the information that is sometimes missing from Chen and Crow's Foot models. All database entity-relationship diagrams used to model tables' relationships and cardinalities do not indicate data restrictions for tables nor do they define tables' field sizes.

The planning sheet is used to plan and design the structure of a table, the data for the table and the restrictions each field of the data table must obey.

There are two main sections in a data table planning sheet. The top section is used to specify the database name in which the table belongs, the table name, and the operational function of the table. The bottom section of the sheet is used to define the specification for each field in the table which include the description of the field, its actual name, its width size where applicable, and the restrictions its data must abide by. The number of rows required in the *Attribute Specification* section of the sheet, beside the first row that contains the field titles, is determined by the number of fields that define the structure of the user table.

The instrument shows how the different rules are defined to enforce data restrictions at the information-level phase during an information system development. The planning sheet is used to enforce the following restrictions, namely: key restrictions (primary, foreign), unique restriction, not null restriction, default restriction, data type restriction, domain restriction, and format restriction. Each column of the *Attribute Specification* section of the instrument is briefly explained for easy comprehension.

Since the *Attribute Specification* section of the planning sheet is the part used to define the data restrictions that a table's data must obey, this article explains each restriction-related column of the section. In the appendix section of this paper, a sample of a completed planning sheet using a fictitious database name and a table name with made-up field requirement that illustrates how data restrictions are established is presented.

Data Restrictions Columns of the Data table Planning Sheet

The *Data Type* column of the planning sheet defines the data type restriction that dictates the type of data, such as date each field of a table must accept and store. Data type restriction is mandatory for each field in a data table.

The *Domain* column defines the acceptable value and manages the legality of the value that must enter a field even if the value satisfies the field's data type. The *Domain* column establishes the restriction that allows only an acceptable value to be entered into the specified field.

The *Format* column defines the restriction that specifies the type of input or output format for a field.

The *Primary Key* column defines the restriction that ensures that a value in a primary key field must not be replicated, except where the field is a part of a composite primary key, and the field must not be null. This type of restriction is used to discriminate one record from another.

The *Foreign Key* column of the planning sheet defines the restriction that a field that serves as the foreign key in a user table must either contain data or be null. If the foreign key field contains data, the data must match some values in the primary key field of the data table the foreign key is referencing.

The *Unique* column defines the restriction that a field must either contain values that are not replicated or store null values. The *Unique* column establishes which field obeys the restriction in a data table.

The *Not Null* column of the planning sheet defines which field must always receive data. The restriction defined by the Not Null column, although it ensures that a value must be entered into a field, permits the attribute to accept replicated values.

The *Default* column defines the restriction that allows an initial value or a default value to be established for a field of a data table. The default value can be initiated either during the table's creation or its alteration.

Note: Both the *Primary key* and the *Foreign key* columns of the planning sheet are typically used to create a controlled redundancy in normalization process. The planning sheet shows the field that serves as the primary key of a user table. The instrument also shows the field that serves as the foreign key of a table if the table is to establish an association or a relationship with other tables in a database. When a data restriction is not defined for a field, the restriction column for that field is left blank. In such a situation, the field accepts data based on the default setting of its data type.

Non Restrictions Columns of the Data Table Planning Sheet

Other columns of the *Attribute Specification* section of the planning sheet that are not used to establish data restrictions in a data

table used by an information system are described below.

The *Description* column is used to describe the function of each field in a user table. The *Name* column is used to provide proper (actual) name for each field. The name assigned to a table's field must be unique. The name, during the physical-level phase implementation, is used to identify the field, except in situations where the name is altered, throughout the life span of the table in the database.

The *Size* column defines, where appropriate, the number of character positions to be assigned or allocated to a field of a user table. Some data type such as the date data type automatically defines a field size to a fixed width size. In such instances, the field's *Size* column assignment is omitted. The *Decimal Position* column of the *Attribute Specification* section is used to note the number of decimal positions, even if it is a zero decimal position, for a field with a numeric data type. All other data types do not require this column.

The *Index* column specifies the field that enhances the efficiency with which data is retrieved from a data table. Although, unique indexes are created automatically for primary key and unique fields, the *Index* column of the planning sheet denotes another field that can be used to search and retrieve data efficiently from a user table.

5. CONCLUSIONS

The data table planning sheet, although rarely used, is one of the designing tools used at the information-level of a database design aspect of an information system development. This paper demonstrates the optimization of an information system's data table using the planning sheet. The tool adopts the data modeling approach of the second level of the planning phase of information engineering and defines the restriction requirement for each field of a user table in a database that supports an information system.

The planning sheet is both a descriptive and a problem-solving instrument that assists in reducing the chance of creating a poor, inconsistent, and unreliable data table for an information system. Using this instrument, an information system's designer properly

and adequately documents the necessary requirement, including the restrictions, for each field of a user table. This documentation becomes one of the guides on which the physical-level phase of an information system's database design is implemented.

For the fact that normalization process, entity-relationship model, and planning sheet serve unique and different functions, it is recommended that the three tools be employed during a database design aspect of an information system development. Making the planning sheet a solid and visible instrument in designing a database will not only increase the understanding of how restrictions are defined for a user table, but will also enhance the way the database of an information system is planned and developed.

