

Comparison of Database Models for Developing a Pavement Performance Analysis System

Choi, Jae-ho*

Abstract

One of the most difficult tasks in pavement management information systems is establishing the links between performance measures of a structure and the design and construction inputs. In-situ pavement performance can be considered a response variable to many project input variables, such as design, construction, and traffic loading effects. If we are to fully understand the component of pavement performance and specify the inputs through design and construction specifications to achieve that performance we must develop quantitative relationship between input and response variables through a scientific, fully integrated Pavement Performance Analysis System (PPAS). Hence, the objective of this study is to design a database model required for developing an effective database template that will allow analysis of pavement performance measures based on design and construction information linked by location. In order to select the most appropriate database model, a conceptual database model (Entity Relationship Model) and dimensional model, which is believed to be the most effective modeling technique for data warehouse project, are designed and compared. It is believed that other state highway agencies could adopt the proposed design strategy for implementing a PPAS at the discretion of the state highway agencies.

Keywords : Pavement Performance, Pavement Performance Analysis System (PPAS), Entity Relationship Model, Dimensional Model

1. Introduction

One of the most difficult tasks in pavement information systems is establishing the links between performance measures of a structure and the design and construction inputs. During the last two decades, Wisconsin Department of Transportation has made significant progress in improving material properties, design practices, and monitoring of pavement performance. The databases that are available at the DOT include detailed material properties, detailed pavement structure information, and detailed pavement indicators and performance measures.

One of the most significant achievements of State Highway Research Program (SHRP) was developing new testing systems that relate mechanistic properties of materials and mixture design methods to in-situ pavement performance. Mechanistic properties are defined through new methods and standardized system for incorporating material, environmental, and other engineering

properties. The results of SHRP could be utilized effectively if a Pavement Performance Analysis System (PPAS) is developed.

In-situ pavement performance can be considered as the response variable given project inputs, such as design, construction, and environmental and traffic loading. The existing processes used by WisDOT and other highway agencies to determine these relationships have been largely based on a collection of experiences and knowledge acquired through years of pavement performance monitoring and continuous specification development. In order to fully understand the component of pavement performance and specify the inputs through specifications to achieve that performance we must develop quantitative relationship between input variables and response variables through the scientific, fully integrated PPAS.

Portions of the system are already in-place, but research is necessary to develop the database model for a PPAS which will allow analysis of pavement performance measures based on design and construction information linked by location. This study divides the research attempts regarding developing a PPAS into two major categories: the conceptual and the logical field. In the conceptual

* 일반회원, 한국건설관리공사, 연구개발처 CM팀, 과장, 공학박사

field, the major issue is to design conceptual modeling formalisms (e.g., Entity Relationship Model) using the network overlay techniques. In the logical field, the research focuses on the development of a logical, implementation independence model (e.g., Dimensional Model) to describe multidimensional data structures, in which the data can be seen as depending on several independent variables. Finally, the comparison between these two models is illustrated to recommend a more appropriate database model for implementing PPAS database template.

2. Overview Model of a PPAS

An overview of the existing databases and a vision for how to link them to achieve the objectives of this study were established after discussions with the DOT engineers. Figure 1 is a schematic representation of how different sources of data located in different systems can be linked to develop a fully integrated PPAS. The upper part of Figure 1 indicates the application systems used by WisDOT. The source files include pavement structure data, design JMF (Job Mixture Formula) data, material QC/QA testing data, performance measures, traffic, and aggregate source testing data. The pavement structure data, JMF, and material QC/QA data are grouped because these sources are subjected to the modification based on the feedback from PPAS.

Nyerges (1990) explains that most transportation information can be referenced using a combination of three schemes: road name and milepoint, control section, and chain (link) and node. Based on this classification, Layer and Base (L&B) database and Pavement Information Files (PIFs) use a combination of road name and reference point (pseudo milepoint) scheme, and control section scheme, respectively.