6. REFERENCES

- Connolly, M. Thomas and Carolyn E. Begg, (2002) Database Systems: A Practical Approach to Design, Implementation, and Management (3rd ed.). Addison-Wesley, New York.
- Elmasri, Ramez and Shamkant B. Navathe (2002) Fundamental of Database Systems (3rd ed.). Addison-Wesley, New York.
- Frick, R. David & Co (2004) Data Integrity (http://www.frickcpa.com/ss7/Theory_DataIntegrity.asp).
- Kroenke, M. David (2003) Database Concepts (2nd ed.). Prentice Hall, New Jersey.
- Modell, Martin (2002) The Various Types of Information Systems Analysis Projects. <http://www.dai-sho.com/pgsa2/pgsa02.html>.
- Pratt, J. Philip and Joseph J. Adamski (2002) Concepts of Database Management (4th ed.). Course Technology, Cambridge.
- Rob Peter and Carlos Coronel (2002) Database Systems - Design, Implementation, and Management (4th ed.). Course Technology, Cambridge.
- Swanson, Marianne and Barbara Guttman (1996) Generally Accepted Principles and Practices for Securing Information

Systems.<http://csrc.nist.gov/publications/nistpubs/800-14/800-14.pdf>.

The University of Texas at Austin (2004) <http://www.utexas.edu/its/windows/database/datamodeling/dm/design.html>.

Wiederhold, Gio (2004) From Data Engineering to Information Engineering <http://www-db.stanford.edu/pub/gio/1994/inf-eng-abstract.html>.

Wu, Jonathan (2004) Ensuring Data Integrity through the Use of Prevent and Detect Controls, Part I <http://www.evaltech.com/wpapers/ensuringdataintegrity.htm>.

7. APPENDIX

For the purpose of this presentation, the following notations are used to explain the *Attribute Specification* section of the sample of the completed planning sheet shown on the following page.

In the *Size* column, the algebraic less than or equal to symbol "<=" is used in combination with a number to assign a width size to a field if the field will contain variable-length data. The symbol indicates that a field can store data with character length of up to the specified character positions. Otherwise the specified character positions denote a fixed-length size for the field.

In the *Domain* column, a hyphen (-) symbol between two value items indicates a data range. For example, the domain specification of 0-9 designates a data range of digit 0 through digit 9. A semi colon is used to separate one data range from another or to separate one item list from another.

In the *Format* column, the character "X" is used to denote non-digit character positions for data to be stored in a field. The character "9" indicates digit character positions for data to be stored in a field, even if the field is defined to store character data type. Such a notation symbol clearly stipulates the positions where digits are allowed in a character data if the data contains digits. A letter-case sub format is used to include a case level (uppercase, lowercase, or mixed case). For example, the format declaration

for the Semester_Code field is "XXXXXXXX9999 (mixed case)".

Although the Semester_Code field is of the character data type, the format allows data such as the phrase "Winter 2004" to be entered into that field where "W" of the word "Winter" is in uppercase and the rest in lowercases. Also the year "2004" is entered due to the fact that the last four character positions of the format are "9999". This indicates that digits are only permitted in those positions.

The expression "Yes" is used to denote primary key, foreign key, unique, not null, and index fields for a user table. A check "√" or the symbol "x" can also be used in place of the expression to denote the fields.

Table 1. Below is a sample of a planning sheet.

Data Table Planning Sheet												
Database Name: _____												
Table Name: _____												
Purpose Of The Table: _____												
Attribute Specification												
Description	Name	Data Type	Size	Decimal Position	Domain	Format	Primary Key	Foreign Key	Unique	Not Null	Default	Index

Table 2. Below is a sample of a completed planning sheet.

Data Table Planning Sheet												
Database Name: <u>Program Dbase</u>												
Table Name: <u>Course</u>												
Purpose Of The Table: <u>Stores courses' information in order to keep track of faculty members assigned to teach the courses and students registered for the courses in the Specified semester</u>												
Attribute Specification												
Description	Name	Data Type	Size	Decimal Position	Domain	Format	Primary Key	Foreign Key	Unique	Not Null	Default	Index
Courses' identification codes	Course_Code	Character	7		0-9; A-Z	XXXX999 (upper case)	Yes					
Course names	Course_Name	Character	<=25		0-9; A-Z; a-z; Space	XXXXXXXXXXXXXXXXXXXXXXXXXXXX Or XXXXXXXXXXXXXXXXXXXXXXXXXXXX99 (Mixed case)				Yes		
Courses' section numbers	Section_Num	Numeric	2	0	0-9	99	Yes					
Credit hours assigned to each course	Credit_Hour	Numeric	1	0	1-5	9				Yes	0	Sat, Nov 4, 8:00 - 8:25, Normandy A
Faculty identification numbers	Faculty_Id	Character	9		0-9	999999999		Yes				
Student identification numbers	Student_Id	Character	9		0-9	999999999		Yes				
Semester identification codes	Semester_Code	Character	<=12		0-9; A-Z; a-z; Space	XXXXXXXXXX9999 (Mixed case)				Yes		Yes
Date semester begins	Start_Date	Date				MM/DD/YY						
Date semester ends	End_Date	Date				DD-Mon-YY						