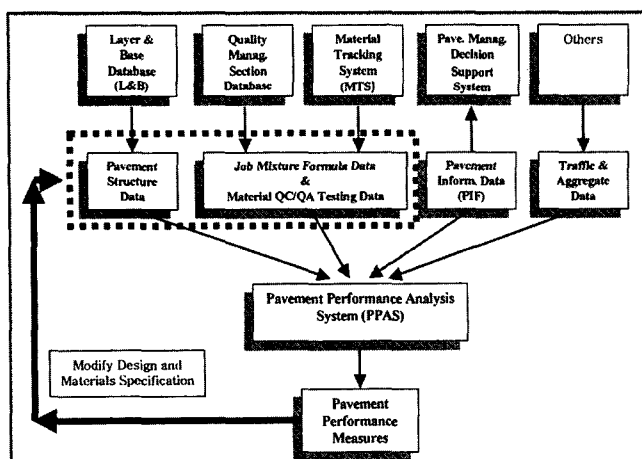


Figure 1. Overview of the WisDOT's PPAS.

Table 1 shows the summary of location reference schemes of databases, which is vital information in integrating the extracted datasets through location referencing. FROM_RP and TO_RP are expressed in reference points whereas FROM_PLUS and TO_PLUS are expressed in miles.

Table 1. Summary of Location Reference Schemes of Each Database

Database	Location Reference Scheme
L&B	FROM_RP and TO_RP
PIF	FROM_RP, FRPM_PLUS, TO_RP, and TO_PLUS
QC/QA	Project ID, Quantity of material placed
MTS	Project ID
AGGREGATE	Project ID
Traffic	FROM_RP and TO_RP

If the location of FROM_RP of a project is known in miles through the conversion process from reference point to milepoint, summing up the location of FROM_RP and FROM_PLUS will identify the location of the starting point of a project.

3. Designing a Conceptual Database Model

This section presents a conceptual database design for PPAS that allows extracted datasets from the existing databases to be related in order to facilitate the pavement performance modeling with respect to material and construction variables. The design of the extracted database should be driven by a clear statement of the ways the database is to be used in the proposed modeling framework. Without a clear statement of the intended use it is not possible to evaluate the usefulness and correctness of the proposed database design. Therefore, five important objectives of the PPAS were established based on the literature search and discussions with the DOT engineers. The important queries for PPAS can be surrogates for illustrating functional requirements of the PPAS. They are as follows:

- 1) Determine the effects of specific design features on pavement performance
- 2) Determine the effects of specific construction features on pavement performance
- 3) Determine the relationship between material properties and distress types
- 4) Determine the effectiveness of pavement maintenance treatment
- 5) Develop mechanisms to re-calibrate coefficients of the PPAS

model based on comparisons between predicted and actual measured values

In addition to these selected objectives, the relationship between as-constructed material and construction input variables, and pavement performance is one of the main challenges involved in estimating future pavement performance. To establish sampling and testing components of variance of quality characteristics for specification improvement, PPAS should be able to correlate quality characteristics with distress over time. Such a system is difficult to achieve using the existing WisDOT information systems since the current QC/QA split sample location is based on the tonnage at the point of sampling (quantity) not on highway location references (Quality Management Program Guide 1999). This requires a plan for including location references in the current quality management section database.

Figure 2 shows an example illustrating how the proposed PPAS can facilitate the analysis for identifying the relationship between as-constructed QC/QA data and field performance. If the design target for asphalt content is 5.0%, and there are significant deviations from the target in several pavement sections where raveling occurred, the regression or Analysis of Variance (ANOVA) models would detect some relationship. It may be learned that certain deviations from target are perfectly acceptable, while other deviations are critical to performance. Since the traffic levels, temperature, and base condition are consistent, these variables could be blocked from the analysis. Then those variables that changed within the project could be evaluated (Schmit 1998).

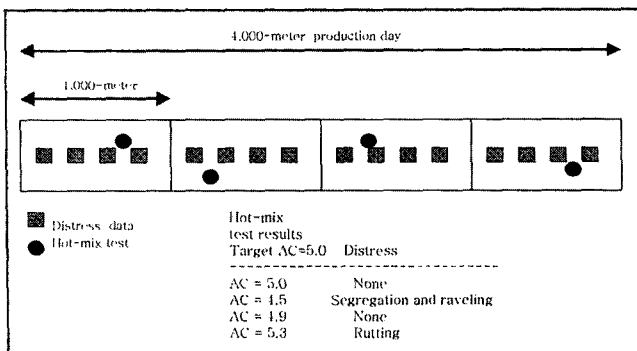


Figure 2. Comparison of as-constructed QC/QA and field performance (Schmit 1998).

3.1 Developing a Strategy for relating the Component Datasets

Figure 3 shows a high-level schematic of a strategy for linking the separate databases. "A", "B", and "C" in figure 3 indicates

components currently lacking in the existing pavement information systems used by WisDOT. "A" indicates that material properties cannot be related to performance measures. "B" indicates that QC/QA data cannot be related to distress data. "C" indicates that field distress data are not available based on sampling location. This is because current method for sampling is mostly based on the quantity of material placed.

In pavement information systems used by WisDOT, the location schemes of L&B, PIF, and traffic databases are linear-oriented (the metric for reference is along the length of a highway) (Nyerges 1990), whereas the MTS database does not use any location referencing schemes. As shown in figure 3, L&B, PIFs, and traffic databases can be related each other using network overlay techniques. The L&B database plays a central role to link material data (Job Mixture Formula) to performance measures because this database includes a common attribute (Project Identification Number) based on which material data and performance measures can be related.

3.2 Network Overlay Process to relate Material Data, Traffic Data, and Performance Measures

In general, different application groups in an organization have difference control sections, one for pavement construction management (e.g., L&B database) and another for highway performance monitoring (e.g., PIFs). Each section is homogeneous with respect to the attribute of interest to that application group. Therefore, when data is to be integrated from different records (e.g., L&B and PIFs) then network overlay techniques using ordered milepoints can derive the common attribute values needed as was indicated by Nyerges (1990). Since L&B database and PIFs use a reference point method to reference the control section, conversion process from reference point to milepoint (mileages) is necessary to be used in the network overlay process.

Figure 4 illustrates the network overlay process used to relate material and construction data in L&B database and performance measures in PIFs given the assumption that reference scheme was converted to milepoints scheme (e.g., A, B, C, and so forth shown in figure 3). Specifically, Co_LAYER entity is the product of network overlay algorithm between "Layer" entity in the L&B database and "Section Description" entity in PIFs. This network overlay process divides the network into smaller sections (e.g., 1, 2, 3, and so forth as shown in figure 4) where the attributes data are homogeneous. For

example, section 1 is associated with PROJ_ID 1234 and PIFs 3722. This section can be associated with material and construction data through PROJ_ID 1234 and with performance measure through sequence number (SEQ) 3722.

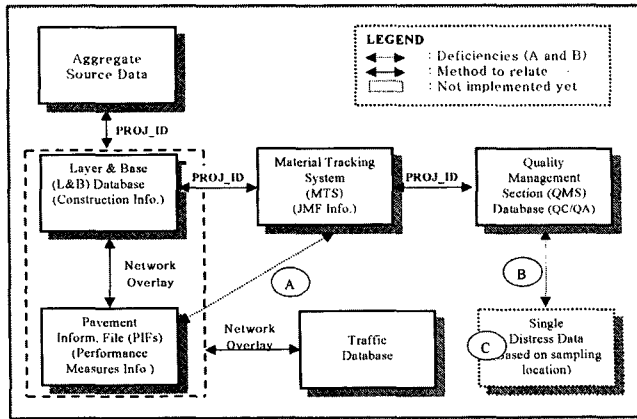


Figure 3. Schematic for relating components datasets.

In order to relate traffic data to a corresponding record in Co_LAYER entity, traffic database needs to be overlaid on the Co_LAYER entity using network overlay technique. It requires converting reference points used in traffic database to milepoints before network overlay is conducted since traffic database uses reference point method (FROM_RP and TO_RP) to indicate the range of a traffic control section. This overlay process will add a new attribute (SITE_ID) which is the foreign key to traffic volume entity to a corresponding record in Co_LAYER entity. This will produce a new entity named Co_LAYER_II whose attributes, therefore, include PROJ_ID, SEQ, SITE_ID, FROM_RP, and TO_RP.

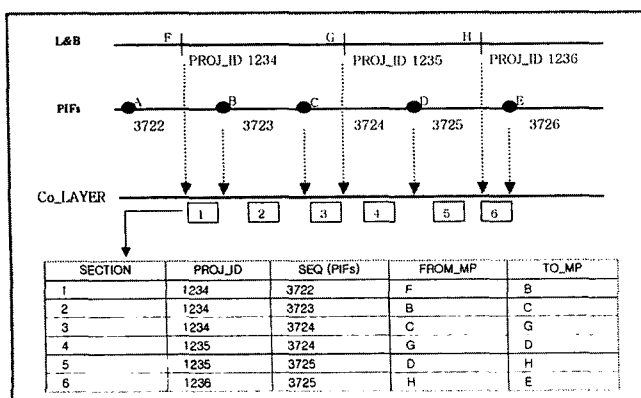


Figure 4. An example of a network overlay between L&B database and PIFs.

3.3 Entity-Relationship (E-R) Model for PPAS

Based on the E-R representation of each database and files (Choi

2002), a conceptual E-R model for developing the PPAS was developed as shown in Appendix I. The entity "Co_Layer_II" is the product resulting from the network overlay processes of "Layer (Layer shown in Appendix I)" entity in the L&B database and "Section Description (SECTION shown in Appendix I)" entity in PIFs and "SITE" entity in traffic database. There are several entities that can be combined into one entity in Appendix A. For example, SECTION, LAYER and Co_LAYER_II entities can be combined into one big denormalized table since minimizing Structural Query Language (SQL) joins is a good way to improve performance execution time. As was indicated before, material data including design JMF, QC/QA, and aggregate data can be associated with the pavement structural data in the L&B database through a project identification number (PROJ_ID). Appendix I shows that they could have the same entity table named "TEST_NO" indicating that material data entities such as QC/QA, JMF, AGGREGATE entities can be related to TEST_NO entity through the attribute TEST_NO.

4. Dimensional Models for Data Warehouse Design

It is required to select an appropriate database model before implementing database template for PPAS. In this section, the dimensional model, which is an alternative to conventional E-R model, was developed since the dimensional model has a number of important data warehouse advantages that the E-R model lacks. Dimensional modeling (DM) is the name of a logical design technique often used for data warehouse that seeks to present the data in a standard framework that is intuitive and allows for high-performance access (Kimball 1998). A data warehousing is an architecture for integrating data resources to be used for reporting and decision support initiatives (Matthias 2000). The data within the data warehouse is accumulated in multidimensional data structures to support direct querying and multidimensional analysis in which the output data can be seen as depending on several independent variables.

Every dimensional model is composed of one table with a multipart key, called the "fact" table, and a set of smaller tables called "dimension" tables. A fact table is the primary table in each dimensional model that is meant to contain measurements of the business. Each dimension table has a single-part primary key that corresponds exactly to one of the components of the multipart key in the fact table. A fact table is surrounded by a halo of dimension tables, where the dimension keys are primary keys in each respective

dimension table.

In this study, fact tables represent a many-to-many relationship with respect to several dimension tables and are only used for modeling the relationship between performance measures and input material and construction variables. In other words, most dimension tables contain many textural attributes (fields) that are the basis for constraining and grouping within data warehouse queries. End-users can always create a row header in a report by dragging a dimension attribute from any of the dimension tables down into the spreadsheet, thereby making a row header.

4.1 Data Warehouse Bus Architecture for PPAS

On the basis of defining the important objectives of the PPAS, an overview of the pavement performance analysis system, called the data warehouse bus architecture, was deigned as shown in Table 2. The data warehouse bus architecture is a matrix that shows dimension tables in columns and fact tables (data marts) as rows. The matrix shows which dimension tables are used in which fact tables. Song et al. (Song 1999) explain that the data warehouse bus architecture was proposed to standardize the development of data marts into an enterprise-wide data warehouse, rather than to cause stovepipe data marts that were unable to be integrated into a whole.

Table 2. Data Warehouse Bus Architecture for PPAS

Data Mart \ Dimension	Layer	Base	Sub base
Material Testing	x	x	x
QC/QA	x	x	x
Aggregate Consensus	x	x	x
Selection Description	x		
PSI	x		
Rigid PDI	x		
Flexible PDI	x		
Layer Detail	x		
Base Detail		x	
Subbase Detail			x
Traffic	x		

As shown in Table 2, the three data marts are combined into an integrated data warehouse through the conformed dimension tables. The dimensions used in multiple data marts are called conformed dimensions (Kimball 1998). Material testing, QC/QA, and aggregate consensus dimension tables are the examples of the conformed dimension tables. Conformed dimension tables make it possible that a single dimension table be used against multiple fact tables in the same database space (Kimball 1998).

4.2 Fact Table Design

In order to develop a fact table based on the data warehouse bus architecture for PPAS, the network overlay algorithm should be implemented on the attribute data on L&B database, PIFs and traffic database to be able to integrate the extracted datasets. This network overlay process divides the network into smaller sections where the attributes data are homogeneous and generates the new entity table (i.e., Layer Fact Table shown in Appendix II (a)), which is required for modeling the relationship between the layer fact table and dimension tables. The attributes (i.e., HWY, FROM_RP, and TO_RP) in layer fact table define one spatially divided pavement segment. The attributes data in denormalized dimension tables are related to the layer fact table through multipart keys (e.g., LAYER_ID, PROJ_ID, and SEQ).

However, the network overlay process is not used to develop base and subbase fact tables because these two sub-layers are not directly related with the performance measures. Therefore, the SEQ attribute in PIFs is not used in base and subbase fact tables as shown in Appendix II (b) and (c). These two fact tables are only used to provide the sub-structure information of a pavement segment specified in layer fact table and can be obtained by dividing each base and subbase master files in L&B database into two parts: key groups (BASE_ID, PROJ_ID) and attributes (HWY, FROM_RP, and TO_RP). Those two sub-layers, information can be provided to end-users through the conformed dimension tables, which enable each fact table to be related to other fact tables.

A set of attributes, HWY, FROM_RP, and TO_RP in fact tables, are called spatial-enabling data (Song 1999). Using a Geographic Information System (GIS) tool could greatly extend the capabilities of this spatially capable data warehouse by allowing the users to view and analyze data using geo-analytical tools. This requires the addition of spatial data (e.g., map data) to the warehouse environment. However, these spatial data will not be fully integrated with the data warehouse, but rather will be managed by GIS (Mennecke et al. 1999). A tight coupling of a GIS and a data warehouse will permit managerial personnel to acquire access to relevant geographical information stored in the appropriate linked geographical data store.

4.3 Complete Dimensional Models

Appendix II shows the complete dimensional models centered on layer fact table, base fact table, and subbase fact table. As was

illustrated before, every fact table represents a many-to-many relationship and contains a set of two or more foreign keys that join to their respective dimension tables. Each dimension is defined by its primary key that serves as the basis for referential integrity with any given fact table to which it is joined. Most dimension table contains many textual attributes (fields) that are the basis for constraining and grouping within data warehouse queries.

In the complete dimensional models, the material testing dimension can be combined with the QC/QA dimension and aggregate source dimension because they all have the same PROJ_ID attribute as a primary key. In this case, although this new attempted dimension would contain exactly the same information as the proposed one in Appendix II, it is recommended to reject this design immediately because the new larger dimension would be unwieldy and would present performance problems and no user interface advantages (Kimball 1998). This also makes data uploading easier because it does not require any pre-transformation or combining process prior to uploading to data warehouse.

4.4 Adding QC/QA and Traffic Subdimensions to Layer Dimension

If the user wants the specified section FROM_RP to TO_RP in Layer Fact table, the system only needs to provide the QC/QA data within that section. The TO_RP in QC/QA dimension of figure 5 is the location where a split sample is collected and a QC/QA data of that sample is recorded. The list of the QC/QA dimension whose TO_RP is less than the TO_RP in the Layer FACT table is the QC/QA attribute data of the specified pavement section end-user is interested in.

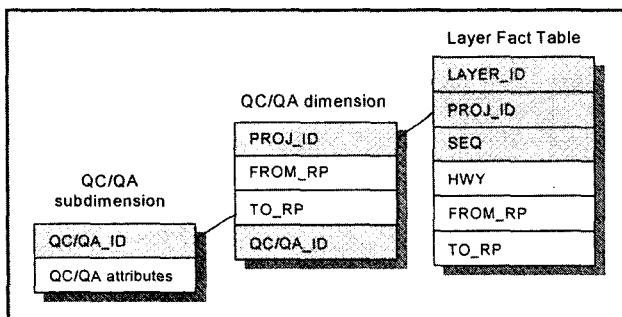


Figure 5. Snow-flaked tables in QC/QA dimension.

The dimensional models require a slight change into snow-flaked model in the QC/QA dimension. The snow-flaked is a data model that is highly normalized with many small entities, taken to fourth or fifth normal form, such that, when laid out visually, the model looks

like a snowflake (Kimball 1998). The reason that snow-flaked QC/QA subdimension table is urged in this study is because of the different level of granularity than the primary QC/QA dimension table.

The possible approach to assign traffic data to the proposed PPAS is to build a subdimension that has the appearance of a snow-flaked table as shown in figure 6. The traffic subdimension is a set of traffic attributes that have been measured for the segments that a PIFs section belongs to. Several PIF sections may share this identical set of traffic attributes because one traffic dimension can include several PIFs sections. It makes sense to isolate this traffic data in a snow-flaked subdimension since the traffic data is available at a significantly different detail and is administered and loaded at different times than the rest of the data in the Section dimension.

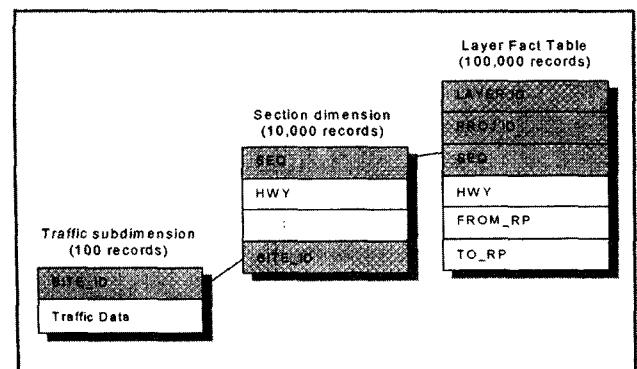


Figure 6. Snow-flaked tables in traffic dimension.

5. Comparison of E-R Model and Dimensional Model

This section is intended to illustrate the basic difference based on the concept of each database modeling technique so that practitioners could choose the best way to model a PPAS. It is known that E-R model is normalized to remove all redundancy and structure the inter-relationship between data and overly complex to the practitioners in the state and local government. Its structure is more efficient in transaction-oriented business.

A dimensional model contains the same information as an E-R model but packages the data in a symmetric format whose design goals are understandability, query performance, and resilience to change. The predictable framework of a dimensional model can make the user interface more understandable, and processing more efficient. The symmetric structure of a dimensional model can accommodate unexpected new data elements without changing

design decision (Kimball 1998). Table 3 shows the comparison between these two techniques.

Table 3. Comparison between E-R Model and Dimensional Model

Criteria	E-R Model	Dimensional Model
Basic Concept	Based on Normalization and Generalization	Based on Star schema (Symmetric Format)
Objective	Transaction-oriented	Efficiency-oriented (Processing more efficient)
Design Goal	Inter-relationship between Data	Resilience to Change, Understandability, Query Performance
Shortfall	Overly complex	Limited Analysis Queries

6. Concluding Remarks

Current WisDOT's pavement management information systems although are of high quality are deficient from an analytical viewpoint because they limit the analysis capability to determine the relationship between performance measures of a structure and the design and construction inputs. Therefore, the formulation and presentation of the strategy for relating the datasets extracted from source databases and files spatially using the network overlay method was presented to develop a conceptual database model (E-R model) required for implementing pavement performance analysis system.

In addition, this study presented the dimensional model, which is an alternative to conventional E-R model, since the dimensional model has a number of important data warehouse advantages that the E-R model lacks. The proposed dimensional models can facilitate the proposed modeling framework by meeting the selected important objectives of the PPAS. To our knowledge, this is the first detailed dimensional data warehouse model specifically targeted for state highway agencies for improving pavement performance analysis. Although the design strategy for modeling a PPAS in this paper is developed based on pavement management information systems and management practices at WisDOT, it could be effectively used by

other state highway agencies which have similar information systems and location reference schemes by adopting the general guideline illustrated in this paper. Next steps will be oriented at developing the software component which interconnects the two technologies: dimensional models for the PPAS data warehousing project and geographic information system.

References

Nyerges, T. L. (1990). Locational Referencing and Highway Segmentation in a Geographic Information System. *ITE Journal*, Vol. 27.

Quality Management Program Guide/Procedure Manual. (1999). Wisconsin Department of Transportation, Div. of Transportation Infrastructure Development, Bureau of Highway Construction.

Schmit, R. L. (1998). *Development of Statistically-Based Methods to Determine QC/QA Testing Levels for Hot-Mix Asphalt Construction*, PhD thesis. Div. Of Const. Engr. And Mange, University of Wisconsin, Madison.

Choi, Jae-ho. (2002). *Framework of Developing Performance-related Specifications at WisDOT*, PhD thesis. Div. Of Const. Engr. And Mange, University of Wisconsin, Madison.

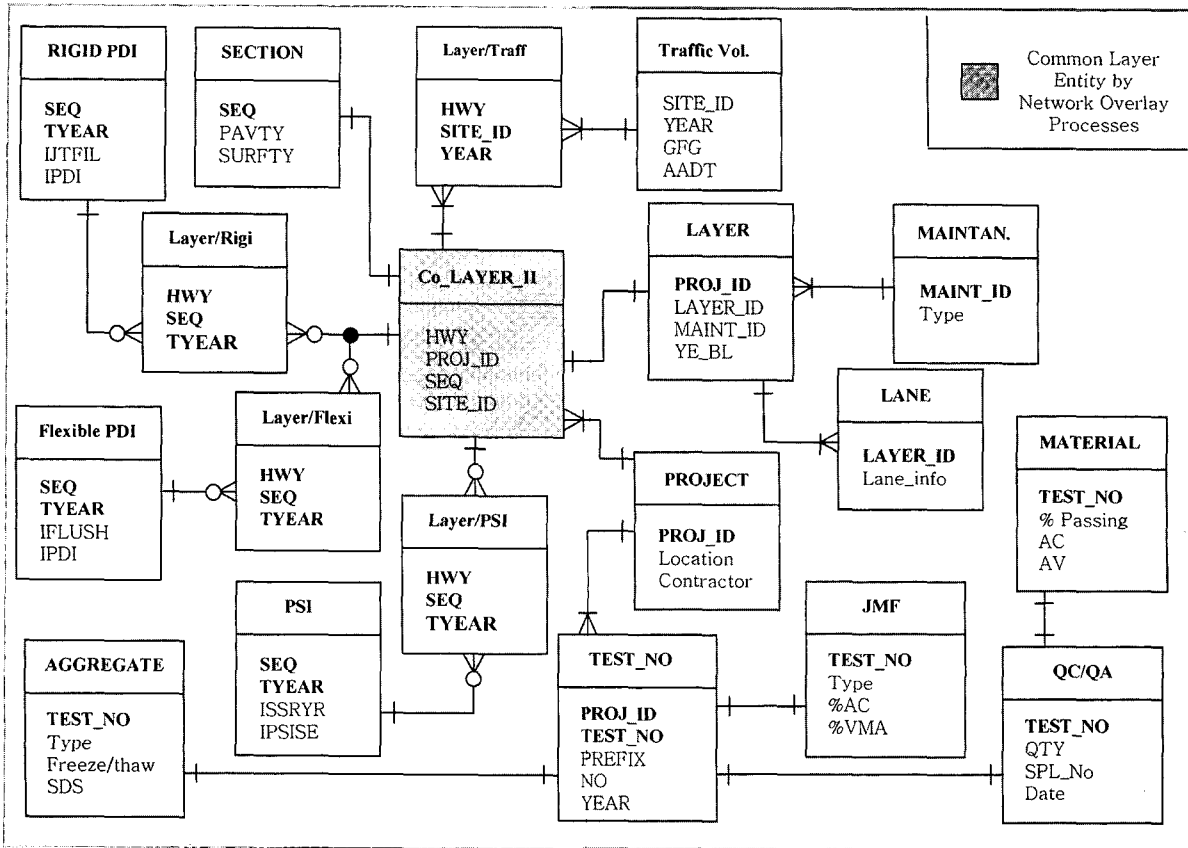
Kimball, R. (1998). *The Data Warehouse, Life Cycle Toolkit, expert methods for designing, developing, and deploying data warehouses*. Wiley, Inc., New York.

Matthias, et al. (2000). *Fundamentals of Data Warehouses*, Springer, New York.

Song, Il-yeol and Kelly Levan-Shultz. (1999). Data Warehouse Design for E-Commerce Environments. *Lecture notes in computer science*. ISSUE 1727, Springer Verlag, New York.

Mennecke, B. and Higgins, G. (1999). Spatial Data in the Data Warehouse : A Nomenclature for Design and Use. *Proceedings of the Americas conference on information systems*. Americas conference 5th, Milwaukee, WI

Appendix I. A conceptual database model (E-R Model) for developing PPAS



Appendix II. Complete dimensional models for pavement performance analysis system

